

ImMesh: An Immediate LiDAR Localization and Meshing Framework

Jiarong Lin*, Chongjiang Yuan*, Yixi Cai, Haotian Li, Yuying Zou, Xiaoping Hong and Fu Zhang

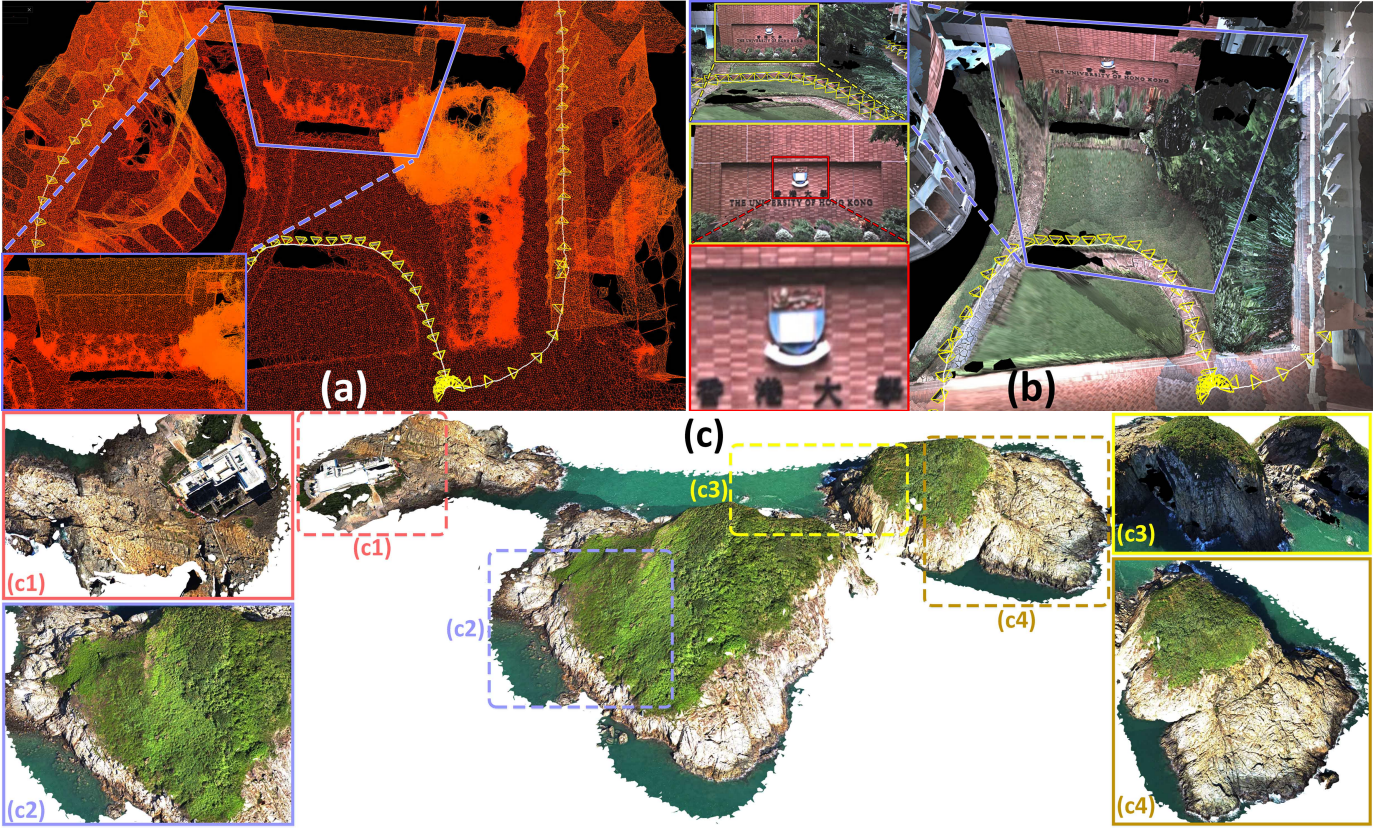


Fig. 1: (a) shows the triangle mesh that is online reconstructed by our proposed work ImMesh, where the white path is our sampling trajectory, and the yellow frustums are the estimated sensor pose. In (b), we use the estimated camera poses (the yellow frustums) of R³LIVE for texturing the mesh with the collected images. Based on ImMesh, we developed a lossless texture reconstruction application, with one of our results shown in (c). Our accompanying video that shows details of this work is available on YouTube: youtu.be/pzT2fMwz428.

Abstract—In this paper, we propose a novel LiDAR-(inertial) odometry and mapping framework to achieve the goal of simultaneous localization and meshing in real-time. This proposed framework termed ImMesh comprises four tightly-coupled modules: receiver, localization, meshing, and broadcaster. The localization module utilizes the preprocessed sensor data from the receiver, estimates the sensor pose online by registering LiDAR scans to maps, and dynamically grows the map. Then, our meshing module takes the registered LiDAR scan for incrementally reconstructing the triangle mesh on the fly. Finally, the real-time odometry, map, and mesh are published via our broadcaster. The key contribution of this work is the meshing module, which represents a scene by an efficient hierarchical voxels structure, performs fast finding of voxels observed by new scans, and reconstructs triangle facets in each voxel in an incremental manner. This voxel-wise meshing operation is delicately designed for the purpose of efficiency; it first performs a dimension reduction by projecting 3D points to a 2D local plane contained in the voxel, and then executes the meshing operation

with pull, commit and push steps for incremental reconstruction of triangle facets. To the best of our knowledge, this is the first work in literature that can reconstruct online the triangle mesh of large-scale scenes, just relying on a standard CPU without GPU acceleration. To share our findings and make contributions to the community, we make our code publicly available on our GitHub: github.com/hku-mars/ImMesh.

Index Terms—Mapping, 3D reconstruction, SLAM

I. INTRODUCTION

Recently, the wide emergence of 3D applications such as metaverse [1, 2], VR/AR [3], video games, and physical simulator [4] has enriched human lifestyle and boosted productive efficiency by providing a virtual environment that alike the real world. These applications are built upon triangle meshes that represent complex geometry of real-world scenes. Triangle mesh is the collection of vertices and triangle facets, which serves as a fundamental tool for objects modeling in most existing 3D applications. It can not only simplify significantly the process and boost the speed of rendering [5, 6] and ray-tracing [7], but also play an irreplaceable role in collision detection [8, 9], rigid-body dynamics [10, 11], dense mapping and surveying [12], sensor simulation [13], etc. However, most existing mesh is manufactured by skillful 3D modelers with the help of computer-aided design (CAD) software (e.g.,

*These two authors contribute equally to this work.

J. Lin, C. Yuan, Y. Cai and F. Zhang are with the Department of Mechanical Engineering, The University of Hong Kong, Hong Kong SAR, China. {jiarong.lin, ycjl, yixicai, haotianl, zycici, fuzhang}@connect.hku.hk

C. Yuan and X. Hong are with the School of System Design and Intelligent Manufacturing, Southern University of Science and Technology, Shenzhen, People's Republic of China. {yuancj2020, hongxp}@sustech.edu.cn

Solidworks [14], blender [15], etc.), which limits the mass production of large-scene meshing. Hence, developing an efficient mesh method that could reconstruct large scenes in real-time draws increasing research interests and serves as a hot topic in the community of 3D reconstruction.

Performing mesh reconstruction in real-time is particularly important in practical usages. Firstly, online mesh reconstruction indeed makes data collection effective by providing a live preview, which is quite important to give a reference for users. Especially for those non-expert users, a live preview can serve as a feedback about which parts of the scene have been reconstructed in good quality already and where additional data is needed. Secondly, online mesh reconstruction can immediately output the mesh of scene once data collection is complete, saving additional post-processing time of offline mesh reconstruction and hence boosts the productivity of mass production. Thirdly, it is particularly important for those real-time applications, especially for fully autonomous robotic applications, a real-time update of mesh can provide better maps with denser representation and of higher accuracy, which can enable the agent to better navigate itself.

Reconstructing the mesh of large scenes from sensor measurements in real-time remains one of the most difficult problems in the fields of computer graphics, 3D vision, and robotics, which require reconstructing the surfaces of scenes with triangle facets that are adjacently connected by edges. This is a challenging problem that needs to build the geometry structure with very high accuracy, and the triangle facet should be reconstructed on surfaces that actually exist in the real world. Besides, a good mesh reconstruction method should also suppress the appearance of holes on the reconstructed surface, and avoid the reconstruction of triangle silver (i.e., the noodle-like triangles that have a shard acute angle). Real-time mesh reconstruction in large scenes is even more challenging as it further requires the reconstruction to operate in an efficient, incremental manner.

In this work, we propose a real-time mesh reconstruction framework termed ImMesh to achieve the goal of simultaneous localization and meshing on the fly. This is a well-engineered system that is comprised of four tightly-coupled modules delicately designed for efficiency and accuracy. To the best of our knowledge, this is the first work in literature that can reconstruct the triangle mesh of large-scale scenes online and with a standard CPU. The main contributions of our work are:

- We propose a novel system that can estimate the sensor pose and reconstruct the mesh of the surrounding environment both online. Its localization is built upon our previous work VoxelMap [16], which can estimate the sensor pose of better efficiency and higher accuracy over its counterparts (e.g., FAST-LIO2 [17], SUMA [18], MULLS [19], Lego-LOAM [20], etc.). Its meshing module implements a novel mesh reconstruction approach, which efficiently reconstructs the mesh in an incremental manner, and can achieve real-time performance in large-scale scenarios on a standard desktop CPU.
- We implement a novel mesh reconstruction method in our meshing module, which directly utilizes the registered LiDAR point as mesh vertices, online reconstructing the

triangle facets (i.e., the indices of three triangle points) in an incremental manner. Specifically, our meshing module first utilizes an efficient hierarchical voxel data structure for fast finding of voxels containing points in new scans. Then, the voxel-wise 3D meshing problem is converted into a 2D one by performing dimension reduction. Finally, the triangle facets are incrementally reconstructed with the voxel-wise mesh pull, commit and push steps.

- We evaluate the runtime performance and meshing accuracy of ImMesh by conducting extensive experiments. We first verify the overall performance by presenting live video demonstrations of how the mesh is immediately reconstructed in the process of data collection. Then we extensively tested ImMesh with four public datasets collected with different types of LiDARs in various scenes. Finally, we evaluate the runtime performance and meshing accuracy of ImMesh by comparing them against existing baselines.
- We additionally demonstrate how real-time meshing can be applied in potential applications by presenting two practical examples. We first show that ImMesh can be applied for LiDAR point cloud reinforcement, which can output the reinforced points in regular pattern, and with higher density and wider FoV compared to raw LiDAR scan. Then, we combined ImMesh and our previous work R³LIVE [21, 22] to achieve the goal of losslessly texture reconstruction of scenes (see Fig. (b)), which is useful for rapid field surveying.
- We make ImMesh publicly available on our GitHub: github.com/hku-mars/ImMesh¹ for sharing our findings and making contributions to the community,

II. RELATED WORKS

In this section, we discuss the related works of mesh reconstruction based on 3D point cloud, which are closely related to this work. Depending on whether the reconstruction processes can perform online, we categorize existing mesh reconstruction methods into two classes: offline methods and online methods.

A. Offline mesh reconstruction

The offline methods usually require a global map in prior, for example, the full registered point cloud of the scene. Then, a global mesh reconstruction process is used to build the mesh. In this category, the most notable works include: methods based on Poisson surface reconstruction (Poisson-based), and methods based on Delaunay tetrahedralization (i.e., 3D Delaunay triangulation) and graph cut (Delaunay-based).

1) *Poisson surface reconstruction (Poisson-based)*: Given a set of 3D points with oriented normals that are sampled on the surface of a 3D model, the basic idea of Poisson surface reconstruction [23, 24] is to cast the problem of mesh reconstruction as an optimization problem, which solves for an approximate indicator function of the inferred solid

¹Our codes will be released as this work is accepted.

whose gradient best matches the input normals. Then, the continuous isosurface (i.e., the triangle mesh) is extracted from the indicator function using the method [25, 26] that is similar to adaptations of the Marching Cubes [27] with octree representations.

Benefiting from this implicit representation, where the mesh is extracted from the indicator function instead of being estimated directly, Poisson surface reconstruction can produce watertight manifold meshes and is resilient to scanner noise, misalignment, and missing data. Hence, in the communities of graphics and vision, these types of methods [23, 24, 28] have been widely used for reconstructing the mesh from given 3D scanned data.

2) *Delaunay triangulation and graph cut (Delaunay-based)*: In the category of offline mesh reconstruction methods, approaches [29]–[31] based on Delaunay tetrahedralization and graph cut are also been widely used for generating the mesh, based on the reconstructed 3D point cloud and the sensor’s poses. The basic idea of this class of methods is first to build a tetrahedral decomposition of 3D space by computing the 3D Delaunay triangulation of the 3D point set. Then, the Delaunay tetrahedra was labeled as “inside” or “outside” with the globally optimal label assignment (i.e., the graph cut). Finally, the triangle mesh can be extracted as the interface between these classes.

Besides these two classes of methods, there exist other offline surface mesh reconstruction algorithms such as the ball-pivoting algorithm [32] that have been proposed in past decades. However, they are usually not the first choice of consideration due to the lower precision and worse efficiency compared to Poisson- and Delaunay-based methods [33].

Unlike these offline mesh reconstruction methods, our proposed work ImMesh can perform online in an incremental manner without the whole point cloud of the scene. Besides, ImMesh also achieves a satisfactory meshing accuracy that is higher than Poisson-based and slightly lower than Delaunay-based (see our experimental results in Section VIII-C).

B. Online mesh reconstruction

1) *Voxel volume-based methods (TSDF-based)*: The online mesh reconstruction method is predominated by TSDF-based methods, which represent the scene in a voxel volumetric theme. These methods implicitly reconstruct the mesh in a two-step pipeline, which first establishes the truncated signed distance to the closest surface of voxels, then extracts the continuous triangle mesh by leveraging the Marching Cubes algorithm [27] from volumes. TSDF-based methods are popularized by KinectFusion [34], with many follow-up works focused on scaling this approach to larger scenes [35, 36], adding multi-resolution capability [37, 38], and improving efficiency [39]–[41]. Since these classes of methods can be easily implemented with parallelism, they can achieve real-time performance with the acceleration of GPUs.

Compared to these methods, our work ImMesh shows several advantages: Firstly, in ImMesh, the triangle mesh is directly reconstructed from the point cloud in one step, while for TSDF-based methods, the mesh is implicitly built in a two-step pipeline (i.e., SDF update followed by a mesh extraction).

Secondly, ImMesh is able to output the mesh in scan-rate (i.e., sensor sampling rate), while the mesh extraction of TSDF-based methods is usually at a lower rate. Thirdly, ImMesh achieves real-time performance by just running on a standard CPU, while TSDF-based methods need GPU acceleration for real-time SDF update. Lastly, TSDF-based methods require adequate observation for the calculation of SDF of each voxel w.r.t. the closet surface, which needs the data to be sampled by a depth sensor of high resolution and moving at a low speed. On the contrary, our work exploits high-accuracy LiDAR points for meshing and is robust to points data of low density.

2) *Surfel-based mesh reconstruction*: Besides TSDF-based methods, another popular approach is to represent the scene with a set of points or surfels (e.g., oriented discs). For example, in work [36, 42, 43], the maps are reconstructed with point-based representation, and its “surface” is rendered with the approaches of “point-based rendering” that originated from the communities of computer graphics [44]–[46]. Besides, in work [47], the high-quality map is reconstructed with surfel-based representations (i.e., use patches), such forms of mapping representation are popularized in works [48]–[51]. To reconstruct a dense map, these classes of methods need a large number of points or tiny patches to represent the surface of models. This is an inefficient representation that has high usage of system memory and computation resources. In contrast, our work reconstructs the surface of models with triangle mesh, which uses triangle facets of proper size that are adjacently connected by edges. It is the most efficient solid-model representation that has been widely adopted in most modern 3D software.

Compared with works reviewed above, our proposed work is in a class by itself, which contains the following advantages:

- It is an online mesh reconstruction method that reconstructs the triangle mesh in an incremental manner. It can achieve real-time performance in large-scale scenes (e.g., traveling length reaches 7.5 km) by just running on a standard desktop CPU.
- It explicitly reconstructs the triangle mesh by directly taking the registered LiDAR points as meshing vertices, performing the voxel-wise meshing operation as each new LiDAR scan is registered.
- It is delicately designed for the purpose of efficiency, and can achieve satisfactory meshing precision comparable to existing high-accuracy offline methods.

III. SYSTEM OVERVIEW

Fig. 2 depicts the overview of our proposed system (ImMesh), which consists of a map structure and four modules that work jointly to achieve the goal of simultaneous localization and meshing in real-time. As shown in Fig. 2, from left to right are: *receiver* (in red), *localization* (in orange), *map structure* (in green), *meshing* (in blue) and *broadcaster* (in purple).

In the rest sections, we will first introduce our *map structures* in Section IV, which will show the detail of the data structures that will be used in other modules. Next, we will

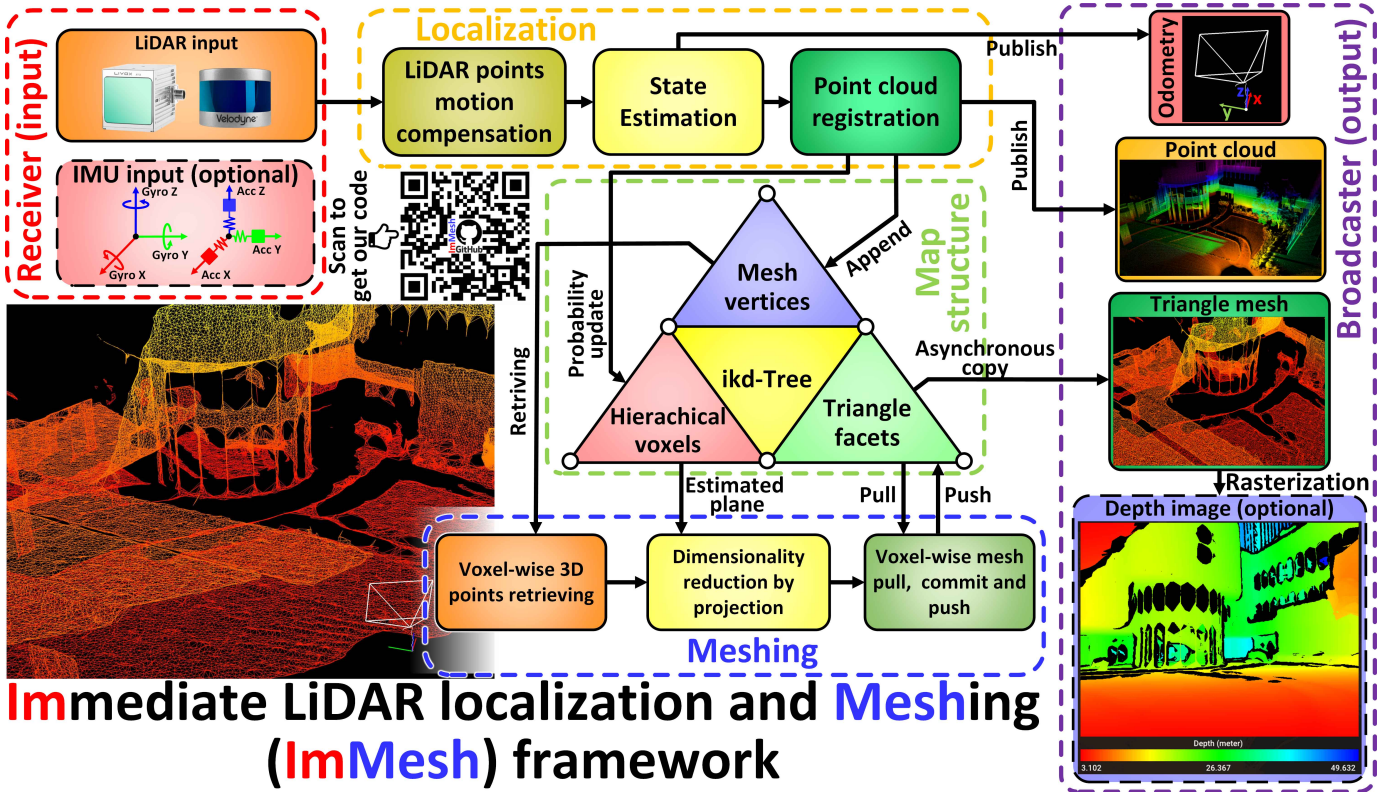


Fig. 2: This figure shows the overview of our proposed work ImMesh, which utilizes the raw input sensor data to achieve the goal of simultaneous localization and meshing. It is constituted by four tightly-coupled modules and a map structure, from left (input) to right (output) are: *receiver* (in red), *localization* (in orange), *map structure* (in green), *meshing* (in blue) and *broadcaster* (in purple).

introduce our receiver and localization module in Section V. Then, we will present how our *meshing* modules work in Section VI. Finally, in Section VII, we will introduce the *broadcaster* module, which publishes the localization and meshing results to other applications.

IV. MAP STRUCTURES

As shown by the *map structures* (in green) in Fig. 2, we design four data structures, including a structure of meshing vertices, a structure of triangle facets, an incremental kd-Tree (ikd-Tree) for k nearest neighbors (kNN) search and down-sampling, and a hierarchical-voxels structure representing the 3D space.

A. Mesh vertices

In ImMesh, mesh vertices are the points that constitute the geometric structure (shape) of mesh. All mesh vertices are stored in a global list. For the i -th entry of the list that represents vertex \mathbf{V}_i , it contains the following elements:

- Its 3D position $\text{Pos}(\mathbf{V}_i) \in \mathbb{R}^3$ in global frame (i.e., the first LiDAR frame).
- The index(id) of this vertex $\text{Id}(\mathbf{V}_i) = i$, which is the unique identification that indicates this point is the i -th point that appended to map.
- The list of pointers to triangles facets \mathbf{T} whose vertices contain \mathbf{V}_i :

$$\text{Tri_list}(\mathbf{V}_i) = \{\text{Ptr}(\mathbf{T}_{i_1}), \text{Ptr}(\mathbf{T}_{i_2}), \dots, \text{Ptr}(\mathbf{T}_{i_m})\} \quad (1)$$

where we use function $\text{Ptr}(\cdot)$ to denote the pointer (i.e., C++ pointer) of (\cdot) .

B. Triangle facets

In ImMesh, a triangle facet describes a small surface that exists in the reconstructed scene. It is maintained online by our *meshing* module (see Section VI) and is asynchronously copied to the *broadcaster* module for publishing. A triangle facet \mathbf{T} contains the following elements:

- The sorted indices $\text{Pts_id}(\mathbf{T})$ of three points that form this triangle:

$$\text{Pts_id}(\mathbf{T}) = \{i, j, k\}, \quad i < j < k \quad (2)$$

- The center $\text{Center}(\mathbf{T})$ and normal $\text{Norm}(\mathbf{T})$ (both in the global frame) of this triangle:

$$\text{Center}(\mathbf{T}) = (\text{Pos}(\mathbf{V}_i) + \text{Pos}(\mathbf{V}_j) + \text{Pos}(\mathbf{V}_k)) / 3 \quad (3)$$

$$\text{Norm}(\mathbf{T}) = \mathbf{n} / (|\mathbf{n}|) \quad (4)$$

$$\mathbf{n} = (\text{Pos}(\mathbf{V}_i) - \text{Pos}(\mathbf{V}_j)) \times (\text{Pos}(\mathbf{V}_k) - \text{Pos}(\mathbf{V}_j)) \quad (5)$$

C. Incremental kd-Tree (ikd-Tree)

We maintain an incremental kd-tree to enable the fast kNN search of mesh vertices. The ikd-Tree is proposed in our previous work [17, 52], which is an efficient dynamic space partition data structure for fast kNN search. Unlike existing static kd-tree (e.g., kd-tree implemented in PCL [53] and FLANN [54]) that require rebuilding the entire tree at each update, ikd-Tree achieves lower computation time by updating the tree with newly coming points in an incremental manner. In ImMesh, we use the ikd-Tree for:

- Downsample the point cloud density to keep the minimum distance between any of two mesh vertices for maintaining the triangle mesh at a proper resolution.

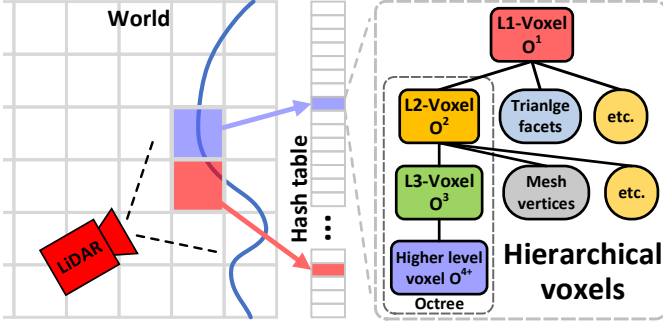


Fig. 3: In ImMesh, the world is partitioned by hierarchical voxels. We compactly store, access, and update the voxels in a spatial hashing scheme.

- Enable the vertex dilation operation in our voxel-wise meshing operation (see Section VI), which can erode the gaps between neighbor voxels.

D. Hierarchical voxels

In our map, we partition the 3D space with hierarchical voxels. As shown in Fig. 3, lower-level voxels contain those of higher levels. These voxels of different levels are designed with different sizes and for various purposes: the lowest level (i.e., $L1$ -Voxel) has the largest voxel size, which partition the 3D space into small regions by uniform grids. Voxels in this layer maintain a hash table of pointers that point to the triangle facet whose center is located inside. This facilitates the *broadcaster* for asynchronous copying of these triangle facets (see Section VII-B). And, the size of the second layer (i.e., $L2$ -Voxel) is much smaller than the first layer, where the voxels in this layer store the mesh vertices that constitute the geometric structure of the mesh. Voxels of this layer allow the *meshing* module to fast retrieve all in-voxel mesh vertices for voxel-wise meshing operations (see Section VI). Lastly, as shown in Fig. 3, the $L2$ -Voxel and its sub-voxels form a typical octree data structure, which is used in our *localization* module for a further split of non-planar point clusters to achieve better pose estimation (see Section V).

1) $L1$ -Voxel O^1 : As illustrated in Fig. 3, we uniformly partition the 3D world into many small regions with $L1$ -Voxel. To avoid large memory consumption in allocating regular volumetric grids (e.g., in kinectFusion [34]), we compactly store, access, and update the voxels with a spatial hashing scheme alike [36]. We map the 3D world space into the hash table via a hash function $\text{Hash}(\cdot)$, where the hash function allows an efficient look-up of voxel blocks with the integer-rounded world coordinates. The array of pointers to Voxel is stored in the hash table.

$$\text{Hash}(x, y, z) = \text{Int_Hash}(x_i, y_i, z_i) \quad (6)$$

$$= \text{Mod}((x_i \cdot p_1) \oplus (y_i \cdot p_1) \oplus (z_i \cdot p_3), n) \quad (7)$$

$$x_i = \text{Round}(x * 100/r_x), \quad y_i = \text{Round}(y * 100/r_y) \quad (8)$$

$$z_i = \text{Round}(z * 100/r_z)$$

where x, y, z are coordinates of 3D space, x_i, y_i, z_i are specified integer rounded world coordinates, r_x, r_y, r_z are the voxel size in three dimensions, \oplus is the XOR operation, and function $\text{Mod}(a, b)$ is the calculation of integer a modulus another integer b . p_1, p_2, p_3 are three large prime numbers for

reducing the collision probability [36, 55], n is the hash table size. In our work, we set the value of p_1, p_2, p_3 and n as 116101, 37199, 93911 and 201326611, respectively.

Notice that the hash table is unstructured, indicating that the neighboring voxels are not stored spatially but in different parts of the buckets (shown in Fig. 3). Besides, for resolving the possible hash collision (i.e., two pieces of data in a hash table share the same hash value), we adopt the technique in [36], using the implementation of `unordered_map` container in C++ standard library (`std`) [56]. In this work, we access a $L1$ -voxel with a given 3D vector $\mathbf{p} = [x, y, z]^T \in \mathbb{R}^3$ by:

$$O^1 = \text{Get_L1_voxel}(\text{Hash}(\mathbf{p})) \quad (9)$$

Shown in Fig. 3, each $L1$ -Voxel contains the voxels of the higher hierarchical layer. To identify the work stage of $L1$ -Voxel, we use a flag to mark the status as either *Sync-required* or *Synced*. These two statuses indicate the update flag related to the data synchronization of triangle facets, as we will use in Section VI-E and Section VII-B.

For each $L1$ -Voxel, it stores and maintains a hash table of pointers pointing to a triangle facet whose center is located in the voxel. These pointers can be efficiently looked up via $\text{Int_Hash}(i, j, k)$ in (6), where i, j, k are the sorted indices (i.e., $i < j < k$) of three mesh vertices. These in-voxel triangle facets are maintained (i.e., added or erased) by the *meshing* module, and are asynchronously copied to *broadcaster* module for publishing to other applications.

2) $L2$ -Voxel O^2 and voxels of higher layer: $L2$ -Voxel is the second biggest container, which stores an array of points that point to all in-voxel mesh vertex, and contains the voxels of higher layer. It is used in both of our *localization* and *meshing* modules; in *localization* module, $L2$ -Voxel stores the in-voxel registered LiDAR points used to constitute planar features for estimating the sensor pose; in *meshing* module, $L2$ -Voxel enables fast retrieval of all in-voxel mesh vertices and provides the local estimated planar norm for projecting the 3D points into the 2D plane.

For a $L2$ -Voxel O^2 , it has a status flag indicating whether it has new points appended. To be detailed, O^2 is marked as *Activated* if this voxel has new mesh vertices registered from the latest LiDAR scan (see Section V-C). And the *Activated* flag is reset as *deactivated* after the voxel-wise meshing operation is performed on this voxel (see Section VI-G).

Similar to (9), we achieve fast access to a $L2$ -Voxel with a given 3D vector $\mathbf{p} = [x, y, z]^T \in \mathbb{R}^3$ through hash tables:

$$O^2 = \text{Get_L2_voxel}(\text{Hash}(\mathbf{p})) \quad (10)$$

where the hash function $\text{Hash}(\cdot)$ in (10) and (9) are distinguished with different voxel size r_x, r_y, r_z in (8).

For voxels of higher layer, e.g., voxel O^3 of the third layer and higher O^{3+} , they are designed to partition the non-planar points (in voxels of the higher layer) with a smaller spatial size (higher resolution), which make them more likely to construct a planar feature for localization, as introduced in the coming section.

Notice that the voxels of $L2$ - and higher levels construct an Octree. We access the voxels of the third layer and higher in a way similar to Octree [57].

V. RECEIVER AND LOCALIZATION

The *receiver* module is designed for processing and packaging the input sensor data. As shown in the red box of Fig. 2, our *receiver* module receives the streaming of LiDAR data from live or offline recorded files, processes the data to a unified data format (i.e., customized point cloud data) that make ImMesh compatible with LiDARs of different manufacturers, scanning mechanisms (i.e., mechanical spinning, solid-state) and point cloud density (e.g., 64-, 32-, 16-lines, etc.). Besides, if the IMU source is available, our *input* module will also package the IMU measurements within a LiDAR frame by referring to the sampling time.

The *localization* module utilizes the input data stream of *receiver* module, real-time estimating the sensor poses of 6 DoF and registering the points to map. Our *localization* module is built upon our previous work VoxelMap [16], which represents the surrounding environment with the probabilistic representation, estimating pose with an iterated Kalman filter by maximum a posterior.

In designing our localization module, we have noticed that a number of works appeared in the literature recently, which utilize the reconstructed mesh for improving the localization accuracy of both visual-slam (e.g., [58]) and LiDAR-slam system (e.g., [59]–[61]). However, in ImMesh, the online reconstructed mesh is not used in our localization module because: 1) our mesh is reconstructed with points that are registered by the localization module, re-using mesh in localization will take more computation efforts and bring extra latency in publishing the estimated pose. 2) the accuracy of our localization module is indeed enough for most of the robotics and surveying applications, which achieve the localization results of better efficiency and higher accuracy compared to its counterparts like FAST-LIO2 [17], SUMA [18], MULLS [19], Lego-LOAM [20], etc. Despite this, we hold a positive attitude toward seeking the possibility of improving the localization accuracy with our online reconstructed mesh in future work.

A. Voxel map construction

Our *localization* is built by representing the surrounding environment with the probabilistic representation, which counts both LiDAR measurement noises and sensor pose estimation errors, and constructs the voxel-volumetric maps in a coarse-to-fine adaptive resolution manner. But, in this work, we mainly focus our attention on how to real-time reconstruct the triangle mesh of the scene, and avoid introducing too many complicated noise analyses that might confuse the reader. We only discuss those processes in *localization* module that are closely related to our *meshing* module in this paper. For the detailed modeling and analysis of LiDAR’s measurement noise, we recommend our readers to our previous work VoxelMap [16].

For a LiDAR sampling point, we first compensate the in-frame motion distortion with an IMU backward propagation introduced in [17]. Let us use ${}^L\mathbf{p}_i$ denote i -th LiDAR sampling point after motion compensation, it is registered to world frame as ${}^W\mathbf{p}_i$ with the estimated sensor pose $({}^W\mathbf{R}, {}^W\mathbf{t}) \in SE(3)$:

$${}^W\mathbf{p}_i = {}^W\mathbf{R} {}^L\mathbf{p}_i + {}^W\mathbf{t} \quad (11)$$

The register LiDAR points are stored inside the voxels (e.g., $L2$ -Voxel), let us consider the distribution of points ${}^W\mathbf{p}_i$ ($i = 1, \dots, N$) that are located inside the $L2$ -Voxel. We have the points covariance matrix \mathbf{A} calculated as:

$$\bar{\mathbf{p}} = \frac{1}{N} \sum_{i=1}^N {}^W\mathbf{p}_i, \quad \mathbf{A} = \frac{1}{N} \sum_{i=1}^N ({}^W\mathbf{p}_i - \bar{\mathbf{p}}) ({}^W\mathbf{p}_i - \bar{\mathbf{p}})^T \quad (12)$$

where the symmetric matrix \mathbf{A} depicted the distribution of points. Let us perform the eigendecomposition of matrix \mathbf{A} :

$$\mathbf{A}\mathbf{U} = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3], \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \quad (13)$$

where $\lambda_1, \lambda_2, \lambda_3$ are the eigenvalues and $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ are the correspondent eigenvectors.

If the minimum eigenvalue λ_3 is less than a specified threshold, which indicates that the points inside this voxel are distributed on a thin planar surface, we regard all points ${}^W\mathbf{p}_i$ ($i = 1, \dots, N$) as a planar feature. Otherwise, this voxel will be further subdivided into voxels of higher level with smaller size (i.e., $L3$ -, $L4$ -, ..., voxel) until: 1) the tiers of layer reach bound (set as tier-5 for our work) 2) the minimum eigenvalue of points covariance matrix \mathbf{A} of a voxel smaller than a given threshold.

If points ${}^W\mathbf{p}_i$ ($i = 1, \dots, N$) inside the voxel indeed forming a planar feature, whose minimum eigenvalue λ_3 of its points covariance matrix \mathbf{A} less than a specified threshold. We represent this planar feature by using its normal vector \mathbf{n} and a point \mathbf{q} that lies in this plane. The normal vector is well known as the eigenvector w.r.t. associated with the minimum eigenvalue λ_3 , i.e., $\mathbf{n} = \mathbf{u}_3$ in (13). And point $\mathbf{q} = \bar{\mathbf{p}}$ is calculated in (12).

B. State Estimation

1) *Point-to-plane residual*: In our *localization* module, we solve the sensor pose by minimizing the Point-to-plane residual. Given a LiDAR point ${}^W\mathbf{p}_i$ predicted in the world frame with the pose prior, we first find which voxel it lies in by hashing with (10). Then, all the contained voxels of higher layers are polled for a possible match with the point. Specifically, let a sub-voxel contains a plane with normal \mathbf{n}_i and center \mathbf{q}_i , we calculate the point-to-plane distance:

$$d_i = \mathbf{n}_i^T ({}^W\mathbf{p}_i - \mathbf{q}_i) \quad (14)$$

If point \mathbf{p}_i lies on the candidate plane with this point-to-plane distance d_i falling within the $3\text{-}\sigma$ bound of the plane measurement noise, we treat this point-to-plane pair as an effective match and add it to the residuals for estimating the sensor pose.

2) *LiDAR pose estimation by maximum a posterior (MAP)*: We build a LiDAR(-inertial) odometry system based on an iterated error-state extended Kalman filter (IESKF) similar to that derived in our previous works [17, 62]. Assume that we are given a state estimation prior, which is provided from a constant velocity assumption for LiDAR-only odometry (e.g., Kitti dataset in our Experiment-1), or from IMU propagation for LiDAR-inertial odometry (e.g., NCLT-dataset,

NTU-dataset, R³LIVE-dataset our self-collected data in Section VIII). This will be fused with the point-to-plane distance matched in Section V-B1 to form a maximum a posteriori (MAP) estimation. Then, we solve this MAP problem by leveraging an IESKF, which leads an optimal state estimation of sensor pose (${}^W_L \mathbf{R}, {}^W_L \mathbf{t}$) that is used for registering the LiDAR point with (11).

C. Point cloud registration

After the state estimation, we perform the point cloud registration for transforming each measurement point ${}^L \mathbf{p}_i$ from LiDAR frame to global frame (i.e., the first LiDAR frame) with (11). This registered point cloud is then used for: 1) Published to other applications with our broadcaster. 2) Use for updating the probabilistic voxel map. 3) Appended to *map structure* that serves as the mesh vertices for shaping the geometry structure of our online reconstructed triangle mesh.

1) *Update of voxel map*: The registered LiDAR points are used for constructing the probabilistic voxel map by updating the point distributions (i.e., \mathbf{A} in (12)), planar parameters (i.e., \mathbf{n}, \mathbf{q}) and the correspondent uncertainties of all possible hierarchical voxels. For the details of this voxel map update, we refer the reader to our previous work [16]. Besides, if a new register point does not lie on an existing $L2$ (or $L1$) voxel, a new $L2$ (or $L1$) voxel will be created and added to the hash table, after, this point will be added to the newly constructed voxel.

2) *Append of mesh vertices*: The registered LiDAR points are also used for forming the meshing vertices in *map structures*. To be detailed, we first leverage a voxel-grid filter for downsampling register LiDAR point cloud. Then, to avoid the appearance of tiny triangles in reconstructing the mesh, we leverage the ikd-Tree (see Section IV-C) for keeping the minimum distance between any of two meshing points. That is, for each register LiDAR point ${}^W \mathbf{p}_i$ in global frame, we search for the nearest mesh vertex in *map structure* with ikd-Tree, if the euclidean distance this point and the searched vertex smaller than a given threshold, we will discard this point. Otherwise, this point will be used for: 1) Constructing a new mesh vertex \mathbf{V}_i , where i is the unique index that indicates \mathbf{V}_i is the i -th appended vertex. 2) Appending the pointer of \mathbf{V}_i to the ikd-Tree. 3) Pushing back the pointer $\text{Ptr}(\mathbf{V}_i)$ to the point array of the $L2$ -Voxel \mathbf{O}_j^2 that \mathbf{V}_i located in. After, the status flag of \mathbf{O}_j^2 is set as *activated* for notifying the meshing module for performing the voxel-wise re-meshing operation (see Section VI).

VI. MESHING

In ImMesh, our meshing module takes the registered LiDAR scan for incrementally reconstructing the triangle mesh on the fly. We explicitly reconstruct the triangle mesh by directly utilizing 3D registered LiDAR points as mesh vertex with two considerations: 1) The points sampled by LiDAR and registered via the ICP-based methods [63, 64] have very high positional accuracy. Hence, they are capable of shaping the geometric structure of the mesh. 2) A LiDAR measurement point naturally lies on the surface of the detected object. That

is, a laser pulse is emitted from the infrared transmitter and reflected by the surface of the detected object. The returned pulse is captured by the receiver, and the ranging distance of the sensor from the surface is finally calculated by counting the time of flight (ToF).

A. Goals and requirements

With the accurate mesh vertices appended from the point cloud registration in Section V-C, the problem of online mesh reconstruction is converted to another goal, which is to seek a proper way for real-time reconstructing the triangle facets with a growing 3D point set. However, to the best of our knowledge, this is a new area in the community that has not been explored yet. Given a set of growing 3D points, our *meshing* module is designed to incrementally reconstruct the triangle facets considering the following four major requirements:

Firstly, precision is our prior consideration. For each reconstructed triangle facet that represents the surface of the scene, we require it to lie on an existing plane.

Secondly, the reconstructed mesh should be hole-less. In the dense reconstruction of the surface triangle mesh, the appearance of holes is unacceptable. To be detailed, these holes lead to the wrong results in the rasterization of the depth image, which wrongly rasterizes the surfaces behind a real object to the front. Consequently, robotic applications based on our meshing result might lead to severe accidents (e.g., crashing into a wall). Besides, the holes on surfaces make the whole reconstructed map unsightly and chaotically.

Thirdly, the reconstruction of triangle mesh should avoid constructing sliver triangles. The sliver triangle (i.e., the noodle-like triangle), as defined in the communities of computer graphic [65]: whose area is so thin that its interior does not contain a distinct span for each scan line, has some undesired properties in the field of computer graphics. For example, these noodle-like triangles would cause some errors in the numerical analysis on them [66]. Besides, these unfavorable properties cause troubles in the pipelines of rendering (e.g., rasterization, texturing, and anti-aliasing [5, 6, 67]), Which leads to the loss of accuracy in calculating (e.g., depth testing, interpolation, etc.) the pixel values distributed near the sharp angle [6, 68, 69].

Lastly, the complexity of triangle mesh reconstruction should be computationally efficient to meet the requirement of real-time applications. The time consumption of each meshing process should not exceed the sampling duration of two consecutive LiDAR frames.

B. Challenges and approaches

To achieve our goals of dense incremental meshing with the four requirements listed above, our system is proposed based on a deep analysis of the challenges. The challenges and corresponding scientific approaches are briefed below:

The first challenge is that the global map is continuously grown by the newly registered LiDAR points, with each update of a LiDAR scan only affecting part of the scene. Hence, for an incremental mesh reconstruction method, it should be able to process only those parts of the scene with new

points appended in. In our work, we incrementally perform the mesh reconstruction with a mechanism similar to *git* [70]. For each incremental mesh update, we first retrieve the data of the voxels with new mesh vertices appended via the *pull* step (detailed in Section VI-E1). Then, an efficient voxel-wise meshing algorithm is executed to reconstruct the mesh with these data. The incremental modifications of newly reconstructed results w.r.t. pulled results are calculated in our *commit* step (detailed in Section VI-E2). Finally, these incremental modifications are merged to the global map via our *push* step (detailed in Section VI-E3).

Given a set of 3D vertices, the second challenge is how to correctly and efficiently reconstruct the triangle facets representing the surfaces of the scene. Since it is hard to directly reconstruct mesh from these mesh vertices in 3D space, our work performs the meshing operation in 2D. To be detailed, for vertices located in a small region (i.e., in $L2$ -Voxel), we first project them into a proper plane (i.e., the estimated plane given by the *localization* module). The mesh of these 2D points is constructed using the 2D meshing algorithms and is recovered back to 3D (detailed in Section VI-D2).

C. Voxel-wise vertex retrieval

1) *Retrieval of in-voxel vertices*: To reconstruct the triangle mesh in an incremental manner, the first step is to retrieve the vertices that need to mesh with the newly added points. In ImMesh, we use the hierarchical voxels (see Section IV-D) for subdividing the 3D space into many regions. The flags that indicate the status of each $L1$ -Voxel are used for identifying whether a $L2$ -Voxel has newly appended mesh vertices (see Section V-C).

Take an *activated* $L2$ -Voxel O_i^2 as an example. We perform a voxel-wise meshing operation to reconstruct the triangle facets with all in-voxel vertices. Since the pointers of these vertices are stored in a pointer array attached to O_i^2 , we address these pointers to retrieve all in-voxel vertices, denoted as $\mathcal{V}_i^{\text{In}} = \{\mathbf{V}_{j_1}, \mathbf{V}_{j_2}, \dots, \mathbf{V}_{j_m}\}$.

2) *Vertex dilation*: In practice, if we perform the meshing operation with only the in-voxel mesh vertices, the gaps between neighborhood voxels will appear due to the absence of triangles facets across voxels, as shown in Fig. 4(b). Motivated by morphological operations (e.g., dilation and erosion) in digital image processing [71], we perform the 3D point cloud dilation for adding neighborhood points of $\mathcal{V}_i^{\text{In}}$ to erode the gaps between voxels, as shown in Fig. 4(a).

For vertex $\mathbf{V}_{i_j} \in \mathcal{V}_i^{\text{In}}$, we perform the radius-search operation by leveraging the ikd-Tree (see Section IV-C) for searching the nearest vertices of \mathbf{V}_{i_j} with their euclidean distance smaller than a given value d_r (usually set as $1/4$ of the size of $L2$ -Voxel). Using $\tilde{\mathcal{V}}_{i_j}$ to denote the searched neighbor vertices and \mathcal{V}_i to denote the dilated vertices, we have:

$$\forall \mathbf{V} \in \tilde{\mathcal{V}}_{i_j}, \quad \|\text{Pos}(\mathbf{V}) - \text{Pos}(\mathbf{V}_{i_j})\| \leq d_j \quad (15)$$

If $\mathbf{V} \in \mathcal{V}_{i_j}$ is not included in \mathcal{V}_i , we add \mathbf{V} by $\mathcal{V}_i = \mathcal{V}_i \cup \mathbf{V}$.

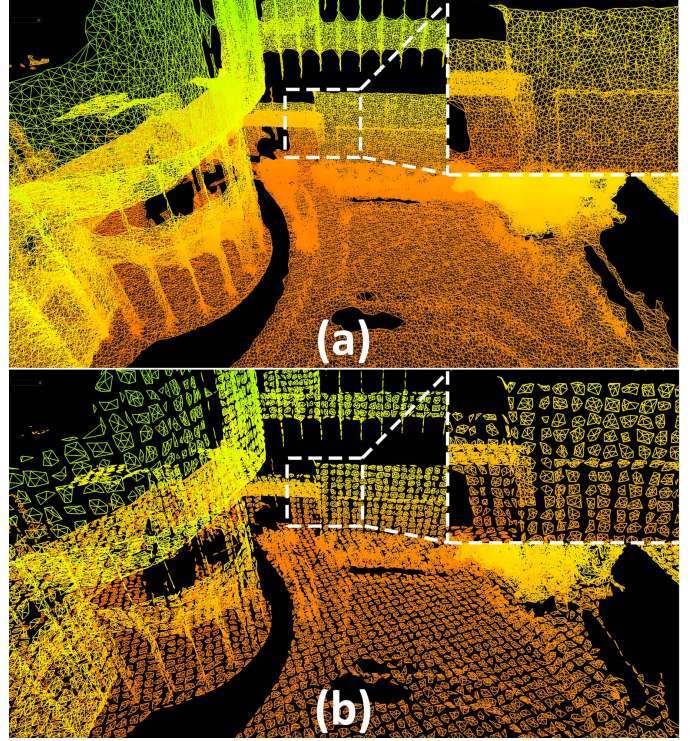


Fig. 4: The comparisons of mesh reconstruction with (a) and without (b) the vertex dilation.

The full algorithm of our voxel-wise vertex retrieval is shown in Algorithm 1.

Algorithm 1: Voxel-wise vertex retrieval of O_i^2

Input : The *activated* voxel O_i^2
Output: The retrieved vertex set \mathcal{V}_i
Start : Copy all in-voxel pointer list to $\mathcal{V}_i^{\text{In}}$.
 $\mathcal{V}_i = \mathcal{V}_i^{\text{In}}$.
1 **foreach** $\mathbf{V}_{i_j} \in \mathcal{V}_i^{\text{In}}$ **do**
2 $\tilde{\mathcal{V}}_{i_j} = \text{RadiusSearch}(\mathbf{V}_{i_j}, d_r)$
3 **foreach** $\mathbf{V} \in \tilde{\mathcal{V}}_{i_j}$ **do**
4 **if** $\mathbf{V} \notin \mathcal{V}_i$ **then**
5 $\mathcal{V}_i = \mathcal{V}_i \cup \mathbf{V}$
Return: The retrieved vertex set \mathcal{V}_i after dilation

D. Dimensional reduction by projection

With the mesh vertices \mathcal{V}_i retrieved from Algorithm 1, we introduce the voxel-wise mesh reconstruction.

1) *Projection 3D vertices on a 2D plane*: Since it is hard to directly mesh with \mathcal{V}_i distributed in 3D space in real-time, we simplify the 3D meshing problem to a 2D one by projecting \mathcal{V}_i on a suitable plane. Based on the analysis of the characteristics of \mathcal{V}_i , we provide two reasons to perform the dimensional reduction by projection, listed as follows: 1) For a 3D point sampled by LiDAR, it is distributed on a surface. Hence, for vertices \mathcal{V}_i retrieved from Algorithm 1 that distributed in a small region (i.e., in a $L2$ -Voxel O_i^2), they tend to form a planar-like point cluster. 2) For these planar-like point clusters, we can approximately mesh them in a 2D view on their lying surface. Imagine a 2D ant climbing on 3D surfaces solving this 3D problem in a 2D view, as shown in Fig. 5.

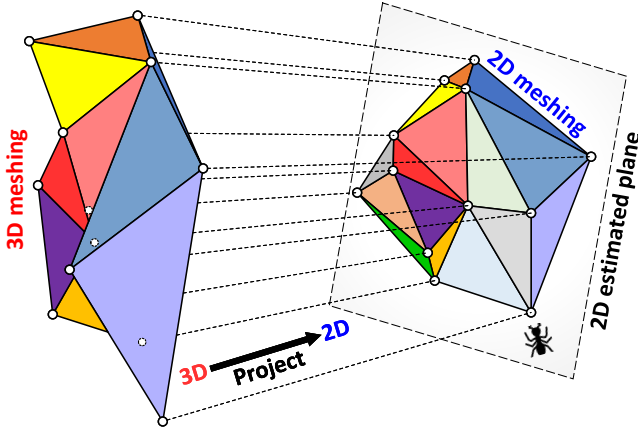


Fig. 5: In ImMesh, we reduce the 3D meshing problem to a 2D one by projecting the 3D points onto an estimated surface plane. Imagine a 2D ant climbing on 3D surfaces solving this 3D problem in a 2D view.

The plane (\mathbf{n}, \mathbf{q}) suitable for projection has already been calculated in our *localization* module in Section V-A. The norm \mathbf{n} of the plane is the eigenvector \mathbf{u}_3 that corresponds to the minimum eigenvalue λ_3 in (13), which is the eigendecomposition of point covariance matrix \mathbf{A} in voxel \mathbf{O}_i^2 . \mathbf{q} is the center points inside \mathbf{O}_i^2 .

Remark: Even though \mathbf{O}_i^2 might be further divided into voxel of lower layer by the *localization* module, the norm \mathbf{n} and \mathbf{q} of \mathbf{O}_i^2 is being updated at each new LiDAR frame.

For each vertex $\mathbf{V}_{i_j} \in \mathcal{V}_i$, we project it to plane (\mathbf{n}, \mathbf{q}) . The resultant 2D point \mathbf{u}_{i_j} is calculated as:

$$\mathbf{p}_{i_j} = [\phi, \rho]^T \in \mathbb{R}^2 \quad (16)$$

$$\phi = (\text{Pos}(\mathbf{V}_{i_j}) - \mathbf{q})^T \mathbf{u}_1, \quad \rho = (\text{Pos}(\mathbf{V}_{i_j}) - \mathbf{q})^T \mathbf{u}_2 \quad (17)$$

where $\mathbf{u}_1, \mathbf{u}_2$ are the other two eigenvectors in (13). We use $\mathcal{P}_i = \{\mathbf{p}_{i_1}, \mathbf{p}_{i_2}, \dots, \mathbf{p}_{i_m}\}$ to denote the 2D point set after projected from \mathcal{V}_i .

2) *Two-dimensional Delaunay triangulation:* After the projection, the dimension of 3D meshing problem is reduced to a 2D one, which can be solved by 2D Delaunay triangulation.

Given a set of 2D point \mathcal{P} , the two-dimensional triangulation problem is well known as introduced in [72, 73], which is to find \mathcal{T} of triangular facets s.t.: 1) Any of two facets are either disjoint or share a lower dimensional face (i.e., edge or point). 2) The set of facets in \mathcal{T} is connected with adjacency relation. 3) The domain $\mathbf{P}_{\mathcal{T}}$, which is the union of facets in \mathcal{T} , has no singularity². With these three useful properties, the 2D Delaunay triangulation has been widely applied for reconstructing dense facets with a given 2D point set (e.g., [74]).

As defined in [75, 76], the Delaunay triangulation $\text{Del}(\mathcal{P})$ of a 2D point set $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m\}$ is the geometric dual of the Voronoi diagram: there is an edge between two points \mathbf{u}_i and \mathbf{u}_j in the Delaunay triangulation if and only if their Voronoi cell $\mathbf{V}(\mathbf{u}_i)$ and $\mathbf{V}(\mathbf{u}_j)$ have a non-empty intersection. $\text{Del}(\mathcal{P})$ yields a triangulation of \mathcal{P} , which is a partition of the

²The union $\mathbf{U}_{\mathcal{T}}$ of all simplices in \mathcal{T} is called the domain of \mathcal{T} . A point in the domain of \mathcal{T} is said to be singular if its surrounding in $\mathbf{P}_{\mathcal{T}}$ is neither a topological ball nor a topological disc (view https://doc.cgal.org/latest/Triangulation_2/index.html of [72] for detail).

convex hull of \mathcal{P} into d -dimensional simplices (e.g., triangle in 2D, tetrahedra in 3D), as shown in Fig. 5.

Remark: The Voronoi cell $\mathbf{V}(\mathbf{u}_i)$ associated with the point \mathbf{p}_i is the region of space that is closer to \mathbf{u}_i than to all other points in \mathcal{P} :

$$\mathbf{V}(\mathbf{p}_i) = \{\mathbf{p} \in \mathbb{R}^d : \forall j \neq i, \|\mathbf{p} - \mathbf{p}_i\| \leq \|\mathbf{p} - \mathbf{p}_j\|\} \quad (18)$$

Considering our requirements in Section VI-A, we chose Delaunay triangulation to reconstruct the mesh for its remarkable properties as follows. Firstly, it is a 2D triangulation providing mesh with no hole leaf in the convex hull of \mathcal{P} , which satisfies our first requirement. Secondly, it naturally avoids sliver triangles by maximizing the minimum angles of the triangles in triangulation, which meets our second requirement. Finally, it is a fast algorithm suitable for real-time requirements. The algorithm complexity of n points is $\Omega(n \log(n))$ in 2D (p.s. $\Omega(n^2)$ in 3D) [77].

Let us use $\mathcal{T}_i = \text{Del}(\mathcal{P}_i) = \{\mathbf{T}_{i_1}, \mathbf{T}_{i_2}, \dots, \mathbf{T}_{i_n}\}$ to denote the triangle facets after the Delaunay triangulation $\text{Del}(\mathcal{P}_i)$. For each triangle facets $\mathbf{T}_{i_j} \in \mathcal{T}_i$, we retrieve the indices with (2): $\{\alpha, \beta, \gamma\} = \text{Pts_id}(\mathbf{T}_{i_j})$, indicating that this triangle is formed with 2D points $\{\mathbf{p}_{i_\alpha}, \mathbf{p}_{i_\beta}, \mathbf{p}_{i_\gamma}\}$. Returning back to 3D space, we constitute a triangle facet \mathbf{T}_{i_j} with vertices $\{\mathbf{V}_{i_\alpha}, \mathbf{V}_{i_\beta}, \mathbf{V}_{i_\gamma}\}$, as shown in Fig. 5.

E. Voxel-wise meshing with pull, commit and push

With the triangle facets \mathcal{T}_i constructed by the voxel-wise meshing operation, we incrementally merge \mathcal{T}_i to the existing triangle facets ${}^G\mathcal{T}$ in the *map structure*. This update is designed with a mechanism similar to *git* [70] (a version control software) that includes *pull*, *commit*, and *push* steps.

1) *Pull:* Given the vertices \mathcal{V}_i obtained from Algorithm 1, we retrieve the triangle facets $\mathcal{T}_i^{\text{pull}}$ from the *map structure*. The algorithm of *pull* step is shown in Algorithm 2.

Algorithm 2: Voxel-wise mesh pull.

Input : The retrieved vertex set \mathcal{V}_i from Algorithm 1

Output: The pulled triangle facets $\mathcal{T}_i^{\text{pull}}$

Start : $\mathcal{T}_i^{\text{pull}} = \{\text{null}\}$

1 **foreach** $\mathbf{V}_j \in \mathcal{V}_i$ **do**

2 Get all vertices related triangle set $\mathcal{T}_{\mathbf{V}_j} = \text{Tri}(\mathbf{V}_j)$

foreach $\mathbf{T}_k \in \mathcal{T}_{\mathbf{V}_j}$ **do**

3 Get triangle vertex index $\{\alpha, \beta, \gamma\} = \text{Pts_id}(\mathbf{T}_k)$

4 **if** $(\mathbf{V}_\alpha \in \mathcal{V}_i)$ **and** $(\mathbf{V}_\beta \in \mathcal{V}_i)$ **and** $(\mathbf{V}_\gamma \in \mathcal{V}_i)$ **then**

5 $\mathcal{T}_i^{\text{pull}} = \mathcal{T}_i^{\text{pull}} \cup \mathbf{T}_k$

Return: The pulled triangle facets $\mathcal{T}_i^{\text{pull}}$

2) *Commit:* In this step, we find out the incremental modifications of the reconstructed triangle facets \mathcal{T}_i (in Section VI-D2) w.r.t. the pulled facets $\mathcal{T}_i^{\text{pull}}$ (from Algorithm 2). These incremental modifications are summarized into an array of mesh facets to be added $\mathcal{T}_i^{\text{Add}}$ and an array of mesh facets to be erased $\mathcal{T}_i^{\text{Erase}}$. The detailed processes of this commit step are shown in Algorithm 3.

Algorithm 3: Voxel-wise mesh commit.

Input : The pulled triangle facets $\mathcal{T}_i^{\text{Pull}}$ from Algorithm 2
The reconstructed triangle facets \mathcal{T}_i

Output: The triangle facets to be added $\mathcal{T}_i^{\text{Add}}$.
The triangle facets to be erased $\mathcal{T}_i^{\text{Erase}}$.

Start : $\mathcal{T}_i^{\text{Add}} = \{\text{null}\}$, $\mathcal{T}_i^{\text{Erase}} = \{\text{null}\}$

- 1 **foreach** $\mathbf{T}_j \in \mathcal{T}_i$ **do**
- 2 **if** $\mathbf{T}_j \notin \mathcal{T}_i^{\text{Pull}}$ **then**
- 3 $\mathcal{T}_i^{\text{Add}} = \mathcal{T}_i^{\text{Add}} \cup \mathbf{T}_j$
- 4 **foreach** $\mathbf{T}_j \in \mathcal{T}_i^{\text{Pull}}$ **do**
- 5 **if** $\mathbf{T}_j \notin \mathcal{T}_i$ **then**
- 6 $\mathcal{T}_i^{\text{Erase}} = \mathcal{T}_i^{\text{Erase}} \cup \mathbf{T}_j$

Return: The triangle facets to be added $\mathcal{T}_i^{\text{Add}}$ and erased $\mathcal{T}_i^{\text{Erase}}$.

Algorithm 4: Voxel-wise mesh push.

Input : The triangle facets that need to be erased $\mathcal{T}_i^{\text{Erase}}$.
The triangle facets that need to be added $\mathcal{T}_i^{\text{Add}}$.

- 1 **Function** `Add_triangle(\mathbf{T}_j)`:
- 2 Get point indices $\{\alpha, \beta, \gamma\} = \text{Id}(\mathbf{T}_j)$
- 3 Construct triangle $\mathbf{T}_j^G = \text{Tri}(\alpha, \beta, \gamma)$ in global map.
- 4 Calculate the center of \mathbf{T}_j^G :
- 5 $\text{Center}(\mathbf{T}_j^G) = (\mathbf{V}_\alpha + \mathbf{V}_\beta + \mathbf{V}_\gamma) / 3$
- 6 Find the L1-Voxel \mathbf{V}^1 that $\text{Center}(\mathbf{T}_j^G)$ located in:
 $\mathbf{V}^1 = \text{Get_L1_voxel}(\text{Hash}(\text{Center}(\mathbf{T}_j^G)))$
- 7 Set the status flag of \mathbf{V}^1 to *Sync-required* (Section IV-D2).
- 8 Add `Ptr(\mathbf{T}_j^G)` to triangle list of L1-Voxel \mathbf{V}^1 .
- 9 Add `Ptr(\mathbf{T}_j^G)` to triangle list of points $\mathbf{V}_\alpha, \mathbf{V}_\beta, \mathbf{V}_\gamma$.
- 10 **Function** `Erase_triangle(\mathbf{T}_j)`:
- 11 Get point indices $\{\alpha, \beta, \gamma\} = \text{Id}(\mathbf{T}_j)$
- 12 Remove `Ptr(\mathbf{T}_j^G)` in triangle list of points $\mathbf{V}_\alpha, \mathbf{V}_\beta, \mathbf{V}_\gamma$.
- 13 Find the L1-Voxel \mathbf{V}^1 with $\text{Center}(\mathbf{T}_j^G)$ via (9):
 $\mathbf{V}^1 = \text{Get_L1_voxel}(\text{Hash}(\text{Center}(\mathbf{T}_j^G)))$
- 14 Set the status flag of \mathbf{V}^1 to *Sync-required* (Section IV-D2).
- 15 Remove `Ptr(\mathbf{T}_j^G)` from triangle list of L1-Voxel \mathbf{V}^1 .
- 16 Delete triangle \mathbf{T}_j^G from memory.
- 17 **foreach** $\mathbf{T}_j \in \mathcal{T}_i^{\text{Add}}$ **do**
- 18 `Add_triangle(\mathbf{T}_j)`
- 19 **foreach** $\mathbf{T}_j \in \mathcal{T}_i^{\text{Erase}}$ **do**
- 20 `Erase_triangle(\mathbf{T}_j)`

3) *Push*: With the incremental modification $\mathcal{T}_i^{\text{Erase}}$ and $\mathcal{T}_i^{\text{Add}}$ from the previous *commit* step, we perform the erasion and addition operations of global triangle mesh facets respectively. The detailed processes of our push step is shown in Algorithm 4.

F. Parallelism

To further improve the real-time performance, we implement our algorithms with parallelism for better utilization of the computation power of a multi-core CPU. In ImMesh, we have two major parallelisms as follows:

The first parallelism is implemented between the *localization* module and the *meshing* module. Except for the point cloud registration in *localization* module, which needs to operate the mesh vertices as the meshing operation, the remaining processes of *localization* module are parallelized

with the *meshing* module. More specifically, once our meshing processes start, the *localization* module is allowed to process the new coming LiDAR scan for estimation of the pose of LiDAR. However, the point cloud registration step is only allowed to be executed after the end of the meshing process.

The second parallelism is implemented among the voxel-wise meshing operation of each *activated* voxel. The voxel-wise meshing operations of different voxels are standalone thus, no conflicted operations exist on the same set of data.

G. The full meshing algorithm

To sum up, our full meshing processes are shown in Algorithm 5.

Algorithm 5: The full meshing process of each update of LiDAR scan

Input : The set of L2-Voxels $\mathcal{V}^2 = \{\mathbf{O}_1^2, \mathbf{O}_2^2, \dots, \mathbf{O}_m^2\}$ that *activated* in Section V-C

Start : The triangle facets that need to be added $\mathcal{T}_i^{\text{Add}} = \{\text{null}\}$, and to be erased in this update $\mathcal{T}_i^{\text{Erase}} = \{\text{null}\}$.

- 1 **foreach** $\mathbf{O}_i^2 \in \mathcal{V}^2$ **do in parallel**
- 2 Retrieve vertices \mathcal{V}_i with Algorithm 1.
- 3 Reconstruct the triangle facets \mathcal{T}_i with \mathcal{V}_i (Section VI-D2),
- 4 Performing voxel-wise mesh pull (Algorithm 2) to get $\mathcal{T}_i^{\text{Pull}}$. ▷ // Mesh pull
- 5 Performing voxel-wise mesh commit (Algorithm 3) to get the triangle facets that need to be added $\mathcal{T}_i^{\text{Add}}$ and erased $\mathcal{T}_i^{\text{Erase}}$. ▷ // Mesh commit
- 6 $\mathcal{T}_i^{\text{Add}} = \mathcal{T}_i^{\text{Add}} \cup \mathcal{T}_i^{\text{Pull}}$, $\mathcal{T}_i^{\text{Erase}} = \mathcal{T}_i^{\text{Erase}} \cup \mathcal{T}_i^{\text{Erase}}$

/* === Mesh push start === */

- 7 **foreach** $\mathbf{T}_j \in \mathcal{T}_i^{\text{Add}}$ **do**
- 8 `Add_triangle(\mathbf{T}_j)` ▷ // In Algorithm 4
- 9 **foreach** $\mathbf{T}_j \in \mathcal{T}_i^{\text{Erase}}$ **do**
- 10 `Erase_triangle(\mathbf{T}_j)` ▷ // In Algorithm 4

/* === Mesh push end === */

- 11 **foreach** $\mathbf{O}_i^2 \in \mathcal{V}^2$ **do**
- 12 Reset status of \mathbf{O}_i^2 as *deactivated*.

/* Remark 1: Line 1~6 are done in parallel for better real-time performance (as mentioned in Section VI-F). */

/* Remark 2: The mesh push step Line 7~10 is different with the voxel-wise operations in Algorithm 4. The $\mathcal{T}_i^{\text{Add}}$ and $\mathcal{T}_i^{\text{Erase}}$ are processed after the parallelism to avoid possible conflicts when operating the same data (i.e., triangle facets in our *mapping* module) (Line 1~6). */

VII. BROADCASTER

In ImMesh, the *broadcaster* module publishes our state estimating results (i.e., odometry) and mapping results (i.e., new registered point cloud and triangle mesh) to other applications. In addition, if the depth image is required, the *broadcaster* module will also rasterize the triangle meshes into a customized depth image (e.g., user-defined resolution and FoV).

A. Broadcast of odometry

The real-time 6-dof sensor pose from *localization* module (Section V-B) is published with the LiDAR frame starting

timestamp at a frequency of the LiDAR sampling rate. Besides, if the IMU source is available, the *broadcaster* module publishes the odometry propagated from the IMU preintegration [78] at the frequency of the IMU sampling rate.

B. Broadcast of triangle facets

Since the triangle facets are stored in an unstructured hash table of $L1$ -Voxels in *map structure*, they can not be directly applied for broadcast. To resolve this problem, our *broadcaster* module maintains a background thread that asynchronously copies the triangle facets from the hash table of each *sync-required* $L1$ -Voxels (set as *sync-required* in Algorithm 4) to a structured array for broadcasting. Then, these *sync-required* voxels are marked as *synced* after the copying. Finally, The *broadcaster* module publishes the refreshed triangle facets to other applications.

C. Rasterization of depth image

Some robotic applications, such as autonomous navigation [79] and exploration [80] tasks, require dense accurate depth images for obstacle avoidance. To meet the requirements of these scenarios, the broadcaster module utilizes the triangle facets from Section VII-B to rasterize a depth image at any customized resolution and FoV, based on the fast implementation of *OpenGL* [67].

1) *Reinforcement of LiDAR point cloud*: With the depth image from rasterization, LiDAR point cloud reinforcement is enabled by unprojecting the 3D points from the depth image. In detail, with the projection matrix and estimated pose used for rasterizing the depth image, the 3D points are obtained (i.e., unproject) w.r.t. each depth value on the depth image. As a result, the 3D point cloud is enhanced with higher resolution and larger FoV than the raw LiDAR measurement scan (see our Application-1 in Section VIII-D).

VIII. EXPERIMENTS AND RESULTS

In this section, we extensively evaluate the performance of ImMesh. Notice that our *localization* module is built upon our previous work VoxelMap [16] with no modification that relative to the state estimation. Hence, the localization precision of this work performs as well as [16]. We recommend our readers get more details about our localization accuracy by referring to the results reported in our previous work.

In this paper, we lead the experiments by evaluating our meshing ability, especially on the runtime performance and accuracy in reconstructing the triangle mesh.

A. Experiment-1: ImMesh for immediate mesh reconstruction

In this experiment, we verify the overall performance of ImMesh toward real-time simultaneous localization and meshing with live video demonstrations. As shown in Fig. 6(b), we record the full process of our data sampling at the campus of the University of Hong Kong (HKU), deploying the ImMesh for simultaneously estimating the sensor pose and reconstructing the triangle mesh on the fly. The full video demonstration of this experiment is available on YouTube: youtu.be/pzT2fMwz428?t=9.



Fig. 6: (a) shows our handheld device for data collection and online mesh reconstruction. (b) shows a snapshot of our accompanying video (on YouTube: youtu.be/pzT2fMwz428?t=9) of Experiment-1, with three time-aligned views of different sources including a screen-recorded view (in red), a camera preview (in yellow), and a third-person view (in blue).

1) *Experiment setup*: Our handheld device for data collection is shown in Fig. 6(a), which includes a mini-computer (equipped with an *Intel i9-10900* CPU and 64 GB RAM), a *Livox avia* 3D LiDAR (FoV: $70.4^\circ \times 77.2^\circ$), and a preview only RGB camera. In this experiment video, three time-aligned views of different sources are presented, including: 1) a screen-recorded view that shows the estimated posed and online reconstructed triangles mesh of ImMesh. 2) a camera preview that records the video stream of the front-facing camera. 3) a third-person view that records the whole process of this experiment.

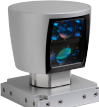



2) *Result and analysis*: As presented in the video, benefits from the accurate uncertainty models of the LiDAR point and plane that counting both LiDAR measurement noise and sensor pose estimation errors in our *localization* module, ImMesh is able to provide the 6-DoF pose estimation of very high accuracy in real-time. What is worth mentioning is, without any additional processing (i.e., loop detection), all of these two trials can close the loop itself after traveling 957 m and 391 m. In addition, with the efficient architecture design and our careful engineering implementation on our *meshing* module, the triangle mesh of the surrounding environment is incrementally reconstructed on the fly. With the live preview of real-time meshing as a reference, it is quite useful to let users know whether the data sampling is sufficient enough for any part of the scene, especially for those non-expert users. At the end of the data sampling, the dense accurate triangle mesh of this scene is already reconstructed. This is why we name our system the **Immediately Meshing** (ImMesh) framework.

B. Experiment-2: Extensive evaluation of ImMesh on public datasets with various types of LiDAR in different scenes

With all the modules delicately designed for efficiency and careful engineering implementations, both the *localization* and *meshing* modules easily achieve real-time performances on a standard multi-core CPU. In this experiment, we statics the average time consumption on four public datasets with the computation platform listed in Section VIII-A1.

The four datasets we chose are: the Kitti dataset [81], the NCTL dataset [82], the NTU VIRAL dataset [83] and the R³LIVE dataset [22]. They are collected in different scenarios ranging from urban structured buildings to field-cluttered complex environments (see Table II), using various types of LiDARs that include mechanical spinning LiDAR

TABLE I: The specifications of LiDARs in four datasets

Dataset	Kitti	NCLT	NTU VIRAL	R ³ LIVE
LiDAR				
	Velodyne HDL-64E	Velodyne HDL-32E	Ouster OS1-16 Gen1	Livox Avia
Scanning mechanism	Mechanical, spinning 64-line	Mechanical, spinning 32-line	Mechanical, spinning 16-line	Solid-state, Risleys's prism
Field of View (Horizontal ^o × Vertical ^o)	360.0 ^o × 26.8 ^o	360.0 ^o × 41.3 ^o	360.0 ^o × 33.2 ^o	70.4 ^o × 77.2 ^o
Points per second ^[1]	1,333,312	695,000	327,680	240,000
Price	\$ 75,000	\$ 8,800	\$ 3,500	\$ 1,599

¹ Only show the point rate of single-return mode.

TABLE II: This table shows the detailed information (e.g., length, duration, scenarios) of each testing sequence, the time consumption of ImMesh in processing a LiDAR scan, and the number of vertices and facets of each reconstructed mesh in Experiment-2.

Sequece	Traveling length (m)	Durations (s)	LiDAR frames	Meshing mean/Std (ms)	Localization mean/Std (ms)	Number of vertices (k)	Number of facets(k)	Scenarios
Kitti_00	3,724.2	456	4,541	32.1 / 12.0	49.0 / 11.7	3,339.4	7,692.7	Urban city
Kitti_01	2,453.2	146	1,101	34.5 / 10.5	51.1 / 18.5	2,033.0	4,046.8	High way
Kitti_02	5,058.9	509	4,661	33.5 / 7.0	36.2 / 9.5	4,390.3	10,028.1	Residential
Kitti_03	560.9	88	801	28 / 7.1	49.0 / 12.2	730.0	1,550.8	Countryside; Road
Kitti_04	393.6	27	271	30.1 / 9.4	42.4 / 12.9	411.7	850.6	Urban city; Road
Kitti_05	2,205.6	303	2,761	29.6 / 8.2	38.7 / 11.5	2,167.4	4,950.3	Residential
Kitti_06	1,232.9	123	1,101	23.1 / 5.6	56.9 / 9.7	886.1	1,889.4	Urban city
Kitti_07	2,453.2	114	1,101	20.7 / 7.4	31.3 / 8.6	764.4	1,710.5	Urban city
Kitti_08	3,222.8	441	4,071	32.4 / 7.8	45.7 / 17.7	3,559.1	7,936.3	Urban city
Kitti_09	1,705.1	171	1,591	34.5 / 7.5	43.1 / 19.2	1,827.4	4,127.5	Countryside; Road
Kitti_10	919.5	132	1,201	23.4 / 6.9	30.9 / 11.9	939.6	2,096.5	Residential
NCLT 2012-01-15	7,499.8	6739	66,889	26.3 / 14.1	21.3 / 9.8	9,659.7	26,608.3	Campus; Indoor
NCLT 2012-04-29	3,183.1	2598	25,819	25.4 / 13.9	19.1 / 5.4	4,820.9	13,483.9	Campus
NCLT 2012-06-15	4,085.9	3310	32,954	24.5 / 14.4	22.3 / 7.7	6,361.0	17,473.5	Campus
NCLT 2013-01-10	1,132.3	1024	10,212	20.2 / 12.5	19.3 / 6.5	2,020.6	5,495.8	Campus
NCLT 2013-04-05	4,523.6	4167	41,651	20.6 / 13.8	26.8 / 11.7	9,582.3	23,982.4	Campus
NTU VIRAL eee_01	265.3	398	3,987	11.2 / 6.7	14.5 / 3.4	597.6	1,380.3	Aerial; Outdoor
NTU VIRAL nya_01	200.6	396	3,949	9.4 / 5.3	10.2 / 1.7	536.8	1,247.6	Aerial; Indoor
NTU VIRAL rtp_01	449.6	482	4,615	12.1 / 8.5	10.9 / 2.6	719.2	2,030.5	Aerial; Outdoor
NTU VIRAL sbs_01	222.1	354	3,542	11.4 / 8.0	17.2 / 3.2	472.5	1,150.4	Aerial; Outdoor
NTU VIRAL tnp_01	319.4	583	5,795	6.3 / 3.7	8.8 / 1.2	155.5	414.0	Aerial; Indoor
R ³ LIVE hku_campus_00	190.6	202	2,022	12.0 / 7.3	11.5 / 3.2	587.1	1,236.9	Campus
R ³ LIVE hku_campus_01	374.6	304	3,043	20.4 / 12.6	17.2 / 6.9	1,323.4	2,862.9	Campus
R ³ LIVE hku_campus_02	354.3	323	3,236	13.5 / 6.4	11.9 / 2.8	867.9	1,913.6	Campus
R ³ LIVE hku_campus_03	181.2	173	1,737	12.2 / 5.7	11.3 / 2.9	550.0	1,130.6	Campus
R ³ LIVE hku_main_building	1,036.9	1170	11,703	16.9 / 14.3	12.5 / 8.0	3,031.2	6,803.6	Indoor; Outdoor
R ³ LIVE hku_park_00	247.3	228	2,285	30.1 / 15.9	12.6 / 3.7	919.5	2,380.2	Cluttered field
R ³ LIVE hku_park_01	401.8	351	3,520	31.5 / 12.2	12.6 / 3.9	1,673.0	3,964.8	Cluttered field
R ³ LIVE hkust_campus_00	1,317.2	1073	10,732	26.0 / 12.8	18.0 / 7.6	4,916.7	11,246.8	Campus
R ³ LIVE hkust_campus_01	1,524.3	1162	11,629	27.1 / 13.9	16.8 / 6.7	5,353.1	12,638.1	Campus
R ³ LIVE hkust_campus_02	2,112.2	1618	4,787	26.7 / 14.5	20.3 / 6.1	1,991.6	4,653.5	Campus
R ³ LIVE hkust_campus_03	503.8	478	16,181	33.6 / 13.3	21.0 / 5.3	7,673.8	18,247.3	Campus

TABLE III: Two ImMesh configurations for two types of LiDARs (i.e., mechanical and solid-state LiDAR).

	Minimum point distance (m)	L1-voxel O ¹ size (m)	L2-voxel O ² size (m)
Mechanical LiDAR	0.15	15.0	0.60
Solid-state LiDAR	0.10	10.0	0.40

TABLE IV: The average/maximum time of *meshing* and *localization* module for processing each LiDAR scan in four datasets.

	Kitti mean/max	NCLT mean/max	NTU VIRAL mean/max	R ³ LIVE mean/max
Meshing (ms)	31.3 / 34.5	24.2 / 25.4	9.8 / 17.2	25.3 / 33.6
Localization (ms)	42.2 / 56.9	22.3 / 26.8	11.9 / 17.2	16.6 / 21.0

of different channels and solid-state LiDAR of small FoV (see the specifications in Table I). Hence, the adaptability of ImMesh is sufficiently validated by extensive tests on these four distinguished datasets.

1) *Experiment setup*: Thanks to the parameter insensitivity of ImMesh, we are able to benchmark ImMesh in four datasets with only two sets of configurations. The two configurations are reasonably required for adapting two classes of LiDARs (i.e., mechanical and solid-state LiDAR), as shown in Table III. Since the 3D points sampled by a solid-state LiDAR are distributed in a small sensor FoV, the accumulated point cloud of solid-state LiDAR usually has a higher density. Therefore, we set the minimum point distance and voxel size for solid-

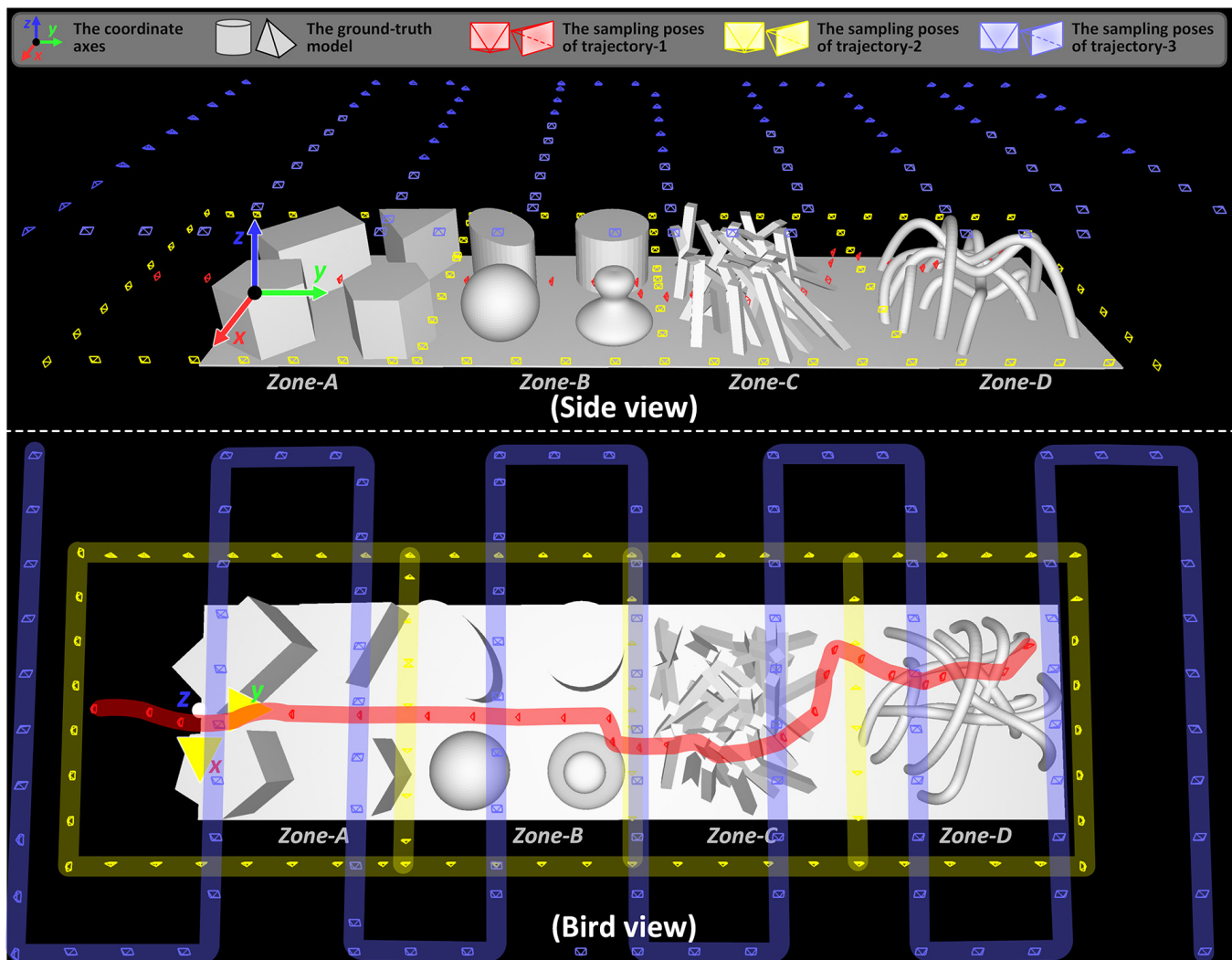


Fig. 7: In Experiment-3, we use CAD software to design a solid model to generate a ground truth triangle mesh as a reference, which contains four zones simulating different scenarios, as the white entity shown in this figure. To simulate the data collecting process with different vehicles, we generate the LiDAR point cloud data by traveling along three distinguished trajectories, whose sampling poses are colored in different colors (i.e., red, yellow, and blue).

state LiDAR 1.5 times smaller than those for mechanical LiDAR, as shown in Table III. For the other setups, we maintained the same configuration except for some necessary adjustments to match the hardware setup.

2) *Result and analysis*: Table II shows the detailed information (e.g., length, duration, scene) of each sequence, the average time consumption of our *localization* and *meshing* module in processing a LiDAR scan, and the number vertices and facets of each reconstructed mesh. From Table II, it is seen that the average cost-time of both *localization* and *meshing* modules are closely related to the density of the input LiDAR scan. To be detailed, the LiDAR of a higher channel has a much higher point sampling rate (see Table I) which causes more data to be processed in each update of a LiDAR frame (e.g., more points in a voxel and more voxels activated in each frame). Besides, for the same set of datasets, the processing time also varies among different scenarios. The sequences sampled in a high-way or field environment (e.g., Kitti_01, Kitti_09) usually have a longer LiDAR sampling range and hence leading to more points per frame to be processed. Thanks to the efficient data structure (e.g., ikd-Tree,

hashed hierarchical voxel) and parallelism strategy, which allows us to perform the state estimation and incremental mesh reconstruction simultaneously, the time consumption of large-scale datasets is bounded in an acceptable value (≤ 35 ms for meshing, ≤ 49 ms for localization).

The average and maximum time consumption of ImMesh in four datasets are shown in Table IV, reflecting that our system satisfies the real-time requirement even with different types of LiDAR and in various scenarios. Notice that the LiDAR sample rate are 10 Hz for all datasets, and our *meshing* and *localization* are run in parallel (see Section VI-F).

C. Experiment-3: Quantitative evaluation of meshing accuracy

In this experiment, we horizontally evaluate the runtime performance and meshing accuracy of ImMesh by comparing it with existing state-of-the-art mesh reconstruction methods.

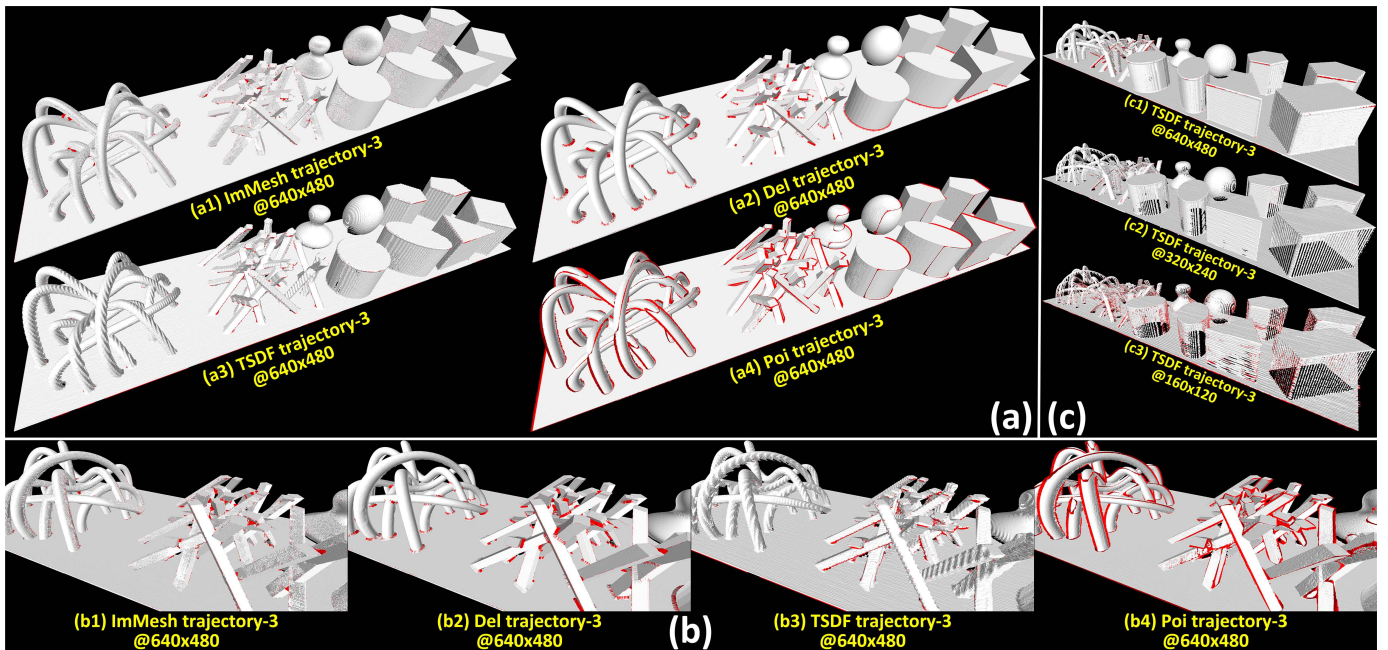


Fig. 8: This figure shows the qualitative results of Experiment-3, with all “positive” facets (correctly reconstructed) colored in white and “negative” facets (wrongly reconstructed) colored in red. (a) and (b) present a set of qualitative results of four candidates under Trajectory-3@640 × 480. (c) shows the reconstructed mesh of *TSDF* feeding with depth images of different resolutions.

TABLE V: The average time consumption of candidates in reconstructing the triangle mesh in Experiment-3.

Method	Time consumption (Unit: second(s))								
	Trajectory-1 @640x480	Trajectory-1 @320x240	Trajectory-1 @160x120	Trajectory-2 @640x480	Trajectory-2 @320x240	Trajectory-2 @160x120	Trajectory-3 @640x480	Trajectory-3 @320x240	Trajectory-3 @160x120
ImMesh (ours)	6.877	6.451	5.522	15.649	14.066	13.206	24.536	18.617	15.055
Del	371.632	132.181	30.366	696.641	353.304	56.765	960.613	323.224	85.008
TSDF	6.064	5.522	5.513	16.191	16.146	16.028	20.544	20.391	20.309
Poi	141.848	78.605	29.610	635.079	198.028	45.280	957.743	310.080	137.976

TABLE VI: The meshing accuracy of four candidates evaluated with *Criteria-1* in Experiment-3.

Method	Criteria-1: Meshing precision			in Zone-A / in Zone-B in Zone-C / in Zone-D in all zones (average)			(Unit: percentage(%))		
	Trajectory-1 @640x480	Trajectory-1 @320x240	Trajectory-1 @160x120	Trajectory-2 @640x480	Trajectory-2 @320x240	Trajectory-2 @160x120	Trajectory-3 @640x480	Trajectory-3 @320x240	Trajectory-3 @160x120
ImMesh (ours)	99.96 / 99.43	99.72 / 97.93	98.65 / 93.82	99.60 / 99.48	98.91 / 98.76	95.98 / 96.27	98.97 / 98.97	96.30 / 96.33	85.21 / 82.84
	98.06 / 98.98	96.06 / 97.15	91.47 / 92.50	98.38 / 99.49	96.51 / 98.09	90.58 / 94.07	98.00 / 99.05	95.24 / 97.53	84.88 / 89.59
Del	97.48	97.48	93.47	99.20	98.09	94.29	98.72	96.31	85.53
	97.62 / 97.38	98.38 / 98.98	97.04 / 97.56	98.49 / 99.27	98.24 / 99.24	97.49 / 98.38	96.28 / 97.27	94.20 / 94.77	92.41 / 91.61
TSDF	92.95 / 97.54	96.20 / 98.86	95.09 / 97.68	94.39 / 99.09	95.31 / 98.96	93.94 / 98.14	94.30 / 98.14	92.75 / 96.31	93.60 / 96.59
	96.39	98.15	96.90	97.83	98.03	97.16	96.47	94.49	93.47
Poi	98.93 / 99.88	99.83 / 97.57	94.18 / 90.42	97.96 / 99.53	96.30 / 96.48	92.60 / 84.98	99.54 / 99.62	99.12 / 99.11	85.88 / 85.45
	93.38 / 99.10	92.78 / 96.30	74.31 / 80.64	90.53 / 98.13	86.02 / 96.24	75.82 / 87.48	92.74 / 99.43	96.60 / 99.21	83.08 / 89.48
TSDF	97.56	96.35	83.55	96.21	93.36	84.97	97.67	98.43	85.76
	96.95 / 97.07	97.04 / 97.30	97.27 / 96.92	96.13 / 97.00	96.31 / 97.10	95.77 / 96.90	95.86 / 96.94	95.78 / 96.23	89.13 / 84.11
Poi	91.14 / 92.11	91.02 / 92.30	88.43 / 91.87	92.27 / 92.00	91.70 / 91.26	91.83 / 92.25	91.72 / 92.44	91.91 / 92.23	86.43 / 89.06
	93.62	93.67	92.38	94.10	94.25	94.35	94.25	93.98	87.06

1) *Prepare of simulated data*: Since the ground truth triangle mesh of the real-world data can not be directly obtained, we use CAD software *SolidWorks* [14] to design a ground truth solid model for reference, as shown in Fig. 7. This solid model we made is constituted of four distinguished zones for an extensive evaluation of the meshing results in different scenes, which include the simple planar zone (Zone-A), simple curvy (bending) zone (Zone-B), complex planar zone (Zone-C), and complex curvy Zone (Zone-D). Each zone has an equal size of length×width×height as 10.0 m×10.0 m×6.5 m.

To simulate point clouds collected by a real LiDAR, we

built a simulator to unproject the “LiDAR” points from the depth images generated from the rasterization of the ground truth models with given poses. In this experiment, we rasterized the depth image with a pinhole projection model of horizontal×vertical FoV as 80° × 60°. Besides, to simulate the LiDAR of different point cloud densities, we rasterized the depth image with three sets of resolutions (see Table V) including 640 × 480, 320 × 240, and 160 × 120. Finally, we designed three distinguished sampling trajectories as shown in Fig. 7. Each trajectory contained a number of manually placed poses for simulating different vehicles in collecting the data.

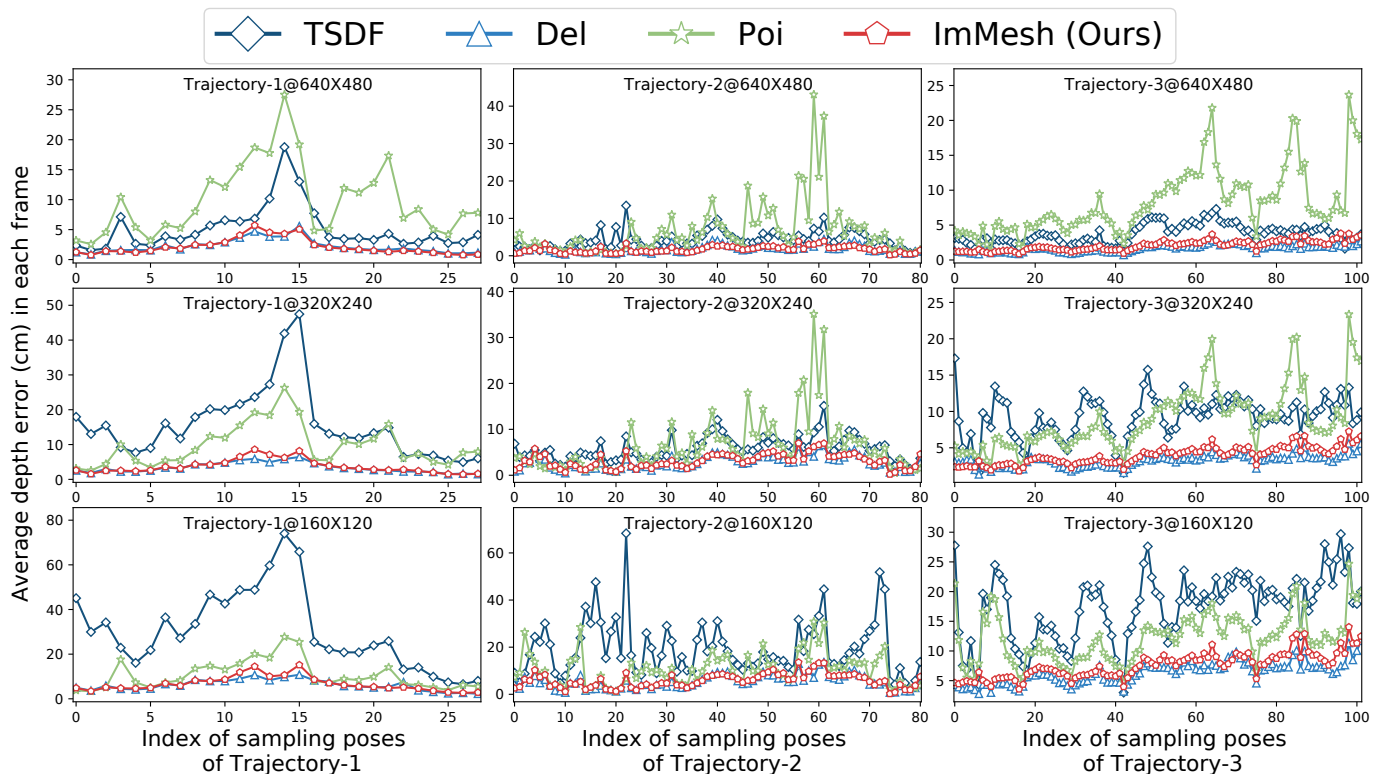


Fig. 10: Re-rendering depth error in each frame of four candidates in Experiment-3.

TABLE VII: The meshing accuracy of four candidates evaluated by *Criteria-2* in Experiment-3.

Criteria-2: Re-rendering depth error (Unit: centimeter (cm))									
Method	Trajectory-1 @640x480	Trajectory-1 @320x240	Trajectory-1 @160x120	Trajectory-2 @640x480	Trajectory-2 @320x240	Trajectory-2 @160x120	Trajectory-3 @640x480	Trajectory-3 @320x240	Trajectory-3 @160x120
ImMesh (ours)	2.146	3.678	6.67	1.686	3.115	5.838	2.076	3.943	7.592
Del	2.216	3.37	6.176	1.832	2.807	5.358	1.674	3.205	6.327
TSDF	5.068	15.643	30.421	4.231	5.652	19.352	3.724	9.167	17.288
Poi	9.844	9.611	10.594	6.546	6.377	10.466	8.21	8.848	12.142

The details of these three trajectories are shown below:

- The trajectory-1 (in red) contains 28 sampling poses, simulating the LiDAR mounted on a car with a height of 1.5 m away from the ground (i.e., $z = 0$ plane). The LiDAR data is collected by moving from Zone-A to Zone-D.

- The trajectory-2 (in yellow) contains 81 sampling poses. It simulates a handheld LiDAR collecting data at the height fixed as 1.5 m. The LiDAR data is collected by traveling in an “∞”-like pattern which sufficiently captures the model’s surface from different views.

- The trajectory-3 (in blue) contains 102 sampling poses, imitating a LiDAR mounted on a drone flying at the height of 8.5 m. The LiDAR data is collected from a tilted bird view by flying in an “S”-like pattern.

Due to the limitation of height in sampling the data, LiDAR in trajectory-1 and trajectory-2 did not capture the ceiling surface of the model. Conversely, LiDAR in trajectory-3 captured the ceiling surfaces but failed to capture the bottom surfaces of the models. Besides, LiDAR in trajectory-1 traveled in one direction; hence only the surfaces facing against the positive Y -axis were captured.

2) *Experiment setup*: In this experiment, we conducted a fair evaluation of meshing ability among our work and existing mesh reconstruction baselines, which includes a TSDF-based

method implemented by *Point cloud library (PCL)* [53] with GPU acceleration, Delaunay triangulation and graph cut based method implemented by *OpenMVS* [84], and the official implementation of Poisson surface reconstruction [23, 24].

We conducted the evaluation of candidates on a desktop PC that equips with an *Intel i7-9700K* CPU, 64Gb RAM, and a *Nvidia 2080 Ti* GPU with 12Gb graphics memory. We feed our ImMesh and TSDF-based (*TSDF*) method with LiDAR points frame by frame. To avoid the pose estimation error that affects the result of meshing, we disable the pose estimation module and feed ImMesh and TSDF with the ground truth poses. For offline mesh reconstruction methods: Delaunay triangulation and graph cut (*Del*) based method and Poisson surface reconstruction (*Poi*), we feed them with the accumulated point cloud of all frames. To avoid the uneven point cloud density which leads to errors in calculating the norm for *Poi*, and to avoid *Del* reconstructing the tiny facets that lead to a biased calculation of accuracy, we leverage a voxel grid filter with a leaf size of $1.0 \text{ cm} \times 1.0 \text{ cm} \times 1.0 \text{ cm}$ to downsample the accumulated point cloud before feeding to *Poi* and *Del*.

Due to and limitation of graphics memory (12Gb for *Nvidia 2080 Ti*), we set the *TSDF* cell size as 0.2 m such that TSDF can utilize the GPU acceleration while preserving satisfying

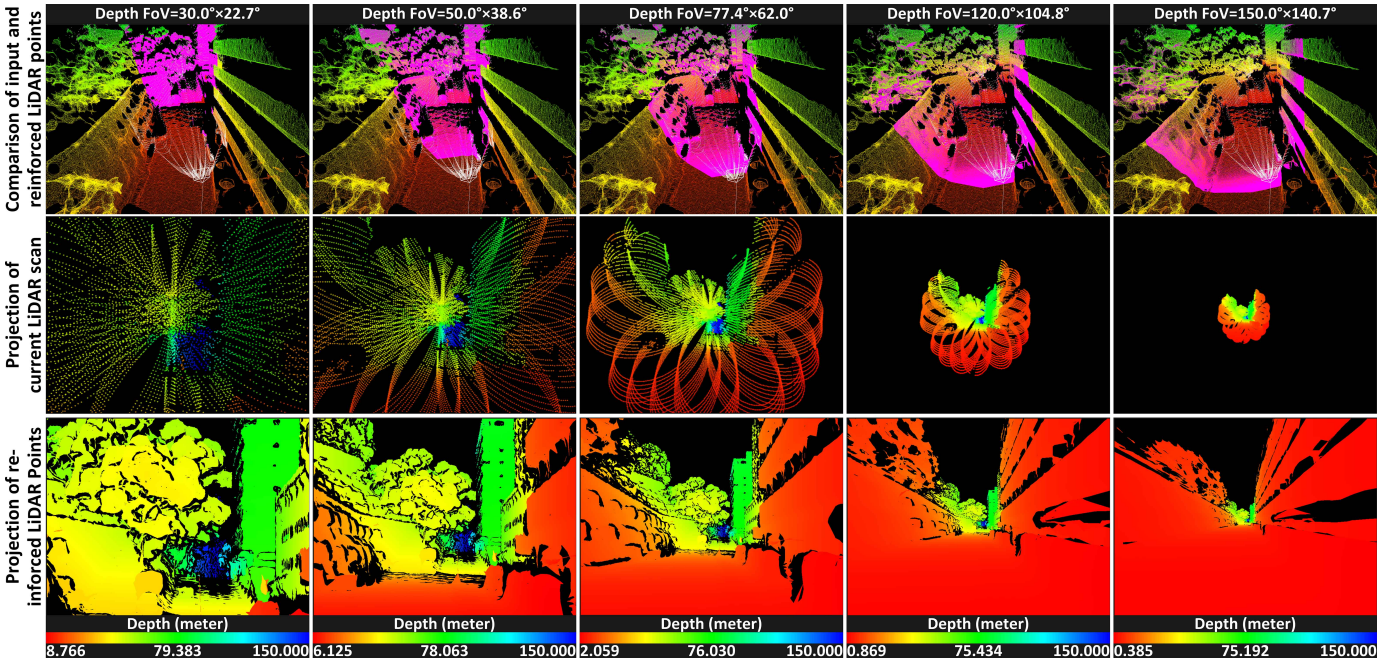


Fig. 11: The first row of images shows the comparisons between a raw LiDAR scan (colored in white) and our reinforced points (colored in cyan) under different sets of rasterizing FoV. The second and third rows of images show the comparisons of raw and reinforced points after projection on the current sensor frame. For more detailed visualizations of this process, please refer to our accompanying video on YouTube: youtu.be/pzT2fMwz428?t=499.

precision in the mesh reconstruction. For our ImMesh, the parameter configuration for solid-state LiDAR is used, as shown in Table III. For *Poi*, we set the octree level as 12 and removed large hulls by deleting facets with one of their edges longer than 15.0 cm. For other configurations of all candidates, we set them as their default configuration.

In this experiment, we horizontally evaluated the meshing accuracy of candidates by comparing their reconstructed mesh with the ground truth models. A set of qualitative results of four candidates under Trajectory-3@640 × 480 are shown in Fig. 8(a and b). To sufficiently and fairly calculate the accuracy by comparing the mesh of the candidate’s and ground truth, two criteria are adopted for counting the difference, shown below:

- *Criteria-1*: For a triangle facet T_i^{can} of a candidate’s reconstructed mesh, we first find out a triangle facet T_j^{gt} of the ground truth model, whose point-to-plane distance from this facet to the center of T_i^{can} is minimum. T_i^{can} is regarded as “positive” if it satisfies both of the following conditions: 1) The point-to-plane distance between T_i^{gt} and $Center(T_i^{can})$ smaller than 5.0 cm; 2) The angular distance between the norm vector of T_i^{can} and the norm vector of T_j^{gt} smaller than 15° . Otherwise, this triangle facet T_i^{can} is treated as “negative”. The ratios of “positive” over the total number of facets in each zone (and the entire simulated scene) served as *Criteria-1* for evaluating the meshing accuracy, as the results are shown in Table VI.

- *Criteria-2*: For each candidate’s reconstructed mesh, it is rasterized into a depth image in the same way as rasterizing the ground truth model to a depth image (for generating the simulation data, see Section VIII-C1). The average depth error of each pixel depth value is calculated between each depth image pair of the candidate and ground truth (i.e.,

re-rendering error), serving as *Criteria-2* for evaluating the meshing accuracy, with the results shown in Figure. 10 and Table VII.

While *Criteria-1* reflects the correctness of candidates in reconstructing the mesh and reflects different performances in different zones, it is unable to count the holes of the mesh. On the contrary, *Criteria-2* reflects the errors caused by holes but can not count the facets out of view (e.g., the facets hide behind other facets). Referring to the results calculated according to *Criteria-1* (i.e., Table VI) and *Criteria-2* (i.e., Figure. 10 and Table VII), we conducted the evaluation and analysis on the meshing accuracy of four candidates.

3) *Results and analysis of runtime performance*: The average time consumption of four candidates is listed in Table V. The online methods *ImMesh* and *TSDF* show a comparative runtime performance, while the offline methods *Del* and *Poi* consume about two orders of magnitude larger than the online methods. Notice that *TSDF* achieves the comparative runtime performance as ours with the acceleration of an *Nvidia 2080 Ti* GPU, which indicates the highest computation efficiency of our ImMesh among the four candidates.

4) *Result and analysis of meshing accuracy*: The results evaluated by *Criteria-1* are shown in Table VI. All candidates show satisfying accuracy in reconstructing the mesh of the simple planar models in Zone-A, followed by the simple curvy model in Zone-B. In complex scenes, all candidates show lower accuracy and achieve worse results in Zone-C, where many square cylinders cross each other, making it hard to reconstruct well. In addition, as the point cloud (i.e., the resolution of depth images) becomes sparser, the accuracy drops responsibly, especially for *TSDF*-based method. Lastly, *Poi* shows a bad accuracy in complex scenes due to the unwanted facets appearing at the sharp edge of the models,

as the facets colored in red shown in Fig. 8(b4 and c4).

The results evaluated by *Criteria-2* are shown in Figure. 10 and Table VII. *Del* achieves the best precision by showing the lowest depth error. Our proposed algorithm ImMesh performs closely to *Del*, followed by *Poi* and *TSDF*. As the graphs shown in each column of Figure. 10, the average depth error of the *TSDF* increases sharply as the resolution of depth images goes down, due to the appearance of the holes on the mesh (as shown in Fig. 8(c)). This unwanted phenomenon that uses *TSDF*-based methods for constructing mesh with depth image of low resolution is also reported in other work [59].

5) *Summary*: We lead the conclusions of Experiment-3 based on the results and analysis discussed in Section VIII-C3 and Section VIII-C4: For offline applications, which only care about quality and neglect time consumption, *Del* is the best choice, and our ImMesh is the second best one. *Poi* shows satisfying results in simple scenes, but it is incapable of reconstructing complex scenes with many sharp edges. For real-time applications, our work ImMesh is the best choice. Even though *TSDF* with GPU implementation can meet the runtime requirement of real-time scenarios, its performance is unsatisfying due to the low meshing accuracy compared to ImMesh.

D. Application-1: ImMesh for LiDAR point cloud reinforcement

Benefiting from ImMesh’s real-time ability to reconstruct the triangle mesh on the fly, depth images can be rasterized from the reconstructed facets online in the current sensor frame. By unprojecting the 3D points from the depth image, point clouds of a regular pattern can be retrieved with wider FoV and denser distribution compared to the original input LiDAR scan. We termed this process as LiDAR point reinforcement.

In this experiment, we demonstrate the LiDAR point cloud reinforcement with a solid-state LiDAR *Livox Avia* with FoV of $70.4^\circ \times 77.2^\circ$. The comparisons between the original points of a LiDAR frame (colored in white) and after our reinforcement (colored in cyan) with different sets of rasterization FoV are shown in Fig. 11. As the white points shown in the first row of Fig. 11, the input LiDAR scan is sparse with an irregular scanning pattern. After the reinforcement, the resultant 3D points colored in cyan are distributed in a regular pattern, with denser density and wider FoV (as the rasterization FoV is bigger than LiDAR’s). To have a better sense of their differences, we present the comparisons of depth images after projection, as shown in the second and third rows of Fig. 11.

In this manner, the LiDAR points after reinforcement can benefit the applications in these scenarios: 1) the reinforced points of denser density and wider FoV enable navigation algorithms to achieve better planning performance and make smarter decisions. 2) it provides unified point cloud outputs neglecting scanning patterns of different LiDARs. Compared to the use of original LiDAR points with specific scanning patterns, using these points of regular patterns potentially benefits learning-based algorithms for better generalization.

E. Application-2: ImMesh for rapid, lossless texture reconstruction

In this application, we show how ImMesh can be applied in applications of losslessly texture reconstruction for rapid field surveying. As shown in Fig. 12(b1~b3), we mounted a *Livox avia* LiDAR and a *Hikvision CA-050-11UC* global shutter RGB camera on a *DJI M300* drone platform.

We collected the data in a mountain field by taking off from Zone-A (see Fig. 12(a)), and flying in a “s”-like pattern trajectory with a traveling distance of 975 m. We leveraged ImMesh for reconstructing the mesh from collected LiDAR data and used *R³LIVE++* [21, 22] for estimating the camera’s poses (as the yellow frustum shown in Fig. 12(a, c1 and c2)). We textured each facet of the reconstructed mesh by the RGB image captured by the nearest camera with the estimated camera pose from *R³LIVE++*. Benefit from the high efficiency of ImMesh and *R³LIVE++*, the total time of reconstructing the RGB textured mesh from this sequence of duration 325 s cost only 686 s, with 328 s for ImMesh, with 330 s for *R³LIVE++* and 28 s for texturing. Fig. 12(a) shows a bird view of our mesh after texturing, with the close-up views of textured mesh in Zone-A, B, and C are shown in Fig. 12(e1, e2, and e3), respectively. In Fig. 12(c1 and c2), we show the altitude of this map by coloring the facets in their height w.r.t. the take-off point (i.e., the ground plane in Zone-A).

As the close-up views shown in the bottom three rows of Fig. 12, the reconstructed mesh (d1~d3) from our ImMesh after texturing (e1~e3) successfully preserves the textures of maps when comparing with the RGB colored point cloud reconstructed by *R³LIVE++* (f1~f3). Since the density of the point cloud is not infinite, *R³LIVE++* is unable to losslessly reconstruct the scene’s radiance by storing radiance information in points with limited density. On the contrary, reconstructing the maps with mesh reconstructed by ImMesh, and texturing the facets with collected images and the camera poses of *R³LIVE++*. The raw color images photoed by the camera are losslessly preserved on the facets of the mesh. Hence this is a lossless manner for reconstructing the texture of the scene. Compared to existing counterparts (e.g., structure from motion (SFM) [12, 30]), this manner shows significant advantages on: 1) It is a reliable solution that does not require GPS measurement. 2) It is a rapid reconstruction method that costs just 2~3 times of data sampling time for reconstructing a scene. 3) It is a lossless texture reconstruction method, while preserving geometry structure of very high accuracy that is constructed from LiDAR’s measurement.

The accompanying video that records the full process of this lossless texture reconstruction is available on our YouTube: youtu.be/pzT2fMwz428?t=622, and an additional trial is shown in our Supplementary Material³.

IX. CONCLUSIONS AND FUTURE WORK

A. Conclusions

In this work, we proposed a novel meshing framework termed ImMesh for achieving the goal of simultaneous local-

³https://github.com/hku-mars/ImMesh/blob/main/supply/Supplementary_material.pdf

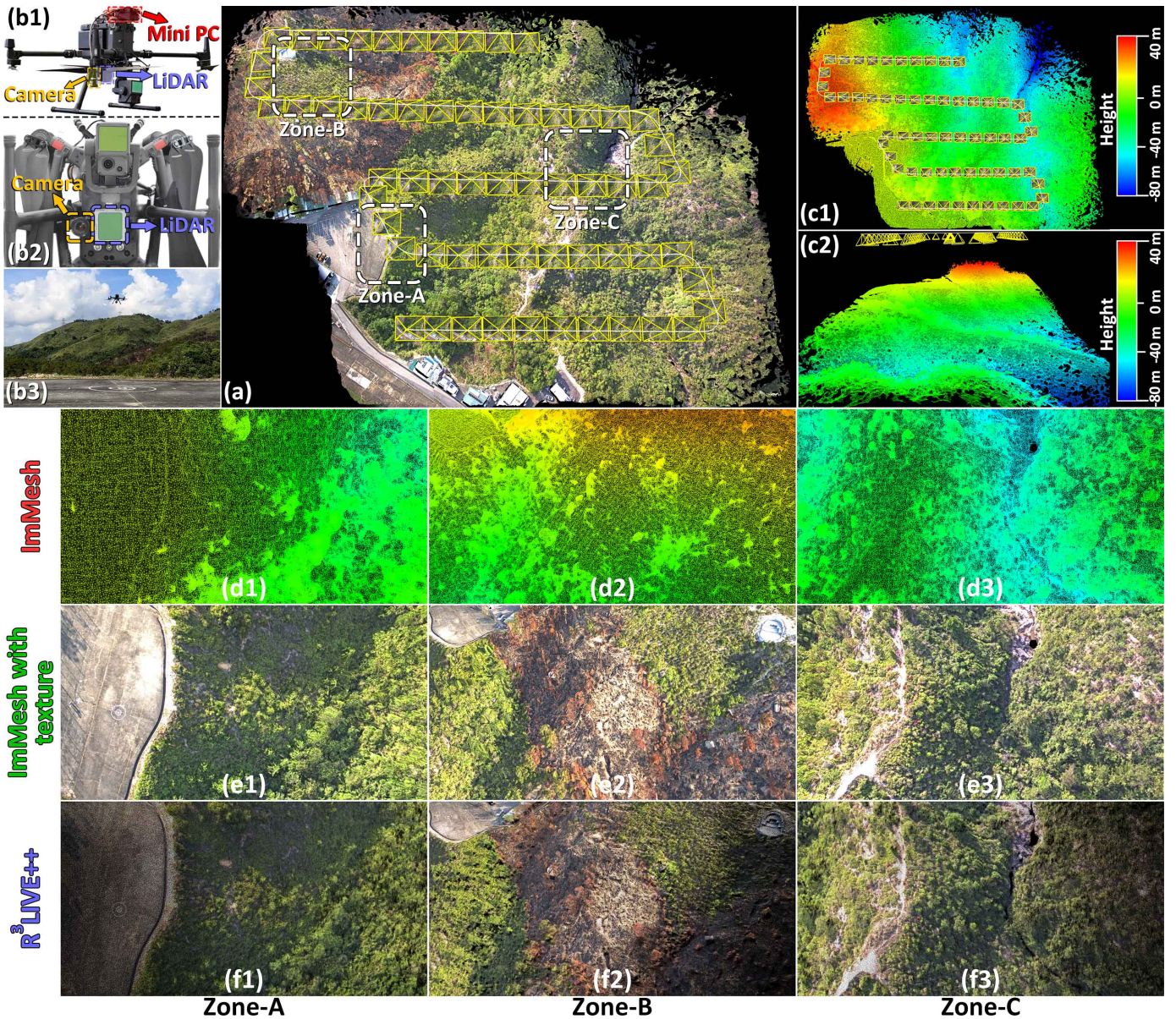


Fig. 12: (b1~b3) show our UAV platform for data collection. (a) show the bird view of our lossless texture reconstruction result. (c1 and c2) show the altitude of this map by coloring the facets in their height w.r.t. the take-off point (i.e., the ground plane in Zone-A). The qualitative comparison of mapping results in Zone-A, B, and C of ImMesh, ImMesh after textured, and R³LIVE are shown in (d~f). To see the detailed reconstruction process of the scene, please refer to our video on YouTube: youtu.be/pzT2fMwz428?t=622.

ization and meshing framework in real-time. To the best of our knowledge, it is the first work in literature to reconstruct the triangle mesh of a large-scale scene in an incremental manner in real-time. In ImMesh, the localization module represents the surrounding environment in a probabilistic representation, estimating the sensor pose in real-time by leveraging an iterated Kalman filter to maximize a posterior. The *meshing module* directly utilizes the registered LiDAR point as mesh vertices, real-time reconstructing the triangle facets in a novel incremental manner. To be detailed, our *meshing module* first utilizes an efficient hierarchical voxel data structure for fast finding of voxels containing newly appended vertices. Then, the voxel-wise 3D meshing problem is converted into a 2D one by performing dimension reduction. Finally, the triangle facets are incrementally reconstructed with *pull*, *commit*, and *push* steps.

In our experiments, we first verified the overall performance by presenting live video demonstrations of how the mesh is immediately reconstructed in the process of data collection. Then we extensively tested ImMesh with four public datasets collected by four distinguished LiDAR in various scenes, which confirmed the real-time ability in all sequences we evaluated. Lastly, we horizontally evaluated the meshing performance of ImMesh in Experiment-3 by comparing it against existing meshing baselines. The results show that ImMesh achieves high meshing accuracy while keeping the best runtime performance among all candidates.

In our applications, we first show how ImMesh can be applied for LiDAR point cloud reinforcement, which generates reinforced points in a regular pattern with denser density and wider FoV compared to raw LiDAR scan. In Application-2, we combined our works ImMesh and R³LIVE++ to achieve

the goal of losslessly texture reconstruction of scenes. Finally, to share our findings and make contributions to the community, we make our code publicly available on our GitHub: github.com/hku-mars/ImMesh.

B. Future work

In ImMesh, we propose a novel framework that can simultaneously localization and meshing in real-time. Further, to realize the goal of lossless texture reconstruction of scenes, our current solution is combining ImMesh and R³LIVE at a system level as presented in our Application-2 (in Section VIII-E), which is indeed a solution but not the elegant one. Hence, our future work would trend to make ImMesh and R³LIVE work in a more tightly combined style. Besides, since our system does not implement any loop correction yet, it drifts gradually due to accumulated localization errors. Our future work will integrate our recent works [85, 86] on loop detection based on LiDAR point cloud, which is able to online detecting the possible loop and then reduce the drift by leveraging the loop correction.

X. ACKNOWLEDGEMENTS

The authors would like to thank DJI Co., Ltd⁴ for providing devices and research found.

REFERENCES

- [1] S. Mystakidis, "Metaverse," *Encyclopedia*, vol. 2, no. 1, pp. 486–497, 2022.
- [2] Y. Wang, Z. Su, N. Zhang, R. Xing, D. Liu, T. H. Luan, and X. Shen, "A survey on metaverse: Fundamentals, security, and privacy," *IEEE Communications Surveys & Tutorials*, 2022.
- [3] P. Cipresso, I. A. C. Giglioli, M. A. Raya, and G. Riva, "The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature," *Frontiers in psychology*, p. 2086, 2018.
- [4] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "Airsim: High-fidelity visual and physical simulation for autonomous vehicles," in *Field and service robotics*. Springer, 2018, pp. 621–635.
- [5] S. Laine and T. Karras, "High-performance software rasterization on gpus," in *Proceedings of the ACM SIGGRAPH Symposium on High Performance Graphics*, 2011, pp. 79–88.
- [6] T. Akenine-Moller, E. Haines, and N. Hoffman, *Real-time rendering*. AK Peters/crc Press, 2019.
- [7] J. Arvo, *Graphics gems II*. Elsevier, 2013.
- [8] P. Jiménez, F. Thomas, and C. Torras, "3d collision detection: a survey," *Computers & Graphics*, vol. 25, no. 2, pp. 269–285, 2001.
- [9] C. Ericson, *Real-time collision detection*. Crc Press, 2004.
- [10] R. Featherstone, *Rigid body dynamics algorithms*. Springer, 2014.
- [11] D. Baraff, "An introduction to physically based modeling: rigid body simulation i—unconstrained rigid body dynamics," *SIGGRAPH course notes*, vol. 82, 1997.
- [12] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4104–4113.
- [13] F. Kong, X. Liu, B. Tang, J. Lin, Y. Ren, Y. Cai, F. Zhu, N. Chen, and F. Zhang, "Marsim: A light-weight point-realistic simulator for lidar-based uavs," *arXiv preprint arXiv:2211.10716*, 2022.
- [14] D. S. SolidWorks, "Solidworks®," *Version Solidworks*, vol. 1, 2005.
- [15] B. O. Community, "Blender—a 3d modelling and rendering package," *Blender Foundation*, 2018.
- [16] C. Yuan, W. Xu, X. Liu, X. Hong, and F. Zhang, "Efficient and probabilistic adaptive voxel mapping for accurate online lidar odometry," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8518–8525, 2022.
- [17] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "Fast-lio2: Fast direct lidar-inertial odometry," *IEEE Transactions on Robotics*, 2022.
- [18] J. Behley and C. Stachniss, "Efficient surfel-based slam using 3d laser range data in urban environments," in *Robotics: Science and Systems*, vol. 2018, 2018, p. 59.
- [19] Y. Pan, P. Xiao, Y. He, Z. Shao, and Z. Li, "Mulls: Versatile lidar slam via multi-metric linear least square," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 11 633–11 640.
- [20] T. Shan and B. Englot, "Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4758–4765.
- [21] J. Lin and F. Zhang, "R³live: A robust, real-time, rgb-colored, lidar-inertial-visual tightly-coupled state estimation and mapping package," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 10 672–10 678.
- [22] —, "R³live++: A robust, real-time, radiance reconstruction package with a tightly-coupled lidar-inertial-visual state estimator," *arXiv preprint arXiv:2209.03666*, 2022.
- [23] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Proceedings of the fourth Eurographics symposium on Geometry processing*, vol. 7, 2006.
- [24] M. Kazhdan and H. Hoppe, "Screened poisson surface reconstruction," *ACM Transactions on Graphics (ToG)*, vol. 32, no. 3, pp. 1–13, 2013.
- [25] J. Wilhelms and A. Van Gelder, "Octrees for faster isosurface generation," *ACM Transactions on Graphics (TOG)*, vol. 11, no. 3, pp. 201–227, 1992.
- [26] R. Shekhar, E. Fayyad, R. Yagel, and J. F. Cornhill, "Octree-based decimation of marching cubes surfaces," in *Proceedings of Seventh Annual IEEE Visualization'96*. IEEE, 1996, pp. 335–342.
- [27] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *ACM siggraph computer graphics*, vol. 21, no. 4, pp. 163–169, 1987.
- [28] M. Kazhdan, M. Chuang, S. Rusinkiewicz, and H. Hoppe, "Poisson surface reconstruction with envelope constraints," in *Computer graphics forum*, vol. 39, no. 5. Wiley Online Library, 2020, pp. 173–182.
- [29] P. Labatut, J.-P. Pons, and R. Keriven, "Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts," in *2007 IEEE 11th international conference on computer vision*. IEEE, 2007, pp. 1–8.
- [30] V. Litvinov and M. Lhuillier, "Incremental solid modeling from sparse and omnidirectional structure-from-motion data," in *British Machine Vision Conference*, 2013.
- [31] M. Jancosek and T. Pajdla, "Exploiting visibility information in surface reconstruction to preserve weakly supported surfaces," *International scholarly research notices*, vol. 2014, 2014.
- [32] F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, and G. Taubin, "The ball-pivoting algorithm for surface reconstruction," *IEEE transactions on visualization and computer graphics*, vol. 5, no. 4, pp. 349–359, 1999.
- [33] R. Wang, J. Peethambaran, and D. Chen, "Lidar point clouds to 3-d urban models : a review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 2, pp. 606–627, 2018.
- [34] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *2011 10th IEEE international symposium on mixed and augmented reality*. Ieee, 2011, pp. 127–136.
- [35] J. Chen, D. Bautembach, and S. Izadi, "Scalable real-time volumetric surface reconstruction," *ACM Transactions on Graphics (ToG)*, vol. 32, no. 4, pp. 1–16, 2013.
- [36] 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.
- [37] O. Kähler, V. Prisacariu, J. Valentin, and D. Murray, "Hierarchical voxel block hashing for efficient integration of depth images," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 192–197, 2015.
- [38] E. Vespa, N. Nikolov, M. Grimm, L. Nardi, P. H. J. Kelly, and S. Leutenegger, "Efficient octree-based volumetric SLAM supporting signed-distance and occupancy mapping," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1144–1151, Apr. 2018.
- [39] O. Kähler, V. A. Prisacariu, C. Y. Ren, X. Sun, P. Torr, and D. Murray, "Very high frame rate volumetric integration of depth images on mobile devices," *IEEE transactions on visualization and computer graphics*, vol. 21, no. 11, pp. 1241–1250, 2015.

⁴<https://www.dji.com>

- [40] M. Klingensmith, I. Dryanovski, S. S. Srinivasa, and J. Xiao, "Chisel: Real-time large scale 3d reconstruction onboard a mobile device using spatially hashed signed distance fields." in *Robotics: science and systems*, vol. 4, no. 1. Citeseer, 2015.
- [41] H. Oleynikova, Z. Taylor, M. Fehr, R. Siegwart, and J. Nieto, "Voxblox: Incremental 3d euclidean signed distance fields for on-board map planning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1366–1373.
- [42] D. Lefloch, M. Kluge, H. Sarbolandi, T. Weyrich, and A. Kolb, "Comprehensive use of curvature for robust and accurate online surface reconstruction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2349–2365, 2017.
- [43] D. Lefloch, T. Weyrich, and A. Kolb, "Anisotropic point-based fusion," in *2015 18th International Conference on Information Fusion (Fusion)*. IEEE, 2015, pp. 2121–2128.
- [44] T. Weise, T. Wismer, B. Leibe, and L. Van Gool, "In-hand scanning with online loop closure," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*. IEEE, 2009, pp. 1630–1637.
- [45] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy, "Real-time 3d model acquisition," *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, pp. 438–446, 2002.
- [46] M. Habbecke and L. Kobbelt, "A surface-growing approach to multi-view stereo reconstruction," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.
- [47] T. Bodenmueller, "Streaming surface reconstruction from real time 3d measurements," Ph.D. dissertation, Technische Universität München, 2009.
- [48] T. Whelan, S. Leutenegger, R. Salas-Moreno, B. Glocker, and A. Davison, "Elasticfusion: Dense slam without a pose graph." *Robotics: Science and Systems*, 2015.
- [49] T. Whelan, R. F. Salas-Moreno, B. Glocker, A. J. Davison, and S. Leutenegger, "Elasticfusion: Real-time dense slam and light source estimation," *The International Journal of Robotics Research*, vol. 35, no. 14, pp. 1697–1716, 2016.
- [50] W. Gao and R. Tedrake, "Surfelwarp: Efficient non-volumetric single view dynamic reconstruction," *arXiv preprint arXiv:1904.13073*, 2019.
- [51] T. Schöps, T. Sattler, and M. Pollefeys, "Surfel-meshing: Online surfel-based mesh reconstruction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 10, pp. 2494–2507, 2019.
- [52] Y. Cai, W. Xu, and F. Zhang, "ikd-tree: An incremental kd tree for robotic applications," *arXiv preprint arXiv:2102.10808*, 2021.
- [53] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 1–4.
- [54] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration." *VISAPP (1)*, vol. 2, no. 331-340, p. 2, 2009.
- [55] M. Teschner, B. Heidelberger, M. Müller, D. Pomerantes, and M. H. Gross, "Optimized spatial hashing for collision detection of deformable objects." in *Vmv*, vol. 3, 2003, pp. 47–54.
- [56] ISO, *ISO/IEC 14882:1998: Programming languages – C++*, Sep. 1998.
- [57] 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, no. 3, pp. 189–206, 2013.
- [58] V. Panek, Z. Kukulova, and T. Sattler, "Meshloc: Mesh-based visual localization," in *European Conference on Computer Vision*. Springer, 2022, pp. 589–609.
- [59] I. Vizzo, X. Chen, N. Chebrolu, J. Behley, and C. Stachniss, "Poisson surface reconstruction for lidar odometry and mapping," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5624–5630.
- [60] M. Dreher, H. Blum, R. Siegwart, and A. Gawel, "Global localization in meshes," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, vol. 38. IAARC Publications, 2021, pp. 747–754.
- [61] M. Oelsch, M. Karimi, and E. Steinbach, "R-loam: Improving lidar odometry and mapping with point-to-mesh features of a known 3d reference object," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2068–2075, 2021.
- [62] J. Lin, C. Zheng, W. Xu, and F. Zhang, "R²live: A robust, real-time, lidar-inertial-visual tightly-coupled state estimator and mapping," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7469–7476, 2021.
- [63] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real-time." in *Robotics: Science and Systems*, vol. 2, no. 9. Berkeley, CA, 2014, pp. 1–9.
- [64] J. Lin and F. Zhang, "Loam livox: A fast, robust, high-precision lidar odometry and mapping package for lidars of small fov," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 3126–3131.
- [65] R. Stevens, *Computer Graphics Dictionary*, ser. ADVANCES IN COMPUTER GRAPHICS AND GAME DEVELOPMENT SERIES. Charles River Media, 2002. [Online]. Available: <https://books.google.com.hk/books?id=XqJcMiPI0C>
- [66] W. Kahan, "Miscalculating area and angles of a needle-like triangle," *University of California, Berkeley*, vol. 94720, 1776.
- [67] M. Woo, J. Neider, T. Davis, and D. Shreiner, *OpenGL programming guide: the official guide to learning OpenGL*. Addison-Wesley Longman Publishing Co., Inc., 1999.
- [68] F. Evans, S. Skiena, and A. Varshney, "Optimizing triangle strips for fast rendering," in *Proceedings of Seventh Annual IEEE Visualization'96*. IEEE, 1996, pp. 319–326.
- [69] D. Hearn, M. P. Baker, and M. P. Baker, *Computer graphics with OpenGL*. Pearson Prentice Hall Upper Saddle River, NJ., 2004, vol. 3.
- [70] J. Loeliger and M. McCullough, *Version Control with Git: Powerful tools and techniques for collaborative software development*. O'Reilly Media, Inc., 2012.
- [71] K. R. Castleman, *Digital image processing*. Prentice Hall Press, 1996.
- [72] A. Fabri and S. Pion, "Cgal: The computational geometry algorithms library," in *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, 2009, pp. 538–539.
- [73] C. D. Toth, J. O'Rourke, and J. E. Goodman, *Handbook of discrete and computational geometry*. CRC press, 2017.
- [74] 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.
- [75] S. Fortune, "Voronoi diagrams and delaunay triangulations," *Computing in Euclidean geometry*, pp. 225–265, 1995.
- [76] J.-D. Boissonnat and M. Yvinec, *Algorithmic geometry*. Cambridge university press, 1998.
- [77] D. Attali, J.-D. Boissonnat, and A. Lieutier, "Complexity of the delaunay triangulation of points on surfaces the smooth case," in *Proceedings of the nineteenth annual symposium on Computational Geometry*, 2003, pp. 201–210.
- [78] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual-inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2016.
- [79] B. Zhou, J. Pan, F. Gao, and S. Shen, "Raptor: Robust and perception-aware trajectory replanning for quadrotor fast flight," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1992–2009, 2021.
- [80] B. Zhou, Y. Zhang, X. Chen, and S. Shen, "Fuel: Fast uav exploration using incremental frontier structure and hierarchical planning," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 779–786, 2021.
- [81] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3354–3361.
- [82] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of michigan north campus long-term vision and lidar dataset," *The International Journal of Robotics Research*, vol. 35, no. 9, pp. 1023–1035, 2016.
- [83] T.-M. Nguyen, S. Yuan, M. Cao, Y. Lyu, T. H. Nguyen, and L. Xie, "Ntu viral: A visual-inertial-ranging-lidar dataset, from an aerial vehicle viewpoint," *The International Journal of Robotics Research*, vol. 41, no. 3, pp. 270–280, 2022.
- [84] D. Cernea, "OpenMVS: Multi-view stereo reconstruction library," 2020. [Online]. Available: <https://cdseacave.github.io/openMVS>
- [85] C. Yuan, J. Lin, Z. Zou, X. Hong, and F. Zhang, "Std: Stable triangle descriptor for 3d place recognition," *arXiv preprint arXiv:2209.12435*, 2022.
- [86] J. Lin and F. Zhang, "A fast, complete, point cloud based loop closure for lidar odometry and mapping," *arXiv preprint arXiv:1909.11811*, 2019.

Supplementary Material: An additional trial of our lossless texture reconstruction based on ImMesh

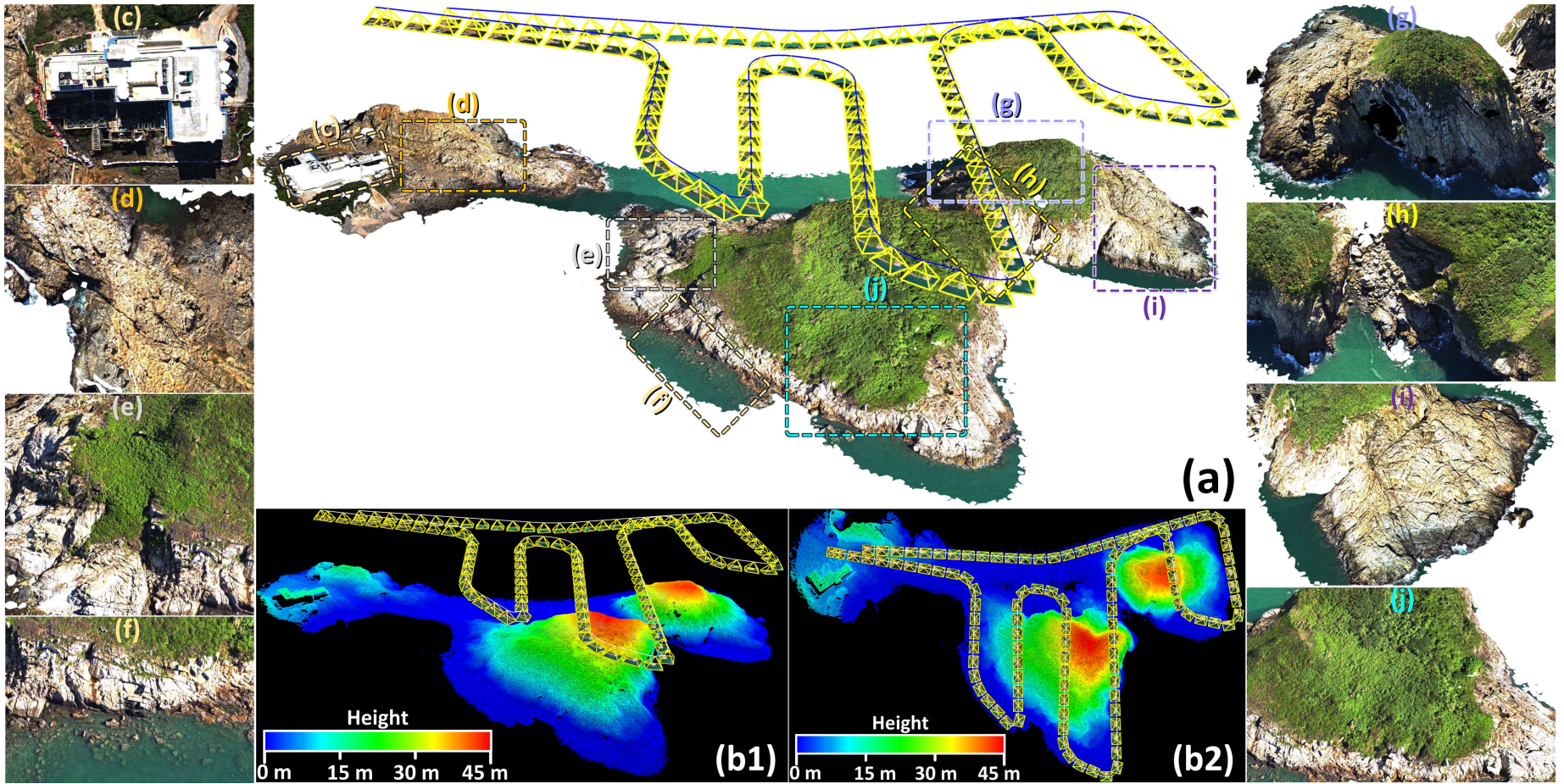


Fig. 1: In this trial, we collected the data by flying over islands in an “B”-like trajectory, as the blue path shown in (a). (b1) and (b2) show the side view and bird view of our reconstructed triangle mesh, where the mesh is colored by their altitude w.r.t. the sea level. (a) show the overview of our lossless texture reconstruction result, where we use the estimated camera poses (the yellow frustums) of R³LIVE++ for texturing the mesh with the collected images. The entire texture reconstruction of this 578 s sequence only costs 1210 s (on Intel i9-10900), with 583 s for ImMesh, 587 s for R³LIVE++, and 40 s for texturing. To see the detailed reconstruction process of the scene, please refer to our video on YouTube: youtu.be/pzT2fMwz428?t=892.