

# Automatic Indoor 3D Surface Reconstruction with Segmented Building and Object Elements

Eric Turner and Avidah Zakhor  
University of California Berkeley  
Berkeley, CA 94720  
{eltturner, avz}@eecs.berkeley.edu

## Abstract

*Automatic generation of 3D indoor building models is important for applications in augmented and virtual reality, indoor navigation, and building simulation software. This paper presents a method to generate high-detail watertight models from laser range data taken by an ambulatory scanning device. Our approach can be used to segment the permanent structure of the building from the objects within the building. We use distinct techniques to mesh the building structure and the objects to efficiently represent large planar surfaces, such as walls and floors, while still preserving the fine detail of segmented objects, such as furniture or light fixtures. Our approach is scalable enough to be applied on large models composed of several dozen rooms, spanning over 14,000 square feet. We experimentally verify this method on several datasets from diverse building environments.*

## 1. Introduction

The ability to automatically and rapidly generate a mesh of building surfaces is important to many fields, such as augmented and virtual reality, gaming, simulation, architecture, engineering, construction, and emergency response services. In this paper, our goal is to generate information-rich virtual models of indoor building environments. These models contain 3D geometry for the interior surfaces of buildings, including both large-scale building surfaces and small-scale features such as furniture.

We aim to improve on existing methods by combining two fundamentally different surface reconstruction techniques for building environments. We first generate a fully detailed volumetric model of the environment. We then combine this model with a highly simplified representation of the same building, which only contains geometry for the floors, walls, and ceilings of the model. This combination allows us to segment volumetric representations of the in-

terior objects in the environment, such as furniture or light fixtures, from the permanent surfaces of the building. We generate accurate, watertight models of objects in the building distinct from the building model itself, as demonstrated in Figure 1. Figure 1a shows a photograph of the scanned environment. Figure 1b shows an octree representation of the complex environment geometry. We use this representation to produce a rich model of the environment, as shown in Figure 1c. The objects of the environment, shown in white, are separated from the building structure, as shown in Figure 1d. We use different meshing techniques for the objects and the building itself to ensure the best representation of each type of surface. The result is a rich model of the environment that represents a whole building based on level, room, or individual objects.

The input scans for our modeling approach come from an ambulatory indoor scanning system [4]. By walking through the indoor environment at normal speeds, we can accurately estimate the trajectory of the system over time and localize the system [5]. Subsequently, we automatically generate a 3D model of the environment with the method described in this paper. This procedure allows us to rapidly move through a large environment, spending only a few seconds in each room yet capturing full geometry information. Since our system is mobile and ambulatory, the resulting point clouds have higher noise than traditional static scanners due to natural variability in human gait. As such, our approach needs to be robust to motion induced scan-level noise and its associated artifacts. We use these scans to form a watertight model of the indoor building environment and the contained objects.

The ability to identify and separate the objects within a model from the rest of the building geometry enables a richer representation of the architecture. Floors, walls, and ceilings tend to be large, planar, and can be meshed efficiently with few triangles. Furniture and other objects in the building tend to have high detail and are often more organic in shape. Thus, it is desirable to mesh these objects with a surface reconstruction scheme that preserves this de-

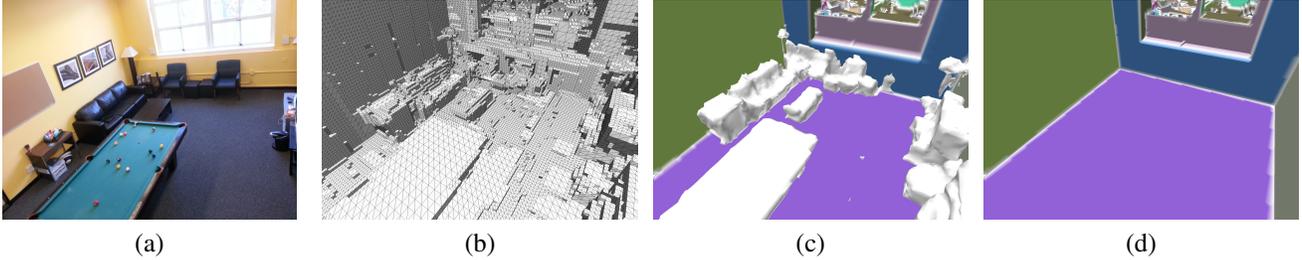


Figure 1: An area modeled by our technique: (a) a photo of the room; (b) the volumetric boundary of room; (c) final mesh with room and objects modeled; (d) final mesh of room only, colored by planar region.

tail. Not only do we use separate meshing techniques for each part of the model, but also we represent the geometry of objects in the environment at a finer resolution than the building surfaces, saving on processing and memory.

This paper is organized as follows: Section 2 provides background information on existing approaches to building modeling. Section 3 discusses the details of each step of our approach. Section 4 shows results of our method applied to a variety of building environments, and Section 5 concludes by highlighting the important aspects of our algorithm.

## 2. Related Work

Many existing building meshing approaches simplify the output model by representing only the primary building surfaces, such as floors, walls, and ceilings [6, 24]. Several techniques first generate 2D floor plans that can be extruded into simple 3D models [1, 16, 22]. These approaches allow for accurate wall geometry and reduce the complexity of the output model. Extruded floor plans also allow for models to explicitly define floors, walls, and ceiling surfaces.

There are several alternate approaches that attempt to capture full geometry detail of the area scanned [3, 8, 14, 21]. These methods include detail of all objects from the scene, not just the primary building elements. Kinect Fusion is an especially popular method [18, 23, 25]. This method represents a model with voxel data using a Truncated Signed Distance Function (TSDF) to generate a mesh using Marching Cubes [13]. In this paper we use a probabilistic field function on octrees, similar to [7, 9], to represent our volumetric model. Our hardware system uses 2D time-of-flight (TOF) lasers to produce much sparser point clouds than Kinect-based scanning, so it is not feasible to employ the same averaging techniques that allow for high-quality Kinect scans. The advantage of these 2D TOF sensors is that they are less noisy, have a longer range, and a wider field of view. The result is that much larger environments can be covered in significantly less time when using these sensors in a sweeping motion [20]. For instance, a backpack-mounted ambulatory system allows an unskilled operator to rapidly walk through an environment to acquire

data, spending only a few seconds in each room, whereas Kinect-based approaches require several minutes per room.

Another important meshing technique is to detect and classify objects in laser scans [2]. Understanding the semantics of a scene enables more sophisticated modeling of detected objects. Many existing techniques segment objects in the environment by explicitly detecting furniture geometry in the scanned data, matching that geometry to a database of known shapes, and retrieving template geometry for the detected item [11, 17]. This approach produces high-quality meshes, since each exported object is represented by a hand-modeled template, but can yield incorrect representations of unknown or unusual objects, which can occur when scanning especially complex scenes or if an object is in an unexpected orientation. In Section 4, we show scans of medical equipment in Figure 8 that would not likely pre-exist in shape databases. Recently, object detection methods from indoor scans have been proposed without the use of training data [15]. Even though this approach can be applied to large datasets it assumes very basic building geometry to segment objects, can only detect objects of certain complexity, is unable to detect very small objects, and requires objects to be repeated often in the environment. We expand on this work by segmenting objects volumetrically, rather than in the point cloud domain.

## 3. Approach

Our proposed method takes scan data from a set of laser sensors as input. We model the uncertainty of the positions of these input scans from a variety of possible noise sources to identify the observed volume likely to be part of the model. We use a volumetric approach to preserve model detail in the presence of noise, where each point in space has some probability of being *interior* or *exterior*. We define *interior* space to be empty or open area that range scans can pass through. We define *exterior* space to be solid material in the environment, including furniture and building structure. We store this volumetric labeling in an octree.

The primary goal of our approach is to use this volumetric information to form two watertight meshes of the envi-

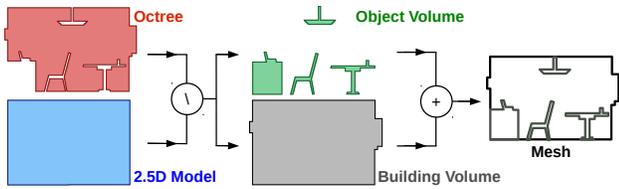


Figure 2: The scanned volume is meshed using two approaches that are combined to separate room geometry and object geometry. The complex geometry from the octree (upper left, in red) and the simple geometry from the 2.5D model (lower left, in blue) are combined to extract the object volume (upper center, in green) and the building volume (lower center, in grey). These volumes are meshed separately and exported (right, in black).

ronment. The first mesh only represents the building geometry, including floors, walls, ceilings, windows, and doors. The second mesh represents the objects in the environment such as furniture, light fixtures, or other items. The block diagram of our segmentation method is shown in Figure 2. We first generate two representations of the same environment. The populated octree represents a complex model of the volume, as shown by the red graphic in the upper-left of Figure 2. We then generate a simplified 2.5D model of the same volume, as shown by the blue graphic in the lower-left of Figure 2. This simplified model is obtained by first generating a 2D floor plan of the scanned area, then extruding the floor plan to form a 2.5D volumetric model that does not represent any interior objects.

We perform a volumetric set difference between these two models, keeping the volume that is labeled *exterior* by the octree and *interior* by the extruded floor plan. This volume represents the objects in the environment and is shown as the green graphic in the upper-center of Figure 2. Similarly, we can denote the union of the *interior* space of both models to be the building geometry, as shown in the lower-center of Figure 2. The object geometry is refined and meshed uniformly to preserve its fine structure. The building geometry is split into planar surfaces and each surface is triangulated efficiently, preserving the sharp corners between floors, walls, and ceilings. These two meshes form the whole environment, shown by the graphic at the right-side of Figure 2. This approach has the added benefit of modeling hidden surfaces, such as the backs of furniture or areas of walls occluded by objects.

In Section 3.1, we discuss how we probabilistically model the input scans. In Section 3.2, we describe how these scans are efficiently combined to generate a unified occupancy estimate for the entire scan volume. In Section 3.3, we detail how the octree is used to produce a simplified building model by first generating a 2D floor plan and then extruding the floor plan into a 2.5D mesh. In Sec-

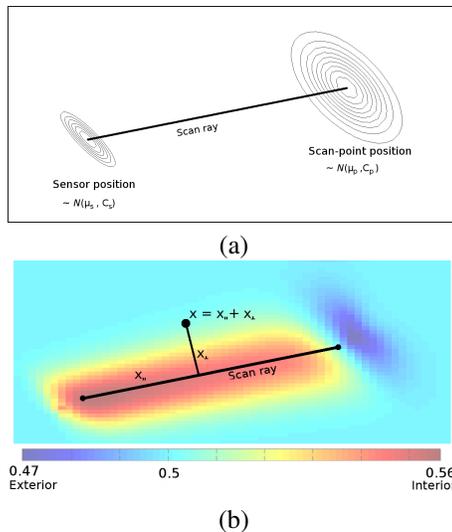


Figure 3: Example curve mapping: (a) the spatial distribution of sensor location and scan-point positions; (b) the computed curve mapping, indicating the areas estimated to be interior and exterior based on this scan.

tion 3.4, we discuss how the complex model and the simplified 2.5D extruded floor plan model are merged to segment the volume delineating objects such as furniture in the environment. Lastly, in Section 3.5, we document how we generate two types of watertight surfaces, one for the objects in the environment and the other for the building constructs. Each of these surface reconstruction techniques are geared to efficiently characterized their respective parts of the building.

### 3.1. Scan Preprocessing

In this section, our goal is to convert the input set of laser scan points into a labeling of space where each location  $\vec{x} \in \mathbb{R}^3$  is assigned a likelihood of being *interior* or *exterior*. First, we form a probabilistic model of each scan point’s position, coupled with the position of the originating sensor. Second, we use this probability model to estimate the scan point’s vote for the *interior* likelihood of each location in space intersected by its scan ray. Once we obtain these estimates for each scan point, we generate an occupancy model of the entire scanned environment, as discussed in Section 3.2.

First, we compute an estimate of the 3D positions for each scan point and the point’s originating sensor. These values are represented as two 3D Gaussian distributions. For each input scan, the sensor position is represented by Gaussian  $N(\mu_s, C_s)$  and the scan point position is represented by  $N(\mu_p, C_p)$ . For tractability, assume scan frame’s distribution is independent from the position of other scan frames. An example of this sensor/scan-point configuration

is given in Figure 3a.

The uncertainty in the position values of each scan point originate from three independent sources of error: the localization estimate, the timestamp synchronization, and intrinsic sensor noise. Localization noise arises from errors in the estimate of the system trajectory, as detailed in [5], and is by far the largest source of error, with typical standard deviations on the order of 20 cm.

Timestamp synchronization errors are due to our system combining measurements from several sensors, whose timestamps need to be transformed to a common clock. Mis-synchronization of timestamps can contribute spatial errors of scan points, especially when the system is moving or rotating rapidly while scanning distant objects. In these cases, an estimate of the scan point’s position changes depending on our estimate of when a scan is taken. However, since our sensors are synchronized to an accuracy of approximately 1 ms, synchronization error is usually the lowest source of noise in the scan points, contributing uncertainty to scan point positions of under 1 cm.

The third source of noise depends on the sensor hardware. Our system uses Hokuyo UTM-30LX sensors, whose intrinsic noise characterization is given in [19]. Typically this noise contributes on the order of 1 to 2 cm to the standard deviation of the positional estimate of scan points. This uncertainty value increases as the range of the point increases, with accurate measurements stopping at a range of 30 m.

The covariance matrices associated with each of these three sources of noise can be added to determine the net uncertainty for the positions of each sensor/scan-point pair. The uncertainties for the positions of each sensor and scan-point are represented with covariances  $C_s$  and  $C_p$  respectively. Next, we use this estimate for each scan point to form a “carve mapping”  $p(\vec{x}) : \mathbb{R}^3 \mapsto [0, 1]$ , which describes the likelihood of any location  $\vec{x} \in \mathbb{R}^3$  of being *interior* or *exterior* based on the position estimates from a scan point. We define  $p(\vec{x})$  as

$$p(\vec{x}) = F_s(x_{\parallel}) f_{\perp}(x_{\perp}) (1 - F_p(x_{\parallel})) + 0.5(1 - f_{\perp}(x_{\perp})) \quad (1)$$

where  $F_s(\cdot)$ ,  $F_p(\cdot)$ , and  $f_{\perp}(\cdot)$  represent one-dimensional marginal distributions derived from the scan ray model, as described below. We split a location  $\vec{x} = \vec{x}_{\parallel} + \vec{x}_{\perp}$ , as shown Figure 3b, where  $\|\vec{x}_{\parallel}\|$  is the distance along the length of the scan ray, and  $\|\vec{x}_{\perp}\|$  is the distance orthogonal to the scan ray. This decomposition allows us to weigh the influence of the scan ray on our estimate  $p(\vec{x})$  so represent a fall-off as  $\|\vec{x}_{\perp}\|$  increases.  $F_s(x_{\parallel})$  is the marginal one-dimensional (1D) Cumulative Distribution Function (CDF) of the sensor position’s distribution along the length of the scan ray, derived from the Gaussian distribution of the sensor position. Similarly,  $F_p(x_{\parallel})$  is the marginal CDF of the scan-point position’s distribution along the length of the ray.

Lastly,  $f_{\perp}(x_{\perp})$  is the 1D Probability Mass Function (PMF) of the lateral position of the scan ray.

The combination of these values in Equation 1 yields the mapping shown in Figure 3b. Values in blue are less than 0.5, indicating a likelihood of a location being *exterior*. As shown, these values occur just past the scan position. Values in red are greater than 0.5, indicating a likelihood of a location being *interior*. As the query location  $\vec{x}$  moves away from the scan ray and  $\|\vec{x}_{\perp}\|$  increases, then  $p(\vec{x})$  approaches 0.5, indicating unknown state. In the next section, we discuss how the estimates from all input scans are used to generate a model of the entire environment.

### 3.2. Carving

Given a carve mapping function for each scan point taken during the data acquisition process, we merge all the scans spatially to obtain a fused probabilistic estimate  $p_f(\vec{x})$  for any point  $\vec{x} \in \mathbb{R}^3$ . The value  $p_f(\vec{x})$  is computed as the maximum-likelihood estimate based on all nearby scans, where any scan whose mean scan-line position is more than 3 standard deviations away from  $\vec{x}$  does not contribute to the estimated value at  $p_f(\vec{x})$ . The final result for the spatial labeling of  $\mathbb{R}^3$  is stored in an octree. The advantage of the octree is that every point in space can be represented, but certain areas can have finer detail than others.

The leaf nodes of the octree contain the compiled probabilistic model of the degree to which that node is labeled as *interior* or *exterior*. Each leaf node  $L$  contains the fused value of  $p_f(\vec{x})$  for all  $\vec{x} \in L$ , variance of the samples of  $p(\vec{x})$  from each intersecting scan ray, and the of number of scans that intersect  $L$ . All these statistics are used later in the pipeline for analyzing the properties of  $L$ . As an example, if  $p_f(\vec{x})$  is 0.5 or less, then the node is considered *exterior*. Nodes that are never intersected by scans are assumed to be *exterior* and are assigned a value of  $p_f(\vec{x}) = 0.5$ . If the value of  $p_f(\vec{x})$  is strictly greater than 0.5, then the node is considered *interior*. The faces between *interior* nodes and *exterior* nodes are considered boundary faces of the octree, and are useful for determining the position of generated meshes. As we discuss in Section 3.5, the position of the mesh between two such nodes is placed to sub-voxel accuracy using our estimates of  $p_f(\vec{x})$  for each node.

We typically use a leaf resolution of 5–10 cm, but the final tree is only stored at full depth in the locations that require it, which are boundaries between interior and exterior spaces. As we discuss in Section 3.4, areas of the environment that are segmented as objects in each room will be re-carved to an even finer resolution, typically 1 cm or less, since these locations are likely to contain high detail.

### 3.3. Simplified Model Generation

In this section, we describe how we use the populated octree from Section 3.2 to generate a simplified model that

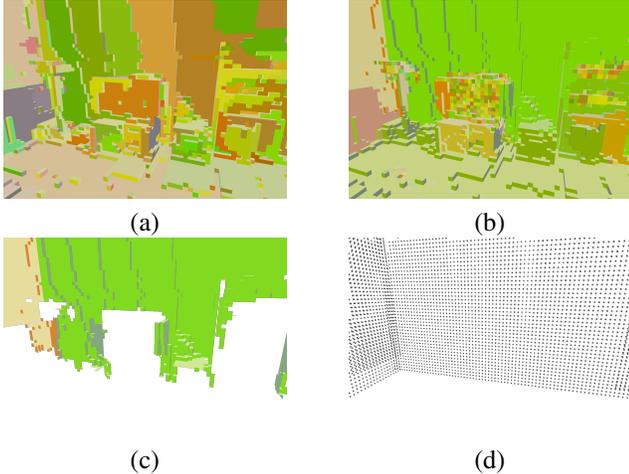


Figure 4: Generating wall samples from an octree: (a) initialize regions on the octnodes’ boundary faces; (b) perform region growing to form large planar regions; (c) filter out wall regions; (d) generate points along planar regions to make wall samples.

only represents the primary building structure. We use an existing technique that produces a 2D floor plan of the environment and extrudes a 2.5D model using the height information of each room [22]. We first need to generate a set of wall samples in the environment, which are a set of points in 2D space that are locations with high likelihood of being wall positions. This set of points is used by the 2D floor plan generation procedure as input data [22]. In prior work, these wall samples are generated by sampling a 3D point cloud of the environment. In this paper, we use the volumetric octree model to generate wall samples. We show that this approach not only produces a floor plan better aligned with the complex geometry of the octree, but also one that is less affected by clutter, such as furniture, than when using the point cloud directly to generate the floor plan.

The first step of generating wall samples from the octree is to identify large planar surfaces. We cluster the boundary faces of the octree into planar regions that represent all surfaces in the model as flat, planar structures. Figure 4a shows the initial boundary regions of a model, with each initial region depicted as a separate color. These regions are formed bottom-up by iteratively consolidating boundary node faces into regions via Principal Component Analysis (PCA) of boundary face positions, using the process described by [21]. This step produces a single planar region for each dominant surface of the model, as shown in Figure 4b. We then filter the regions based, keeping only the surfaces that are within  $5^\circ$  of vertical and at least 1 m tall, as shown in Figure 4c. The output regions have gaps corresponding to the portion of the walls hidden behind any furniture in the model. To counteract these occlusions, we ex-

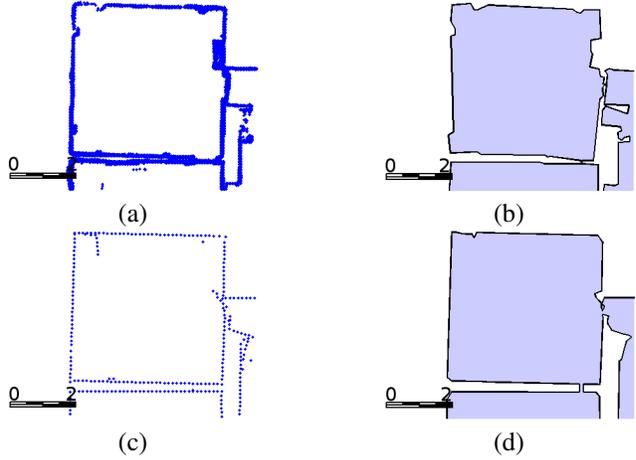


Figure 5: Comparison of wall samples and floor plans: (a) wall sampling generated from original point cloud; (b) corresponding floor plan; (c) wall sampling generated from octree; (d) corresponding floor plan. All units are in meters.

pand the represented geometry of each wall to include any *exterior* points that are within the 2D convex hull of each wall planar region. Figure 4d shows a set of these points, sampled uniformly, across the wall plane. Once we obtain these 3D wall positions, we use these generated points to estimate 2D positions of vertical surfaces. Since the input to this method is a uniformly sampled set of points along the surface, the produced 2D wall samples more faithfully represent the environment, as demonstrated in Figure 5. Since wall samples are generated from connected surfaces and not raw scan points, there are with fewer artifacts due to clutter and furniture.

Once we use the octree to generate the wall samples for a model, we can feed those samples into the 2D floor plan generation method discussed in [22]. This method produces a watertight 2D model that defines the scanned area. Figure 5 shows a comparison of floor plans generated with the octree versus floor plans generated with the raw scan points. Figures 5a and 5b show the wall samples and the floor plan generated from the raw scan points, respectively. Obstacles in the environment such as furniture are not properly removed, causing the output walls to be noisy. Figures 5c and 5d show the wall samples and floor plan generated from the octree. Since the region merging allows more sophisticated separation of large planar surfaces, the furniture in the environment is properly removed and the output model is cleaner. As an example, the upper-right corner of this model contains a large bookcase, which appears in the wall sampling in Figure 5a and is propagated to the floor plan in Figure 5b. This bookcase is correctly removed from the wall samples generated from the octree in Figure 5c, which means the wall is more appropriately represented in

the floor plan shown in Figure 5d.

Height information is stored in this floor plan, so a 2.5D model can be extruded, resulting in a simplified representation of the floors, walls, and ceilings in the environment. Since both the 2.5D floor plan and the 3D octree are volumetric representations of the environment, we can segment the objects in the environment by searching for locations that are *interior* in the floor plan, but *exterior* in the octree. These locations coincide with furniture and other objects that are removed by the 2.5D floor plan construction.

This wall sampling approach is less direct than using the original scans, but it has a number of advantages. If a wall is partially occluded by a large object, such as a tall bookshelf, then we can use the portions of the wall on either side of the object to confirm that the entirety of the wall is represented. Using only the raw scans, the portion of the wall behind the bookshelf would be under-represented in the final output, causing errors in the floor plan. By generating these wall samples using the octree as input, we can ensure that the final floor plan is well-aligned with the octree geometry. This alignment is important for the steps described in Section 3.4, where we combine the simplified model back into the octree geometry. The effect of misalignment can be seen by comparing Figures 6b and 6c. In Figure 6b, the octree was segmented using a floor plan generated directly from the point cloud and not the octree. As such, parts of the back wall and window are mislabeled as objects and kept in the output. However, in Figure 6c, the octree was segmented using a floor plan generated via the technique described in this section. The back wall is no longer mislabeled and only the actual furniture in the environment are segmented as objects.

### 3.4. Merging Models

Both the extruded floor plan and the original octree are volumetric models of the environment, so we can classify the overlapping volumes into three categories. First, locations that are *exterior* in the octree yet *interior* in the extruded floor plan are objects or furniture in the environment. Locations labeled *exterior* by both models are considered part of the building structure. Lastly, all locations labeled *interior* by the octree are considered open space interior to the building, regardless of the extruded floor plan’s labeling. Volume intersected by the boundary of the 2.5D floor plan is considered *exterior*, since these represent the primary building surfaces and not objects within the building. Using this segmentation, we can now consider the objects in the building separately from the 2.5D building structure. In Figure 6a, we see the original octree leaf nodes of a scanned environment. By performing a set difference of the octree volume from the volume of the 2.5D model of the environment, we can extract the furniture and other objects. Figure 6b shows the segmentation using an unaligned floor plan

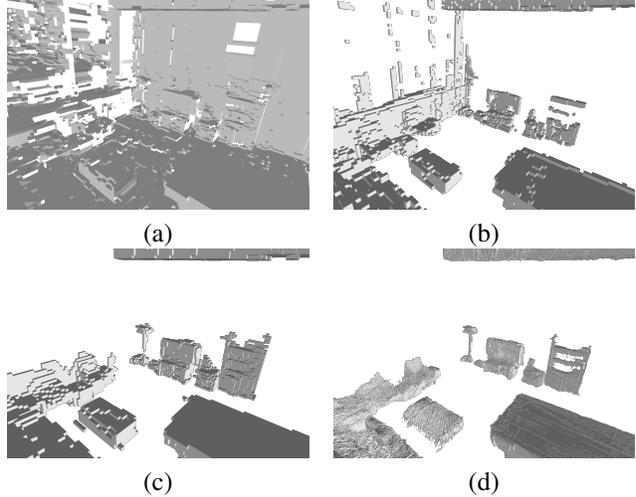


Figure 6: Example of aligning floor plan to segment objects: (a) original octree nodes, at a leaf resolution of 6.25 cm; (b) segmented objects using unaligned floor plan; (c) segmented objects using aligned floor plan; (d) the segmented objects are re-carved to a leaf resolution of 0.8 cm.

and Figure 6c shows the result with a fully aligned floor plan. The unaligned floor plan was generated directly from the raw point cloud of the scans [22], whereas the aligned floor plan was generated with our method described in Section 3.3. With a properly segmented representation of the room’s objects, we can re-carve the nodes of the octree containing object geometry, since these locations tend to have finer detail than the rest of the model. Figure 6d shows an example of this re-carving, which has been refined from the original resolution of 6.25 centimeters to a new resolution of less than a centimeter.

Once we have fully merged these models, each node of the octree is labeled as either object geometry or room geometry. Since the object geometry is represented volumetrically, we can easily represent watertight models of each individual object based on connected components. In addition to refining the resolution of object nodes for more accurate representation, we can use this labeling to adjust how we generate a mesh for each portion of the environment.

### 3.5. Meshing

After segmenting the octree geometry into objects and rooms, we can mesh each separately. Objects such as furniture and light fixtures tend to have higher detail than the room-level geometry. The room-level geometry tends to be composed of large, planar surfaces. We use a dense meshing technique to represent the object geometry, which preserves detail and curves in the geometry. For the room-level geometry, we identify planar regions and mesh each plane efficiently with large triangles. Figure 7 shows an example

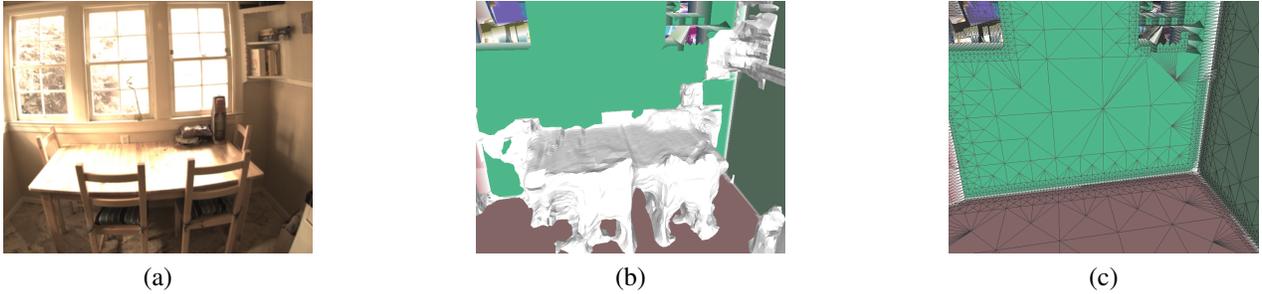


Figure 7: Example meshing output of residential area: (a) photo of area; (b) all reconstructed geometry; (c) geometry of room surfaces only, colored by planar region.

of applying both of these methods to a given model. Figure 7a shows a photograph of the scanned area (a kitchen table) and Figure 7b shows the final output of all meshing approaches combined.

To mesh building geometry, we first partition the boundary faces of the octree into planar regions, in a similar fashion to the method described in Section 3.3 for wall sampling. Each planar region represents a set of boundary faces along with fitting plane geometry. To generate a watertight mesh, we find the intersection points between each pair of neighboring planes and insert vertices for our output mesh.

Planar region fitting on voxel data was performed in [21]. They intersect the fitting planes of each region to determine the locations of output mesh vertices shared by multiple regions. This process may produce artifacts or self intersections at locations where two nearly-parallel regions are neighbors. We instead use a pseudo-intersection point that is closer to the original corner position. If we took the intersection point of all planes, the vertex position may be under-constrained if some of the regions are close to being parallel. We perform singular value decomposition of the space of plane normal vectors to determine if this basis is under-constrained. Any under-represented dimensions in a vertex’s position are set to the original node corner position. This process produces connecting vertices between planar regions that reside as close to the geometric intersection of the fitted planes as possible, without producing degenerate artifacts in the final mesh.

Once the locations of vertices shared by two or more planar regions are computed, then the interior area of each region is triangulated using a 2D variant of isosurface-stuffing [12], as described in [21]. An example of this meshing technique can be seen in Figure 7c. This method to represent building elements is important for features that do not follow the 2.5D assumption, such as windows or door-frames. As shown in Figures 1d and 7c, the planar mesh of the building surface still provide geometry for features such as window recesses.

When meshing object geometry, we use a variant of Dual Contouring [10], since it works well with adaptively-sized

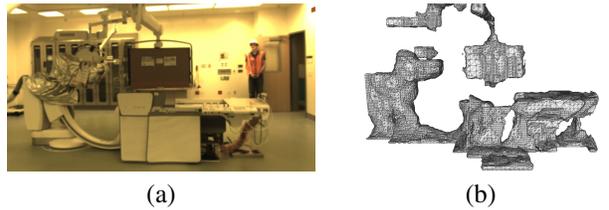


Figure 8: Example scan of equipment in hospital’s hybrid operating room: (a) picture of scanned area; (b) object model triangulation of operating table and equipment.

nodes in an octree and represents both curved and sharp features in the output geometry. Since our data labels are divided into node centers of the tree, rather than node corners, we perform dual contouring by mapping each boundary face of the octree to a vertex in the mesh and each corner of the octree into a face in the meshed output. The vertex position of the mesh is offset from each node face based on the stored probability value  $p_f(\vec{x})$  of that node. This offset positions the mesh at the  $p_f = 0.5$  isosurface, which provides sub-voxel accuracy for the generated surface position.

An important aspect of meshing these two segments separately is to ensure watertightness of building and object models. The surfaces of walls hidden behind any occluding objects are still meshed, even though they are never directly scanned. This effect can be seen in Figure 7c. Similarly, the hidden surfaces of objects are also fully meshed.

## 4. Results

Our technique generates models of large scanned environments and still preserves fine detail of objects in those environments. In this section, we discuss the advantages of our method and show example results. All models shown were generated on an Intel Xeon 3.10 GHz processor.

In Figure 8, we show results for a scan we generated of several rooms in a hospital operating area. This model contains four rooms, covering a total of 1,937 square feet, and was scanned using our hardware system in 2 minutes

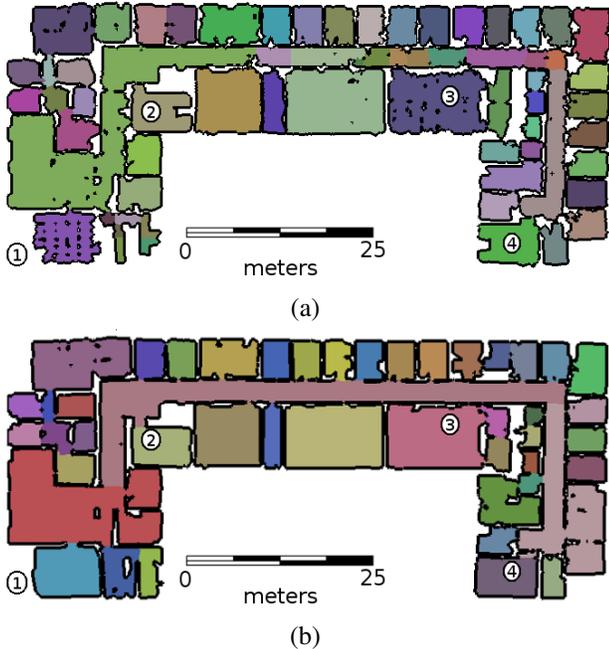


Figure 9: Comparison of floor plans of large office environment: (a) floor plan generated from raw point cloud scans using [22]; (b) floor plan generated from octree proposed here. The octree floor plan has much fewer artifacts due to clutter or furniture, as shown in areas “1” through “4”.

47 seconds. Processing this model took a total of 6 hours and 8 minutes. The room shown in Figure 8a is an operating room, which contains several medical scanners affixed to the ceiling. Our approach automatically segments the geometry of the scanners from the rest of the building and generates a mesh for the entire environment. Figure 8b shows the segmented geometry of the operating table and medical equipment. This object would be difficult to model with techniques that semantically classify geometry to form a mesh, since it is unlikely that shape libraries would have many examples of such a usual device. Since our technique does not need to classify the shape, we can still generate an accurate representation of its geometry.

As shown in Section 3.3, one of the by-products of our approach is a 2D floor plan of the scanned environment. When compared to floor plans generated from raw point cloud scans [22], our proposed approach yields not only better alignment to the complex geometry, but also a cleaner floor plan. We demonstrate this contrast with Figure 9, which represents a scan of a 14,079 square foot office area with over 50 rooms. This data acquisition took 25 minutes and processing took 12 processor hours. An example floor plan from the previous method is shown in Figure 9a. This floor plan has several artifacts caused by clutter and furniture in the environment. Locations “1” and “3” show

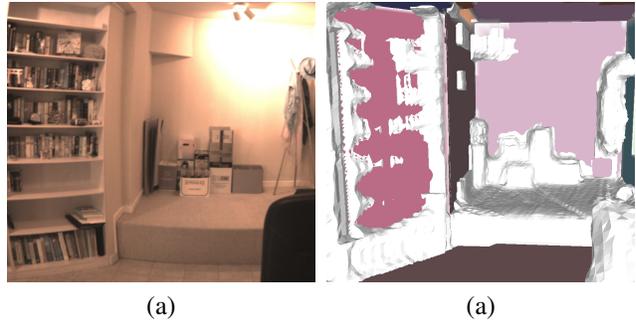


Figure 10: Example mesh of bookcase and boxes: (a) photograph of scanned area; (b) generated mesh, showing both building and object geometry.

rooms filled with large amounts of objects, causing holes in the floor plan. Locations “2” and “4” show rooms with a single large object that occluded part of a wall, causing an incorrect notch in the floor plan. The floor plan of the same environment generated by our proposed technique is shown in Figure 9b. This floor plan correctly separates the rooms of the environment and does not have the same artifacts as in the previous method. Additionally, long hallways in the building are correctly represented as one room, rather than being split into several small segments.

Our approach has some limitations. Since we perform volumetric intersections with an extruded 2.5D model based on a floor plan, our output relies on assumptions this model makes about the scanned environment: that each room has fixed floor and ceiling heights. If a room’s ceiling is not horizontal, then it is approximated with a horizontal surface. Figure 10 shows a set of boxes on top of a raised platform next to a bookcase. The raised platform is identified as a separate object, since it is at a different elevation than the rest of the floor in this room. This figure also shows a floor-to-ceiling bookcase. Since the position of fitted walls is found by looking for vertical surfaces, it is difficult to accurately gauge the depth of the shelves, especially when they are filled, so only part of the bookcase is segmented as an object. Other parts of the model are still meshed correctly in the presence of these issues.

## 5. Conclusion

We present a robust method of surface reconstruction designed for indoor building environments. Our method takes input laser scans with high noise and forms detailed models of the objects in the scanned area, as well as models of the building structure itself. Our technique can handle diverse building environments including residential apartments, office areas, and hospitals. Our approach allows partitioning of building geometry into levels, rooms, and individual objects. This degree of segmentation is an important step to

automatically generating richly defined Building Information Models that represent all aspects of the environment.

## References

- [1] R. Cabral and Y. Furukawa. Piecewise planar and compact floorplan reconstruction from images. *Computer Vision and Pattern Recognition (CVPR)*, pages 628–635, 2014. 2
- [2] C. Cadena and J. Kosecka. Semantic segmentation with heterogeneous sensor coverages. *IEEE International Conference on Robotics and Automation (ICRA)*, 2014. 2
- [3] A.-L. Chauve, P. Labatut, and J.-P. Pons. Robust piecewise-planar 3d reconstruction and completion from large-scale unstructured point data. *CVPR*, 2010. 2
- [4] G. Chen, J. Kua, S. Shum, N. Naikal, M. Carlberg, and A. Zakhor. Indoor localization algorithms for a human-operated backpack system. *3D Data Processing, Visualization, and Transmission*, May 2010. 1
- [5] N. Corso and A. Zakhor. Indoor localization algorithms for an ambulatory human operated 3d mobile mapping system. *Remote Sensing*, 5(12):6611–6646, October 2013. 1, 4
- [6] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Manhattan-world stereo. *Computer Vision Pattern Recognition (CVPR)*, pages 1422–1429, 2009. 2
- [7] C. Hernandez, G. Vogiatzis, and R. Cipolla. Probabilistic visibility for multi-view stereo. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, November 2007. 2
- [8] C. Holenstein, R. Zlot, and M. Bosse. Watertight surface reconstruction of caves from 3d laser data. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, September 2011. 2
- [9] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard. Octomap: an efficient probabilistic 3d mapping framework based on octrees. *Autonomous Robots*, 34(3):189–206, 2013. 2
- [10] T. Ju, F. Losasso, S. Schaefer, and J. Warren. Dual contouring of hermite data. *ACM Transactions on Graphics (TOG)*, 2002. 7
- [11] Y. M. Kim, N. J. Mitra, D.-M. Yan, and L. Guibas. Acquiring 3d indoor environments with variability and repetition. *ACM Transactions on Graphics*, 31(6), November 2012. 2
- [12] F. Labelle and J. R. Shewchuk. Isosurface stuffing: Fast tetrahedral meshes with good dihedral angles. *ACM Transactions on Graphics*, 26(3):57, July 2007. 7
- [13] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4):163–169, July 1987. 2
- [14] L. Ma, T. Whelan, E. Bondarev, P. H. N. de With, and J. McDonald. Planar simplification and texturing of dense point cloud maps. *2013 European Conference on Mobile Robots (ECMR)*, pages 164–171, September 2013. 2
- [15] O. Matusch, D. Panozzo, C. Mura, O. Sorkine-Hornung, and R. Pajarola. Object detection and classification from large-scale cluttered indoor scans. *Computer Graphics Forum*, 33(2):11–21, 2014. 2
- [16] C. Mura, O. Matusch, A. J. Villanueva, E. Gobbetti, and R. Pajarola. Automatic room detection and reconstruction in cluttered indoor environments with complex room layouts. *Computers and Graphics*, 44:20–32, November 2014. 2
- [17] L. Nan, K. Xie, and A. Sharf. A search-classify approach for cluttered indoor scene understanding. *ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH Asia*, 31(137), November 2012. 2
- [18] R. Newcombe, A. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. *Mixed and Augmented Reality (ISMAR)*, pages 127–136, 2011. 2
- [19] F. Pomerleau, A. Breitenmoser, M. Liu, F. Colas, and R. Siegwart. Noise characterization of depth sensors for surface inspections. *2nd International Conference on Applied Robotics for the Power Industry*, pages 16–21, September 2012. 4
- [20] M. Smith, I. Posner, and P. Newman. Adaptive compression for 3d laser data. *The International Journal of Robotics Research*, 30(7):914–935, June 2011. 2
- [21] E. Turner and A. Zakhor. Watertight planar surface meshing of indoor point-clouds with voxel carving. *International Conference on 3D Vision*, June 2013. 2, 5, 7
- [22] E. Turner and A. Zakhor. Floor plan generation and room labeling of indoor environments from laser range data. *International Conference on Computer Graphics Theory and Applications*, (9), January 2014. 2, 5, 6, 8
- [23] T. Whelan, M. Kaess, M. Fallon, H. Johannsson, J. J. Leonard, and J. McDonald. Kintinuous: Spatially extended kinectfusion. *CSAIL Technical Reports*, July 2012. 2
- [24] J. Xiao and Y. Furukawa. Reconstructing the world’s museums. *EECV 2012 Lectures in Computer Science*, 7572:668–681, 2012. 2
- [25] Q. Zhou and V. Koltun. Dense scene reconstruction with points of interest. *ACM Transactions on Graphics*, 32(112):112, July 2013. 2