Indoor Scene Reconstruction using Feature Sensitive Primitive Extraction and Graph-cut

Sven Oesau, Florent Lafarge, Pierre Alliez

Indoor Scene Reconstruction using Feature Sensitive Primitive Extraction and Graph-cut

Sven Oesau, Florent Lafarge & Pierre Alliez

Inria Sophia Antipolis - Méditerranée

Abstract

We present a method for automatic reconstruction of permanent structures, such as walls, floors and ceilings, given a raw point cloud of an indoor scene. The main idea behind our approach is a graph-cut formulation to solve an inside/outside labeling of a space partitioning. We first partition the space in order to align the reconstructed models with permanent structures. The horizontal structures are located through analysis of the vertical point distribution, while vertical wall structures are detected through feature preserving multi-scale line fitting, followed by clustering in a Hough transform space. The final surface is extracted through a graph-cut formulation that trades faithfulness to measurement data for geometric complexity. A series of experiments show watertight surface meshes reconstructed from point clouds measured on multi-level buildings.

Keywords: Indoor scenes, 3D reconstruction, laser scanning, multi-scale line extraction, graph cut, energy minimization.

1. Introduction

In recent years the reconstruction of 3D architectural scenes has received more and more attention. Blueprints and precise models are often needed for the architecture, engineering and construction application domains. As the physical geometry of buildings often differs from the original plans, the reconstruction of a precise 3D model of the interior is a common need. Such reconstruction is often a manual or semi-automatic time-consuming process.
Compared to the reconstruction of outdoor environments, indoor scene reconstruction is still in an early stage. Applying outdoor reconstruction methods to indoor scenes is not practical as indoor impose different challenges than outdoor scenes. The outside of a building can often be described by a single or few cuboids, and the amount of clutter hiding part of the geometry is rather low. In contrast, interior space often has a more complex geometry and a high quantity of clutter. Further hurdles come from the varying point density and anisotropic sampling.

1.1. Related work

Surface reconstruction has been an active research topic for decades. Despite the wide variety of methods, they often perform unsatisfactorily for extracting a surface representation from indoor scenes. They commonly result in a single surface representation approximating the entire point cloud, which is unsatisfactory for further specialized processing such as semantic labeling into floors, ceilings and walls. A favored approach would provide a classification of the point cloud into permanent structures and clutter. General surface reconstruction methods instead assume that

the point cloud is created from a single surface whereas indoor scenes are usually composed of planar parts and arbitrarily shaped clutter.

The RANSAC-based algorithm proposed by Schnabel et al. (2007) provides a valid solution for extracting several types of shapes from an oriented point cloud, i.e., from points with oriented normals. While missing normals can be estimated by robust feature preserving methods (Boulc'h and Marlet, 2012), adjusting the parameters of Schnabel's algorithm to detect primitives in indoor scenes is often a trial-and-error process. Another drawback for indoor scene reconstruction is that Schnabel's algorithm is not robust to highly variable point density and strong anisotropy. For the reconstruction of permanent structures from indoor scans, we advocate for considering a domain-specific knowledge. Common knowledge assumptions are piecewise planar permanent structures and Manhattan-World scenes, i.e., exactly three orthogonal directions: two for the walls and one for floors and ceilings. We classify the existing approaches into three categories: *Planar primitive detection*, *Volumetric primitive fitting* and *Mesh-based reconstruction*.

Planar primitive detection. Sanchez and Zakhor (2012) recently proposed a method for modeling building interiors as an arrangement of planar polygons. They classify the points based on estimated normals by assuming a Manhattan-World scene. The planar primitives are reconstructed by a region growing segmentation, least-squares fitting by employing RANSAC and alpha shapes (Edelsbrunner et al., 2006) to reconstruct polygons. Additionally, they employ model fitting of a staircase model to unsegmented points. The results presented in form of a polygon soup show that they can deal with multi-level buildings. There is however no structure information such as adjacency of primitives.
Another primitive detection approach has been proposed by Budroni and Boehm (2010); Budroni and Boehm. The authors apply sweeping techniques to identify Manhattan-World directions and to locate wall segments. The floor plane is decomposed by the so-detected segments and marked as inside or outside based on the point density. The output is a watertight model extracted from the ground plan. However, resilience to missing data is not addressed.
Another approach proposed by Okorn et al. (2010) uses the Hough transform to extract wall segments from the point cloud. Although the majority of walls are detected, they are often not captured in their full extent. The output of the algorithm is a set of unconnected wall segments that are used to classify the points into permanent structures and clutter. However, neither structural relations nor volumes of the indoor space are provided.
Adan and Huber (2011) build on top of this approach and focus on the modeling of wall surfaces. A learning mechanism is proposed for segmenting and labeling wall surfaces into parts either occluded by clutter, or occupied (solid), or empty (e.g. windows).

Volumetric primitive fitting. Xiao and Furukawa (2012) introduce a *constructive solid geometry* (CSG) based method to reconstruct large-scale indoor environments for visualization purposes. The authors propose a greedy algorithm for creating a CSG model, guided by an objective function devised to both measure quality and control detail. In order to reconstruct different floor and ceiling heights they first decompose the point cloud into horizontal slices and apply their method first in 2D, then in 3D to merge the 2D models. Although this approach can deal with more than two wall directions, it performs well on scenes mostly consisting of orthogonal or parallel structures. This limitation is mostly due to the primitive generation for the CSG model extraction: They detect linear wall segments and combine parallel or orthogonal ones for generating can-

didate primitives to be used by the CSG model. Jenke et al. (2009) proposed a greedy algorithm reconstructing the free space volume. This approach is based on Schnabel's algorithm to detect planar primitives. The primitives are first structured in a graph, and a graph matching is applied to locate the cuboid candidates for the greedy reconstruction. Due to the fitting of cuboids their approach is limited to rectangular geometry. In presence of clutter or missing data the graph matching may fail as it requires at least five planar primitives to fit a cuboid.

Mesh-based reconstruction. Newcombe et al. (2011) and Izadi et al. (2011) propose a surface reconstruction method from point clouds acquired by a low cost consumer-grade hand held scanner. By designing their approach for parallel execution on GPUs they achieve real-time performance. Reconstructing an indoor scene using a general surface reconstruction method and presenting the result as a surface mesh allows for a general and detailed representation. However, no semantic classification of the scene into wall, floor, ceiling or clutter is provided. Other work devised to semantize indoor scenes (Kim et al., 2012; Shao et al., 2012), focus on understanding clutter in indoor environments.

1.2. Contributions and overview

Our approach reconstructs volumetric models of indoor spaces, by extending a previous work (Oesau et al., 2013). Oesau et al. (2013) partition the bounding box into cells using walls detected by a Hough transform. An empty/solid space labeling is formulated as an energy minimization and solved using graph-cut. The final model is extracted as the union of all cells labeled empty. In this work the level of detail of the reconstructed model is limited by the construction of the space partitioning. By introducing a multi-scale line fitting we enhance the alignment of the space partitioning with the permanent structures, which is particularly relevant for highly detailed structures. Experiments from defect-laden data and different sensors, including a scene recorded by a Kinect handheld sensor, are provided. Our new approach substantially increases the amount of details and overcomes some of the above mentioned limitations:

- **Arbitrary wall directions:** The model is not restricted to the Manhattan-World geometry and deals with planar wall detection for arbitrary vertical directions. The only assumption is that floors and ceilings are horizontal.
- **Multi-level buildings:** Our approach reconstructs an entire building with multiple levels in a single optimization step without requiring *a priori* knowledge about the levels.
- **Missing and outlier data:** 3D space partitioning into volumetric cells and labeling of the cells by global energy minimization provides resilience to missing data and outliers.
- **Raw data:** To be as general and applicable as possible, only dense raw point sets and knowledge about vertical direction are required. Nevertheless, when oriented normals or knowledge about the scanning device position are provided, they can be used to further improve robustness.

Our method takes as input a point cloud $P = \{p_1, ..., p_n\} \in R^3$ and consists of two main steps depicted in Figure 1:

1. **Space partitioning:** The bounding box of P is partitioned into volumetric cells by using detected permanent structures as splitting planes. The vertical distribution of the point cloud is first analyzed in order to separate the points into slices containing horizontal/vertical structures and clutter. Each horizontal slice containing

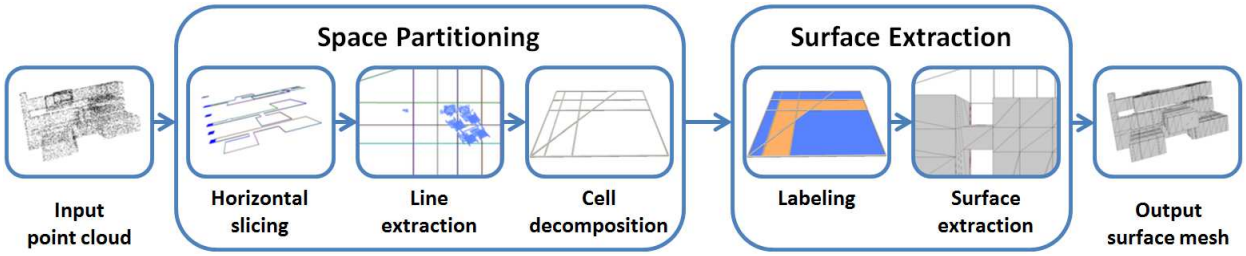

Figure 1: **Method overview**. The bounding box is split into cells using detected permanent structures. The final surface is obtained through labeling the cells in empty or solid space, and extracting the interfaces between empty and solid cells.

wall structures is then analyzed in 2D after vertical projection, in order to extract the wall line segments. Line extraction is performed through a multi-scale, feature-preserving line fitting method, followed by global clustering in a Hough transform space in order to favor alignment of the walls. Such alignment reduces the complexity of the volume partitioning.

2. **Surface extraction:** The volumetric cells created in previous step are labeled into either solid or empty space, respectively for permanent structures (walls, floors, ceilings) or outside. The final reconstructed surface is then deduced from the labeled cells.

2. Space partitioning

The goal of this step is to provide a 3D space partitioning of the bounding box of P into volumetric cells that well align to the empty and solid space. Using the permanent structures for partitioning the space provides such alignment. The following sub-steps describe the extraction of directions of horizontal and vertical permanent structures to create a 3D space partitioning.

2.1. Horizontal slicing

Walls are assumed to be vertical and perpendicular to floor and ceiling, and the permanent structures are assumed to be piecewise linear along the vertical direction. At every height where the permanent structures change we assume the presence of horizontal planar structures, i.e., floor or ceiling. In this step the point cloud is vertically partitioned into horizontal slices containing important horizontal structures and vertical slices containing vertical structures, i.e., walls. To detect the horizontal structures, we extend an idea from Okorn et al. (2010). The scanning of horizontal structures generates a high number of samples sharing similar heights. Hence, horizontal structures show up as peaks in the point distribution along the vertical axis. The presence of clutter challenges the detection as it shows up as noise in the distribution. To filter the clutter, we remove the points whose normals are not parallel to the vertical direction (Figure 2). If normals are not available, they are estimated through local principal component analysis (PCA) applied to a local neighborhood.

For extracting the peaks from the point distribution we create a histogram. The bin size has to be manually specified, a default value of 5-10 cm is suggested. The peaks in the distribution are located through mean

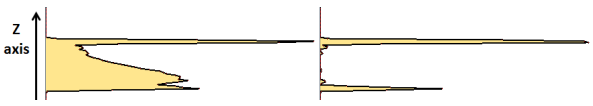

Figure 2: **Cluttered and uncluttered distributions.** Comparing point distributions along vertical direction n_z without filtering by normals (left) and with filtering by normals (right, $|n_i \cdot n_z| > 0.98$). The chosen bin size is 8 cm and the distribution is split into 69 bins.

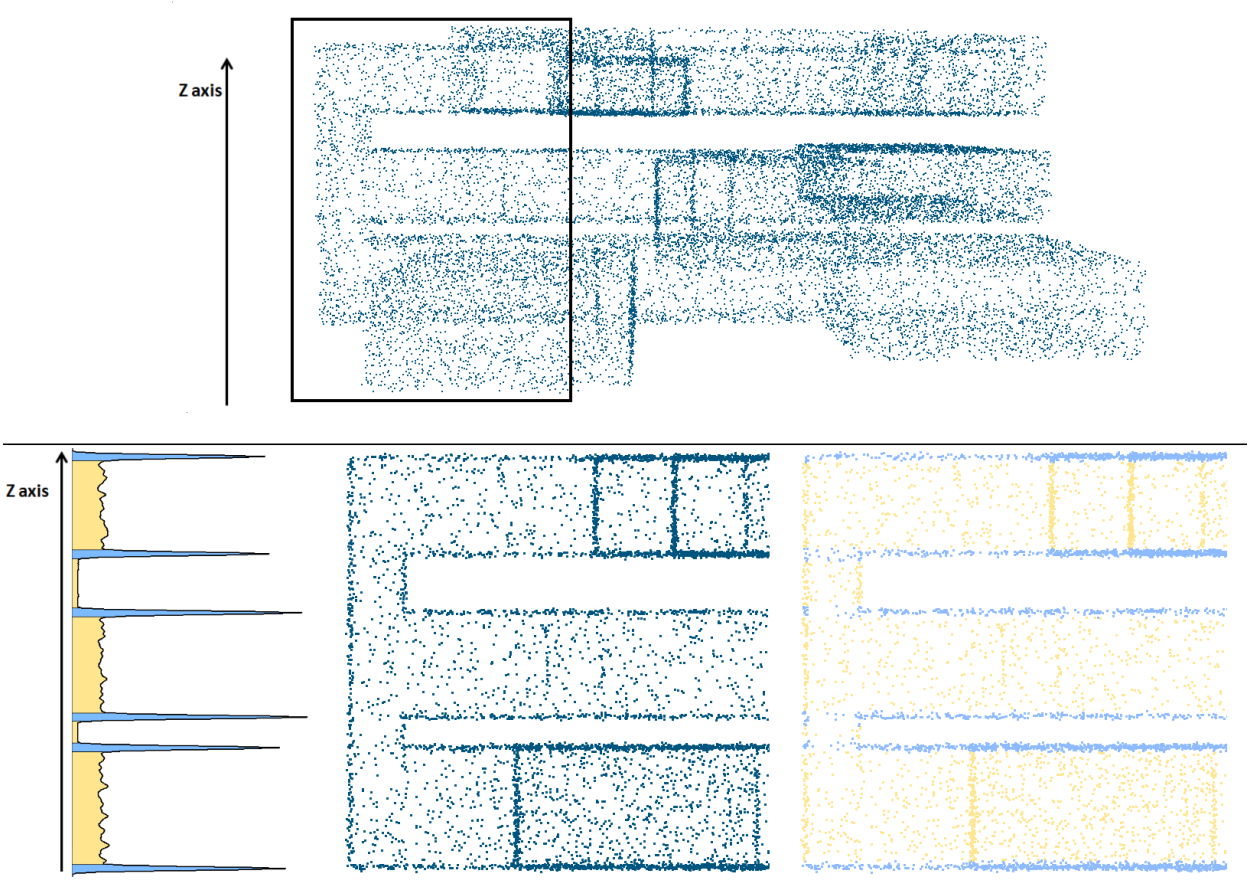

Figure 3: **Horizontal slicing applied to a synthetic scene.** Top: Input point cloud. The excerpt shown below is highlighted by a black box. Left: Distribution along vertical direction. The point cloud is split horizontally within a small range around the peaks. Ranges for horizontal structure-slices are depicted in blue, and yellow for wall-slices. Middle: Excerpt of original point cloud in side-view. Right: Side-view of excerpt colored by slice type.

shift (Cheng, 1995). Mean shift is an iterative algorithm similar to gradient descent for detecting maxima in density functions. For discretized spaces, mean shift is able to detect the interpolated maxima and yields a higher precision than the discretization of the underlying space. Before locating the peaks in the distribution via mean shift a Gaussian smoothing is applied. We use a flat kernel for mean shift. By choosing a kernel and bin size, close peaks in the distribution are inherently merged by mean shift. Each bin in the distribution is used as a starting point for the mean shift. We iterate until convergence or a maximum of 10 iterations has been reached. To ensure a minimum distance between two peaks, close peaks within a small distance h are clustered. We choose a default value of twice the bin size for h. The z coordinate of each maximum is denoted by $m_i, i \in \{1, .., N_m\}$.

The point cloud is now split at points around the peaks into horizontal structure-slices, containing the peaks and representing floor and ceiling, and into wall-slices, covering the remaining parts representing the walls. The split points are thus selected based on the gradient of the distribution. The number of points in the bins of the distribution is assumed to drop significantly around the

horizontal structures. The split points are located by walking through the bins in the distribution in both directions from the peak until the gradient is significantly smaller than at the first step adjacent to the peak. An example of selected split points is depicted by Figure 3.

If the split points are not located within a certain range around the peak, the peak is considered to be clutter and removed. As every local maximum in the distribution is detected, peaks that are not significant, i.e., that do not stand out to their neighborhood by a certain ratio, are removed. As close peaks are merged by mean shift and splitting points are chosen within a maximum range it is guaranteed that split points of adjacent peaks do not overlap. The type of slice, wall-slices or horizontal structure-slices, is thus alternating along the vertical direction. For the following steps only the wall-slices are considered. However, both types of slices are used for the graph-cut optimization used for model extraction.

2.2. Line extraction

The line extraction step aims at identifying wall segments from each wall-slice created in the previous step. The identification is split into three steps: (1-Filtering) the points are projected in the horizontal plane and clutter is filtered; (2-Line fitting) a multi-scale line fitting method is used to generate a line hypothesis for each point representing the local wall direction; (3-Clustering) the points are locally clustered into wall segments, and these segments are globally clustered into wall directions through a Hough transform.

1-Filtering: The detection of wall segments in indoor scenes is hampered by the presence of clutter: The clutter occludes the sight of the scanning device so that parts of the wall geometry are not sampled. When detecting permanent structures in the point cloud, clutter may be classified as permanent structure or vice versa. To avoid these cases, clutter is removed by filtering the points of the wall-slices by their normals. As walls are assumed to be vertical, points with a non-horizontal normal are filtered. More specifically, each point p_i in the wall-slices with normal n_i is removed when $|n_i \cdot n_z| < t$, where n_z is the unit vector along the vertical axis. A threshold of $t = 0.02$ corresponds to a deviation of a few degrees of the normal from the horizontal plane and provides a strict filtering of clutter while preserving samples on wall geometry, see Figure 4.

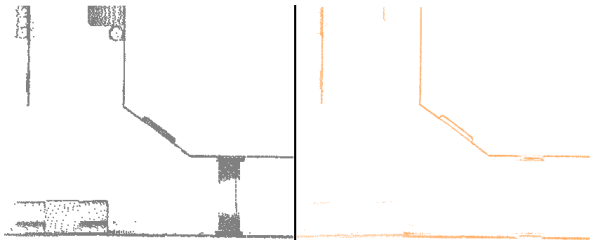

Figure 4: **Clutter removal.** Left: Wall-slice of a real data set showing a corridor scene in top-view. The scene contains clutter that challenges the wall detection. Right: Most of the clutter gets removed by filtering the points by their normals.

In order to gain robustness to missing data caused by occlusion and to reduce the problem to 2D, each wall-slice is projected vertically into the horizontal plane. Since walls are nearly vertical they get projected into line segments. Depending on the acquisition method the point distribution may exhibit a strong anisotropy. To remove anisotropy and speed up the following steps, we perform downsampling through an *occupancy grid* with uniform grid-size τ, provided as a parameter. In an occupancy grid each grid cell is labeled either empty or occupied. Downsampling is performed via replacing all points inside the cell by the averaged point location. This restricts the spatial point density to the grid size and approximates the original point coverage, see Figure 5. A small grid size τ allows for preservation of details in highly sampled areas, whereas a large grid size permits the reconstruction of sparsely sampled areas.

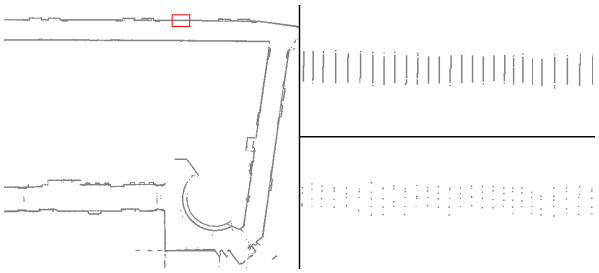

Figure 5: **Resampling for anisotropy removal.** Left: Overview of wall-slice of real data set. Upper right: Original point distribution of wall section in red box. Strong anisotropy is caused by the acquisition method of the scanning device. Lower right: After downsampling, the anisotropy is significantly reduced.

2-Line fitting: The wall direction at a point p_i may be estimated from the simplified projected points Q via local line fitting to a spatial neighborhood:

$$N_{p_i,r} := \{p \in Q : \|p - p_i\|_2 < r\}. \qquad (1)$$

However, there are unknowns: the local level of detail, ie., the wall geometry, and the scale of the neighborhood considered for fitting. For longer wall segments without details, the neighborhood for estimating the wall direction should be as large as possible to provide stable results in presence of noise. However, for points close to corners the size of the neighborhood must be small and therefore results in imprecise estimation, see Figure 6 (left). Furthermore, a corner has a higher level of detail as it cannot be described well with a single line.
In order to obtain a wall direction adaptive to the local scale and level of detail, several hypotheses are made at multiple scales. For each point the hypotheses are evaluated to find the one that best describes the local wall geometry. The idea behind the line fitting is that on a blueprint, walls compose a piecewise linear polyline separating the indoor space and the solid space, ie., permanent structures. Thus for each of the points there exists a small neighborhood in which the local wall configuration consists of either one straight wall or two walls in a corner. We exclude non-manifold configurations by assuming a non-zero wall thickness. For each scale of each point two hypotheses are proposed: a single line hypothesis to describe simple straight walls and a two-line hypothesis for detecting corner configurations.
The two-line hypothesis is created via RANSAC. Two pairs of points are randomly selected from the neighborhood and used to construct two lines l_1 and l_2.

The quality of the hypothesis is measured in the least-squares sense by:

$$S = \sum_{i}^{n} \min_{j \in \{1,2\}} dist(p_i, l_j)^2, \qquad (2)$$

where $dist(p, l)$ is the ℓ^2 distance between point p and line l. After a series of samplings, the pair of lines with the lowest sum of squared residuals S is selected as the best pair for the current scale.
As the local level of detail is not known several line hypotheses are established at multiple scales in order to select a level of detail. The scales, i.e., the ranges of the neighborhoods used for the hypotheses, are selected as multiples of the grid size τ. The minimum number of points in the neighborhood is set to five, as for a smaller number there always exists a perfect alignment of two lines. The smallest scale considered relevant is $N_{p_i,2\tau}$, as even smaller neighborhoods are unlikely to contain five or more points.
To find the proper line hypothesis for a point, the largest suitable scale for each type of hypothesis, single line and two lines, is first selected. This selection is performed by generating each type of hypothesis at increasing scales, starting with $N_{p_i,2\tau}$. If a hypothesis matches a certain quality criterion, the scale is increased by doubling the range $N^i_{p_i,r} \rightarrow N^{i+1}_{p_i,2r}$. The largest suitable scale is considered to be the largest scale before the quality criterion fails,

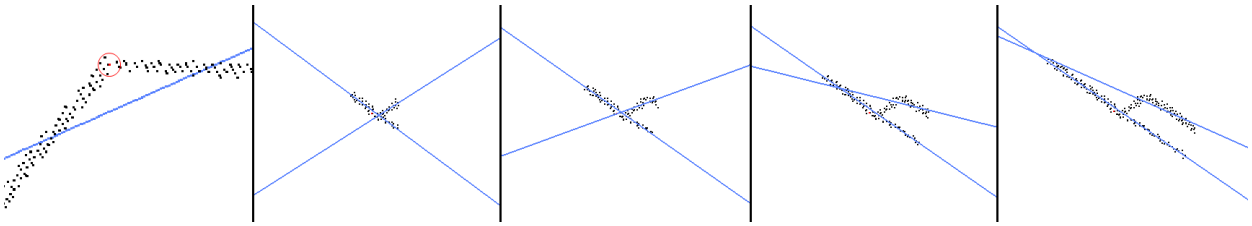

Figure 6: **Line hypotheses at different scales.** Left: Least-squares fitting of a line to the neighborhood of a point (red) located in a corner. Next four images: Fitting of two lines on four scales to a door frame depicted in the center of the scene shown by Figure 4. For the two smaller scales, see left and mid-left images, the neighborhood is well covered by the chosen lines. For larger scales however, see mid-right and right images, the geometry is too complex to be covered by two lines.

see Figure 6. The quality of the hypothesis is measured by the maximum Euclidean distance from a point p in $N_{p_i,r}$ to the closest line. A parameter ϵ is introduced to control the quality criterion:

$$\max_{p \in N_{p_i,r}} \left(\min_{j \in \{1,2\}} dist(p, l_j)^2 \right) < \epsilon^2, \qquad (3)$$

where ϵ denotes the tolerance distance between a point and the closest line. A high tolerance value deals with high noise in the input data and vice versa.

Among the two types of hypothesis (single or two lines), the hypothesis with largest scale is selected. When the two scales are equal the single line hypothesis is favored. There is one exception to this. If the two line hypothesis consists of two almost collinear lines, we consider the local wall configuration to be a single straight wall and discard the two line hypothesis.

Finally, a wall direction, i.e., a line l_i, is assigned to every point p_i by choosing the closest line of the selected hypothesis. The wall direction estimation process is illustrated by Figure 7 (left).

3-Clustering: For linear wall parts and corners the multi-scale line estimation of wall directions is satisfactory, but for small wall segments the assigned wall directions are inhomogeneous, see Figure 7 (middle). For robust clustering of points into wall segments, we adopt an idea from Huang et al. (2013) to sharpen and separate the assigned wall directions of the points. Through bilateral filtering, originally introduced for signal processing by Smith and Brady (1997), the points are classified around sharp features into disjoint clusters by their normals. We apply the bilateral filter as an iterative filter updating the wall direction of one point with the weighted average of the wall directions of the surrounding points. The scale s, selected during the previous step, is used for determining the neighborhood:

$$n_i' = \sum_{p_j \in N_{p_i,s}} n_j \, \theta(dist(p_i, l_j)) \, \psi(n_i, n_j). \qquad (4)$$

The normals n_i and n_j in the original formulation are replaced by the assigned wall directions. θ is the spatial weighting function, providing a high weight for points that are close to the chosen line l_i as they are likely to be part of the same wall segment. A Gaussian weighting function is chosen, parameterized by the ℓ^2 distance of the current point p_i to the line l_j fitted to neighboring point p_j:

$$\theta(d) = e^{-\frac{d^2}{\sigma^2_{spatial}}} \quad with \quad \sigma_{spatial} = 2\epsilon^2. \qquad (5)$$

We then define a similarity function ψ favoring points with similar wall direction. The similarity function from Huang et al. (2013) is adapted by using the wall directions instead of the nor-

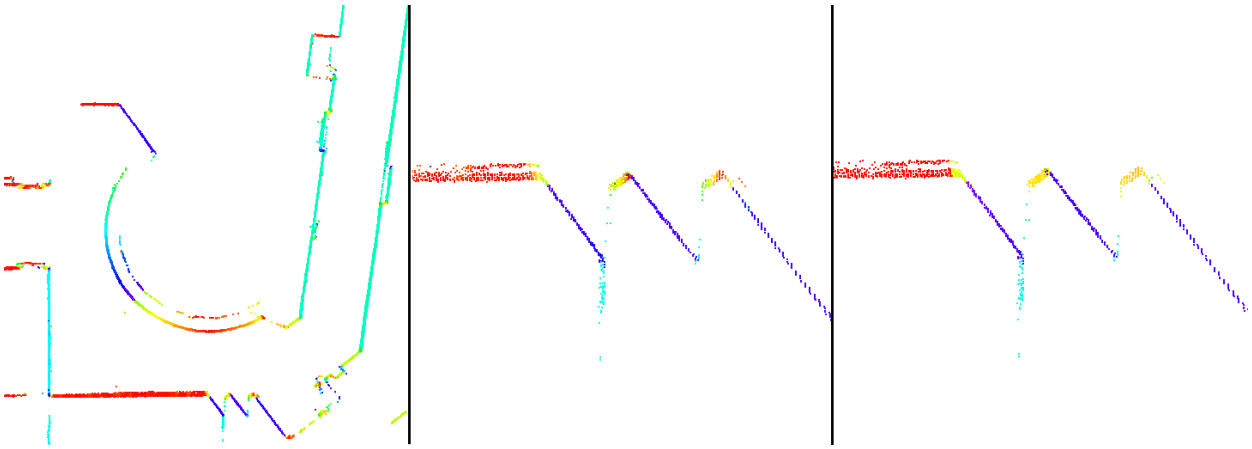

Figure 7: **Multi-scale line hypothesis.** Wall directions depicted by color in real data scene. Left: Excerpt of scene containing a corridor. The circular wall section is matched accurately. Middle: Some smaller segments have an unsharp diffuse fitting. Right: Segments are separated after filtering.

mal vectors:

$$\psi = e^{-\left(\frac{1-n_i^T n_j}{1-cos(\sigma)}\right)^2} \tag{6}$$

with σ set to 15°. We next extract the wall segments through region growing. The bilateral filtering and region growing clusters neighboring points into wall segments. As many wall segments are nearly collinear, we use a Hough transform for global clustering of the wall segments into lines, see Figure 8.

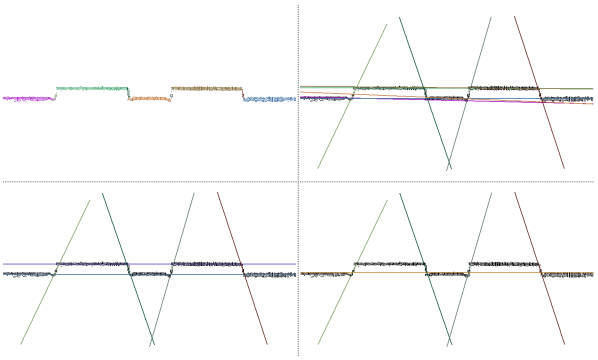

Figure 8: **Global clustering of wall segments.** Upper left: Clustered wall segments. Upper right: Lines extracted without global clustering. Lower left: Global clustering using Hough transform leads to reduction of the number of lines. Lower right: Over-simplification induced by coarse resolution for Hough Accumulator and a high value for ϵ.

The Hough transform (Davies, 2005) is used for robust 2D line extraction through accumulation and extraction of local maxima in discretized parameter spaces denoted as Hough Accumulator. In our framework the angle of each line with respect to the x axis and the distance between the line and the center of the bounding rectangle of the scene are chosen as parameters. Each cluster is added to the Hough Accumulator with a fitted least-squares line and a number of votes corresponding to the number of points in the cluster.
The main wall directions are so extracted one by one, starting with the global maximum in the Hough Accumulator. For each extracted maximum a line segment is fitted to cover the span of all points contained in the corresponding cell of the Hough Accumulator. All other clusters that match the line are removed from the Hough Accumulator and the next peak is detected. A cluster is considered matching a line when its points have a maximum distance ϵ to the line. This procedure is repeated until all clusters have been extracted from the Hough Accumulator.

2.3. Cell decomposition

We partition the bounding box by first splitting the horizontal cross section of the bounding box into single 2D cell decomposition and stacking copies of that 2D cell decomposition vertically to yield the 3D space partitioning. For constructing the 2D cell decomposition we use the line segments extracted from the Hough Accumulator. In architecture wall directions are often shared within and even across different levels of the building. To consolidate for wall segments that were not detected on all floors, we combine line segments from all levels to create the 2D cell decomposition. During combination of lines from different levels very similar ones are clustered as they represent the same wall direction.

We consider line segments with a certain extent to be a shared wall direction and extend them to lines crossing the complete bounding box. We then use the CGAL arrangement data structure (Agarwal and Sharir, 1998) for partitioning the horizontal plane by these lines into a 2D cell decomposition.

Line segments with a smaller extent are considered as local details and are only used to split cells in the decomposition locally. A default value for the minimum line extent of 1m has proven to be suitable.

The 3D space partitioning is then created by vertically stacking copies of the 2D cell decompositions, one for each wall-slice. Each copy is vertically extruded at the height of the corresponding peak in the distribution of the associated wall-slice, see Figure 9.

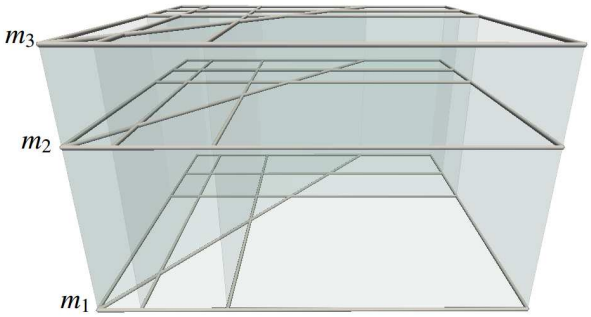

Figure 9: **3D space partitioning.** The 2D cell decomposition is extended in 3D through stacking and vertical extrusion. m_i denote the height of the detected peak during the horizontal slicing step.

3. Surface extraction

The final model is extracted from the 3D cell decomposition through labeling the cells as either empty or solid space. We formulated this binary labeling problem as a global energy minimization, solved through a graph-cut algorithm (Boykov et al., 2001). Such global minimization provides robustness against defect-laden data.

An undirected graph G is defined by a set of vertices V and a set of edges E:

$$G := (V, E) \quad with \quad E \in V^2. \tag{7}$$

Graph G is embedded into the 3D cell decomposition by associating each vertex to a volumetric cell. The edges of G connect all pairs of cells that share a vertical or horizontal face of the cell decomposition. We then use a graph-cut solver to minimize the following energy:

$$\min_{l \in \{0,1\}^{|V|}} \sum_{i \in V} D_i(l_i) + \alpha \sum_{(i,j) \in E} R_{i,j}(l_i, l_j), \tag{8}$$

where V denotes the vertices and E denotes the edges of G. The label l_i of a cell is either set to 0 for solid space or to 1 for empty space. D_i denotes the data term used for each label l_i assigned to the cell i in order to favor data faithfulness. $R_{i,j}$, referred to as the regularization term, represents a pairwise cost for connected cells. The parameter α is used to trade regularity for data faithfulness. The data and regularization terms are determined with the horizontal structure-slices and wall-slices.

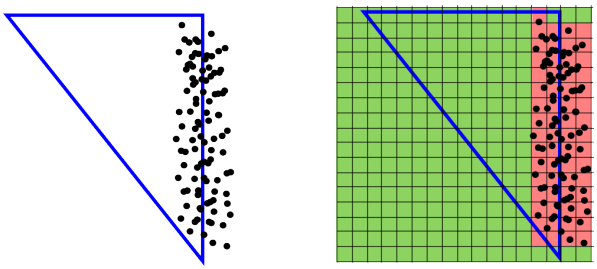

Figure 10: **Estimation of point coverage.** Left: Number of points in face is high, but coverage is low. Right: Usage of occupancy grid. Ratio of #occupied to #total grid cells provides approximation of coverage.

3.1. Regularization term

The regularization term $R_{i,j}$ is defined to favor a final model with low complexity. The penalty for different labels between adjacent cells is thus set to be proportional to the area of the shared face, see Eq.9. $A_{i,j}$ denotes the surface area of the shared face between cells i and j. For scale normalization the area is divided by the area of the horizontal cross section of the bounding box. As observed by Xiao and Furukawa (2012), approaches that penalize the surface area tend to miss thin details such as walls. To preserve thin structures, a weight is thus introduced in order to lower the cost of different labels between adjacent cells where permanent structures are expected. The weight is added to the regularization term without eliminating the favoring of a low complexity.

The weights between vertically adjacent cells lower the cost if horizontal structure is expected in the vicinity of the face. The presence of horizontal structure is estimated by the presence of points close to the face in the horizontal structure-slices. However, due to noise as well as imprecise cell alignment and varying scanning resolution, taking only the point density close to the face into account is ineffective, see Figure 10. Our experiments show that the coverage of the face provides a better solution. To evaluate the point coverage, an occupancy grid is used. The grid size is chosen as 2τ, where τ is the size of the occupancy grid filter used for downsampling during line extraction. The weight $\omega_{i,j}$ is then defined via the ratio of occupied grid cells to total grid cells.
The weights of horizontally adjacent cells are determined similarly. The shared face is vertically projected into an edge. The coverage of the edge by the points of the corresponding wall-slice is determined by discretizing the edge into bins of size 2τ. The weight $\omega_{i,j}$ is then defined as the ratio of number of occupied bins to total number of bins. This results into the following regularization term in cells i and j:

$$R_{i,j}(l_i, l_j) = \begin{cases} 0, & l_i = l_j \\ (1 - \omega_{i,j}) \cdot A_{i,j}, & l_i \neq l_j \end{cases} \quad (9)$$

However, for thin cells multiple faces might be close to the same points. As each face is generated by one specific line extracted during the line extraction step, only points of clusters fitted to this line are considered for that face. The cost for different labels might thus be zero only when the face or edge is fully covered. Such regularization term satisfies the condition of submodularity stated by (Kolmogorov and Zabih, 2004), see eqn. 10.

$$R_{i,j}(0,0) + R_{i,j}(1,1) \leq R_{i,j}(1,0) + R_{i,j}(0,1). \quad (10)$$

3.2. Data term

To provide a cost for each combination of cell and label, an estimation is done for each cell whether it belongs either to the empty or solid space. If the scanning device position is provided, this can easily be estimated by making use of the *free space* between each scanned point and its associated scanning origin. The problem of identifying inside and outside is recurrent for surface reconstruction (Lafarge and Alliez, 2013).
Our rationale is, that a ray cast from a point has an odd number of intersections with the geometry if the point is in empty space and an even number if it is in solid space. For counting intersections between a ray and the point cloud

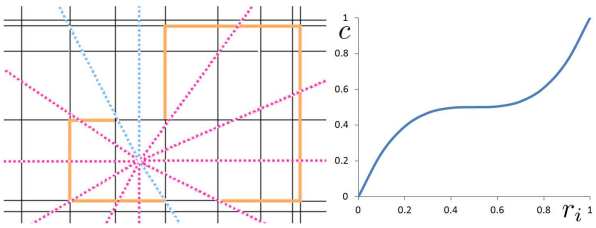

Figure 11: **Data term.** Left: Ray-casting to predict empty or solid label for a cell. The cell decomposition is depicted in black and edges with $\omega_{i,j} > \frac{1}{2}$ are depicted in orange. Rays indicating empty, i.e., odd number of intersections, are depicted in pink versus blue. Right: Inverse sigmoid function f turning ratio r_i of rays indicating empty space to total number of rays into c.

one solution would be to define a surface from the point cloud. However, clutter and missing data are likely to interfere and distort the number of intersections. The weights used for the regularization term are used as they are meant for indicating the presence of permanent structure. Rays are thus cast from each cell center and each intersection with a face with an assigned weight $\omega_{i,j} > \frac{1}{2}$ is counted as an intersection with a permanent structure, see Figure 11.

To improve stability at the cost of computation time, a higher number of rays can be send. The ratio of rays r_i for cell i is defined as the ratio of rays with an even number of intersections to the total number of rays. It is mapped with an inverse sigmoid function $f : [0, 1] \rightarrow [0, 1]$, see Figure 11. f is chosen so that a value of r_i is mapped conservatively to a cost c. Ratios not clearly indicating either empty or solid space are mapped to the same cost for both labels. In this way the label depends on the regularization term.
Due to the different sizes of cells, larger cells receive a higher penalty from the regularization term as they have a larger surface area. In order to eliminate this bias the cost of the data term is scaled by V_i. V_i is defined as the volume of the cell i that, for scale normalization, is divided by the volume of the bounding box. This leads to the final data term function:

$$D_i(l_i) = \begin{cases} c \cdot V_i, & l_i = 0 \\ (1 - c) \cdot V_i, & l_i = 1 \end{cases} \quad (11)$$

After the labeling is done, every cell is marked as either empty or solid space. The final 3D model is extracted as the set of faces between different labeled adjacent cells.

4. Experiments

We evaluate our method on a synthetic multi-level data set, two measurement data sets obtained by two types of laser scanners and one data set acquired with a Kinect sensor. The algorithm is implemented in C++ using the Computational Geometry Algorithms Library CGAL (2012). For energy minimization we use the Graph Cut Library (Boykov et al., 2001).

Cory 5th floor. In the measurement data set, [1] the point cloud is sampled on a hallway exhibiting non Manhattan-World geometry, including curved walls and archways. Some data are missing as several ways and rooms have partially been scanned without being entered. The scene also contains many fine details such as doors and tilted windows, and clutter such as couches and curtains. To reduce the memory footprint, the point cloud is downsampled by selecting every other point. The reconstructed model covers the structure of the hallway and partly captures details like doors, see Figure 12. The lintels of the doors are not sampled densely enough to appear as peaks during height analysis, and hence are not detected as horizontal structures. The height of the doors is extended to the next significant horizontal structure, here the ceiling, see lower right excerpts in Figure 12. The archways have not been reconstructed as our method is not suited to reconstruct surfaces that are non-planar or neither vertical nor horizontal. The circular wall

[1] `http://www-video.eecs.berkeley.edu/research/indoor/`

is approximated with planar surfaces and only exhibits minor artifacts although our method is not designed to detect and reconstruct non-planar walls, see upper right excerpts in Figure 12. The reconstruction in sparsely sampled areas, i.e., rooms and sideways, is incomplete. This is mostly due to the absence of walls, as it impacts both the space partitioning and the energy minimization.

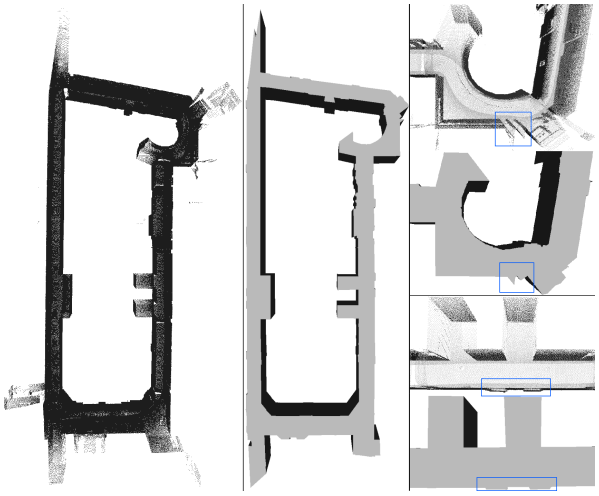

Figure 12: **Reconstruction of the Cory 5th floor.** Left: Input point cloud. Middle: Reconstruction of the indoor space. Upper right: Circular wall segment and details in the outer part. Points colored by estimated normals. Lower right: Another excerpt of the input point cloud, sampled on doors in the corridor and on an archway.

Inria Euler building. Using a Leica Scanstation P20, we performed an acquisition in the Euler building at Inria Sophia Antipolis. The data sets consists of 6 registered scans showing several adjacent rooms including the entrance, a conference room, a lecture room and a small corridor. It features walls scanned from both sides, clutter like chairs and tables as well as vegetation captured outside. The input data and the reconstructed model are depicted by Figure 13. The thin walls have been reconstructed, while the clutter outside does not interfere with the reconstruction.

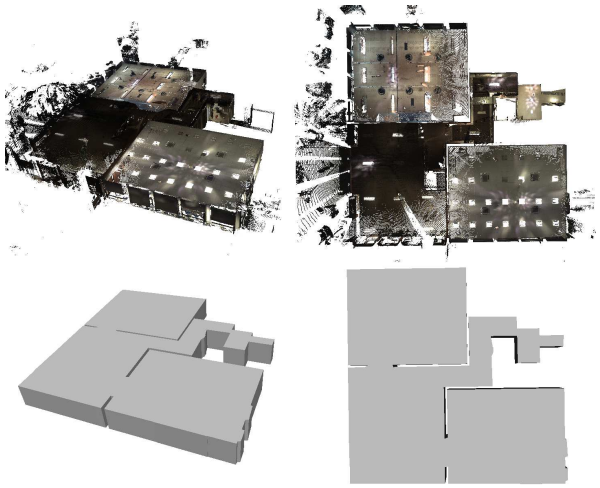

Figure 13: **Reconstruction of the Euler building entry area.** Top row: Input point cloud in perspective and top view. Bottom row: Reconstructed model in perspective and top view.

Synthetic data set. For evaluation we use as ground truth a synthetic data set created using Google Sketchup. We then use Meshlab for sampling a point cloud from the exported triangle mesh. A small amount of uniform noise (5 mm) is added after sampling. The ground truth model features a multi-floor, non Manhattan-World scene. The reconstructed model shows that different wall directions are reconstructed accurately, see Figure 14. Reconstruction of non-orthogonal wall directions is depicted by the upper two close-ups. The reconstruction of different height levels is shown by the lower two close-ups. The mesh is colored by the Hausdorff distance from the result to the ground truth showing a maximum distance of 2.3 cm. A small deviation in the reconstructed wall directions was caused by global clustering in the Hough transform space. Such merging incurs a small deviation visible in the two red colored corners of the reconstructed model. The chosen parameters and timings are listed in Table 1.

Kinect data set. Point clouds acquired by the consumer-grade Kinect sensor are challenging our algorithm as they contain many defects

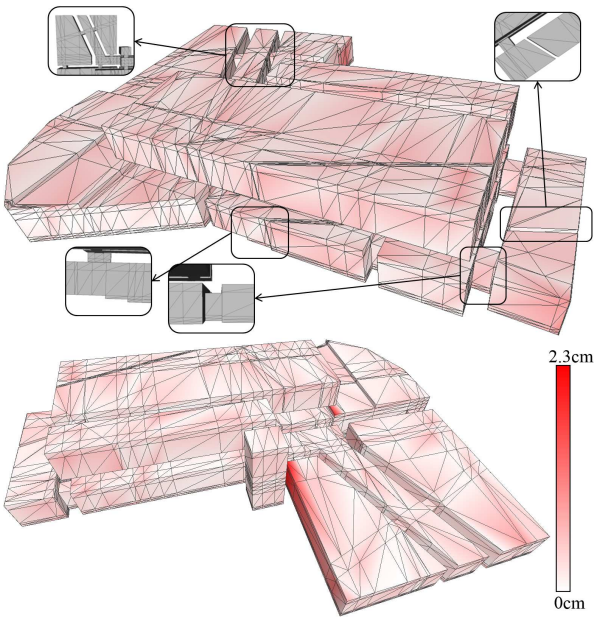

Figure 14: **Reconstruction from a synthetic data set.** Two views on the same reconstruction are shown, colored by the Hausdorff distance from the result to the ground truth.

(Khoshelham and Elberink, 2012). The scene exhibits a single office room containing clutter such as a desk, some cupboards and a bicycle. As the scan covers only a small part of the floor and no ceiling at all, no significant peaks show up in the horizontal distribution and the scene is processed as a single level. Compared to commercial-grade laser scanners the high level of noise requires a high tolerance value ($\epsilon = 6$ cm) for clustering of the wall directions. The input data and reconstructed model are shown by Figure 15. Most of the room is correctly reconstructed except a corner entirely occluded by a bicycle and a cupboard. The corridor could not be reconstructed due to very sparse sampling.

Comparison with Oesau et al. (2013). We evaluated the improvement over previous work using the *Cory 5th floor*, as it contains different level of details, varying from a large corridor to small niches and a circular wall segment. While both methods are able to reconstruct the main corridor, our new method, reinforced by a more accurate line detection strategy, exhibits a much higher level of detail. The circular wall is reconstructed with many small linear segments with our method, whereas the former method only provides a rough approximation. Results from both methods are shown by Figure 16.

4.1. Parameters

The parameters have a direct impact on the quality and level of detail of the reconstructed model. The bin size for the horizontal slicing affects the detection of horizontal structures. It should be larger than the typical amount of noise, but still allow for several bins between horizontal structures in the height histogram. In general a value of 5-10 cm is suggested. However, for scenes containing just ceiling and floor the method is tolerant to the choice of the bin size.
τ relates to the point density in the scanned data and is used for the downsampling. A high value improves the reconstruction of sparsely sampled regions, while removing detail from highly sampled regions due to the downsampling. Technically, τ should be chosen as the

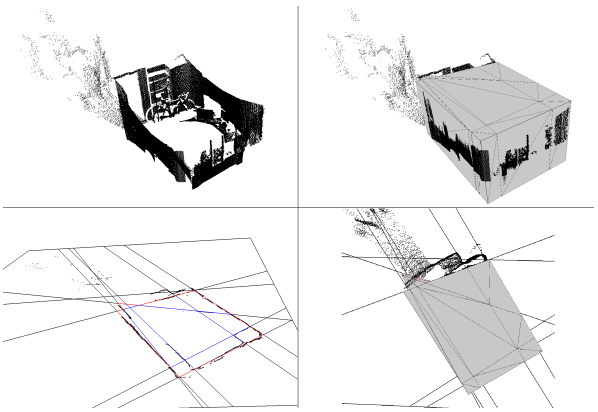

Figure 15: **Reconstruction of Kinect-recorded indoor scene.** Upper left: Input point cloud featuring an office room. Upper & lower right: Reconstructed model. Lower left: The labeled cell decomposition exhibits a good alignment of the room with the exception of the occluded corner.

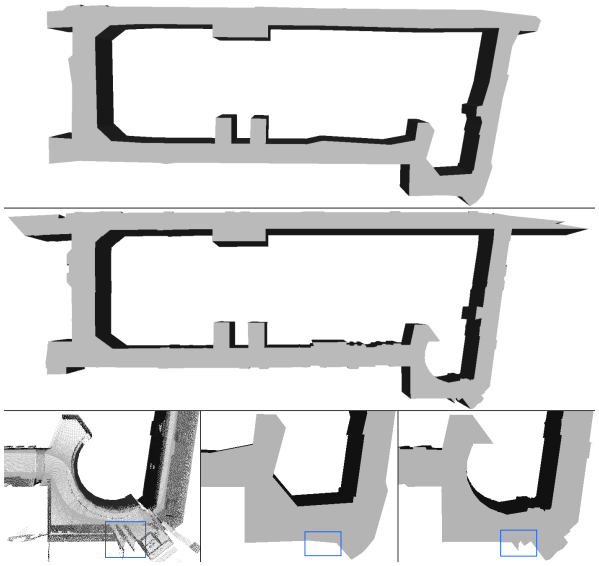

Figure 16: **Comparison with Oesau et al. (2013).** Top: Reconstructed model from previous work (Oesau et al., 2013). Middle: Result from the presented model. Bottom: An excerpt of the circular area is shown. In contrast to the original method many more details are recovered.

minimal sampling resolution, i.e., largest distance between neighbored points in the area of interest. In usual cases this parameter will be chosen in a range between 0.75 and 4 cm.

ϵ is used for the multi-scale line fitting and clustering of extracted wall directions. It indirectly relates to the number of cells generated in the space partitioning. High values of ϵ lead to reconstructed models with low levels of detail (Figure 17), in short computation times, especially for model extraction. A value of at least twice the range of scanner noise is recommended.

The Hough transform is used to cluster the detected wall segments. The effect of the chosen resolution is depicted in Figure 8. As the clustering is also restricted by the choice of ϵ, a coarser resolution of the Hough Accumulator can be used as a starting point for parameterization: $2^\circ \cdot \tau$.

α is used to trade data faithfulness for regularity in the energy minimization formulation. For densely scanned data a low value of α enforces faithfulness to the input data. For incomplete data a high value of α provides regularity by filling gaps but may over-simplify the model (Figure 17). Note that α is only used for the last step, i.e., labeling. An adjustment and extraction of a new result requires only few seconds of computation time.

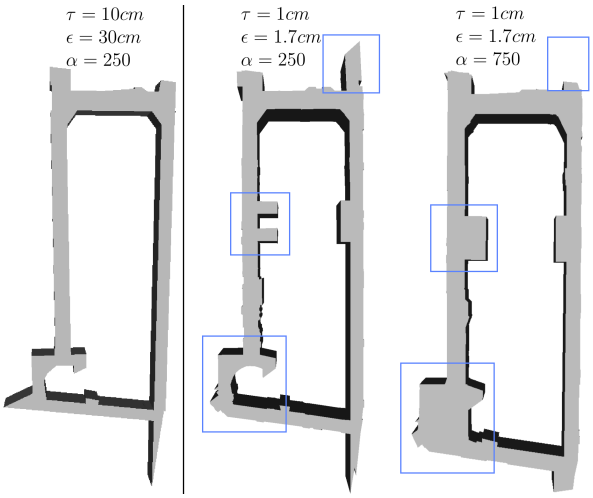

Figure 17: **Impact of parameters.** Left: Reconstruction of the Cory 5th floor with $\tau = 10\,cm$ and $\epsilon = 30\,cm$. Such low level of detail requires 7.2 s. Middle & right: Respectively 250 and 750 for α. The higher regularization $\alpha = 750$ leads to a lower level of detail through area minimization.

4.2. Accuracy

We evaluate the accuracy of our method by comparing the reconstructed model to the synthetic multi-level data set used as ground truth. Figure 14 illustrates that different wall directions are reconstructed accurately: the symmetric Hausdorff distance is 2.6 cm and the one-sided Hausdorff distance from the result to the ground truth is 2.3 cm (Figure 14).

4.3. Performances

Running times and parameters are provided in Table 1. Timings are measured on an Intel Core i7 920 with 16 GBs RAM. The time spent to estimate the normals is omitted. The most time consuming steps of the algorithm

are the stochastic ray-casting to determine the data term and the multi-scale line fitting. A kD-Trees is used to speed up the calculation of the edge weights for the regularization term (see section *Regularization Term*).

4.4. Robustness

We first evaluate the robustness of our approach against sparse sampling with downsampled instances of the *Cory 5th floor* (50%, 20%, 10% and 5% of the original point cloud, generated by selecting every 2nd, 5th, 10th or 20th point of the original point cloud). The parameters set for reconstruction are adapted to each downsampled version. As the point density is lowered by downsampling the parameter τ used for adjusting the grid size during line extraction is increased. The tolerance ϵ used for clustering is also increased, as line fitting is less accurate on sparse point clouds. While small values for τ and ϵ allows for a detailed reconstruction of high resolution scans the method tends to generate more bumpy artifacts. Higher values of τ and ϵ may miss details, but provide lower computation times and more regular reconstructions. Timings and parameters for the lower resolution data sets are provided in Table 2.

The reconstructions from the downsampled instances are shown in Figure 18. The reconstruction of the 50% data set provides the same amount of details as the original data set. In the lower resolution data sets the connected rooms and sideways are imprecisely recovered or not at all, as they are even sparsely scanned in the original mesh. The recovery of details for the lower resolution data sets is depicted in the right column. Note that the doors and niches in the corridor are extended to the height of the ceiling.

Robustness to noise is evaluated by adding noise to the *Cory 5th floor* data set. More specifically, each point of the input point cloud is displaced along a random vector, with maximum length set to 5 mm, 25 mm and 50 mm.

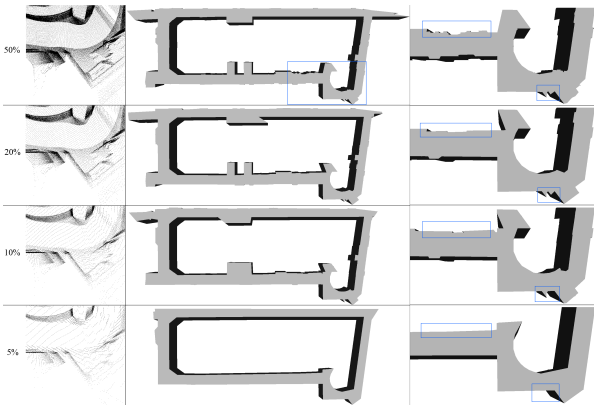

Figure 18: **Reconstruction at lower resolutions.** Each row of the image shows the reconstruction of one downsampled data set (from top to bottom): 50%, 20%, 10% and 5%. Left column: Excerpt of the input data. Left-mid column: Overview over the reconstructed model. Right column: Close-ups showing amount of detail recovered.

The noise is thus added to the noise of the measurement device, estimated between 3 and 5 mm for common laser scanners. Figure 19 illustrates that noise impacts the level of detail of the reconstruction without altering the coarse structures. A small amount of noise (5mm, the usual range of laser scanners) has minor impact on the small details. With higher amount of noise less details are reconstructed but the coarse structure of the architecture is correctly recovered, except for sparsely scanned areas such as the outer rooms and the two hallways. As the accuracy of the line fitting process is impacted by noise, the parameters τ and ϵ are adjusted to the level of noise. The bin size for the vertical distribution (0.1), $\alpha = 250$ and the resolution for the Hough Accumulator $1^\circ \cdot 0.8\,cm$ have not been changed.

Robustness to outliers is evaluated by adding random points uniformly distributed within the bounding box of the *Cory 5th floor* data set: respectively 1% and 5% outliers. The reconstruction of those modified data sets are shown in Figure 20. The corridor is reconstructed correctly despite the presence of outliers in both data sets. However, the amount of recovered details is lowered and the circu-

	Synthetic	*Cory 5th floor*	*Euler*	Kinect
#points	1,000,000	4,391,604	2,305,938	86,694
Vertical distribution - bin size	1 cm	8 cm	8 cm	x
τ	2 cm	1 cm	0.75 cm	1 cm
ϵ	2.5 cm	1.7 cm	1 cm	6 cm
Hough res: $angle \cdot distance$	$1^\circ \cdot 0.25\,cm$	$1^\circ \cdot 0.8\,cm$	$1^\circ \cdot 0.8\,cm$	$3^\circ \cdot 5\,cm$
α	20	250	250	250
Spatial partitioning	63 s	192 s	36 s	6.9 s
Model extraction	58 s	206 s	66 s	15.2 s

Table 1: **Running times.** Chosen parameters and running times (single thread). Timings for model extraction include the ray-casting performed to compute the data term.

	50%	20%	10%	5%
#points	4.391.604	1.756.642	878.321	439.161
τ	1 cm	2 cm	2 cm	4 cm
ϵ	1.7 cm	3cm	3.5 cm	5 cm
Spatial partitioning	192 s	70 s	58,8 s	16.8 s
Model extraction	206 s	68,4 s	58,5 s	12.1 s

Table 2: **Parameters and running times of lower resolution data sets.** The bin size for the vertical distribution (0.1), $\alpha = 250$ and the resolution for the Hough Accumulator $1^\circ \cdot 0.8cm$ do not vary.

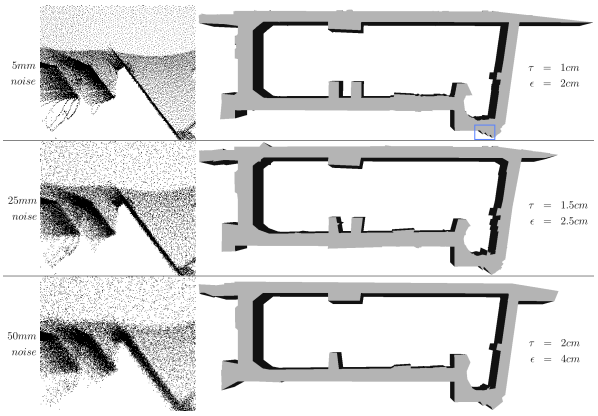

Figure 19: **Robustness to noise.** Reconstruction of the *Cory 5th floor* with added noise. Left: Closeups on the outlined rectangle. Right: Reconstructions and corresponding parameters.

lar area exhibits minor geometric inaccuracies. The horizontal slicing step shows to be effective for removing outliers. Points are filtered by their normals on separation into wall-slices and horizontal structure-slices and, therefore, a major part of the outliers are removed. Some outliers close to the original points remain and show some impact on the precision of the reconstruction, mostly in sparsely scanned areas or small details.

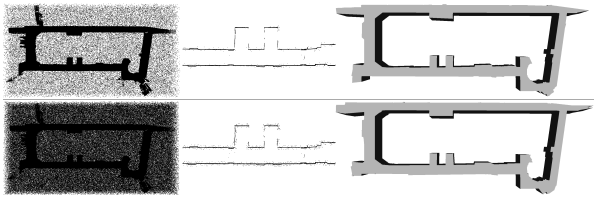

Figure 20: **Robustness to outliers.** Reconstruction of the *Cory 5th floor* with added outliers. Top: 1% outliers. Bottom: 5% outliers. Left: Point cloud with outliers. Middle: Wall-slice generated by horizontal slicing with only some added points close to the walls remaining. Right: Reconstructed model of the indoor space.

4.5. Limitations

Our algorithm is designed to handle arbitrary vertical wall directions. For cases of partially scanned rooms, where some walls are not sampled at all, our method relies on other wall directions detected in the scene or on the borders of partially scanned parts.

These cases may result in improper data consolidation, while algorithms strictly restricted to Manhattan-World scenes often yield more plausible reconstructions. Note, however, that other approaches fail on these cases Jenke et al. (2009); Xiao and Furukawa (2012).

Global clustering in a Hough transform space performs a regularization as the Hough Accumulator determines the main wall directions and aligns the extracted lines to these directions. We notice, however, that such regularization may hamper the reconstruction of fine details (Figure 12) and requires parameter adjustments. The latter trial-and-error process is time consuming and suggests to investigate automatic parameter selection.

The algorithmic complexity of the ray casting is quadratic in the number of cells of the space decomposition, this number being itself related to the level of detail adjusted through a tolerant or selective line clustering. The computation times thus increase rapidly with the level of detail sough after.

Locating the splitting heights for the point cloud (Section 2.1) depends on the amount of points detected on the horizontal structures. For large scenes such horizontal structures require a very large number of sample points. The door lintels in the *Cory 5th floor*, for instance, are sparsely sampled and therefore the height of doors is extended to the ceiling (Figure 12, lower right).

Finally, some buildings contain non-vertical walls, non-horizontal floors and ceilings, stepped floors and ceilings, and even non-planar structures. One of such examples with stepped floors and non-horizontal ceilings is the auditorium of the Delft University of Technology (Figure 21). Our approach fails on such cases with stepped floors and large amount of clutter.

5. Conclusion

We proposed a new method for indoor scene reconstruction. Through detecting the permanent structures in a cluttered scene we reconstruct a model of the indoor space with satisfactory tradeoff between accuracy and low complexity. Our first technical contribution is a multi-scale, feature-preserving approach for detecting walls as line segments, followed by global clustering in a Hough transform space in order to align the reconstructed model with the permanent structures. Through global energy minimization we label the cells of a 3D space partitioning in order to reconstruct a watertight model consolidating missing data. Our experiments show that our method is moderately robust to noise and outliers, and generates satisfactory results from data measured with Kinect sensors.

As future work, we will investigate the use of non-planar primitives and non-vertical walls in order to both improve accuracy and decrease complexity by simplifying the space partitioning. We will also further investigate regularization through the Hough transform. Detecting and reinforcing canonical geometric relationships (parallel, collinear, orthogonal, coangular) and structural regularities may improve robustness and provide semantic information about the structuring of indoor spaces.

6. Acknowledgments

The authors wish to thank Astrium for funding this work. Pierre Alliez is supported by an ERC Starting Grant "Robust Geometry Processing" (257474). We also thank the *Video and Processing Lab* of *University of California, Berkeley* and Prof. Sisi Zlatanova from the *Delft University of Technology* for making the data sets available.

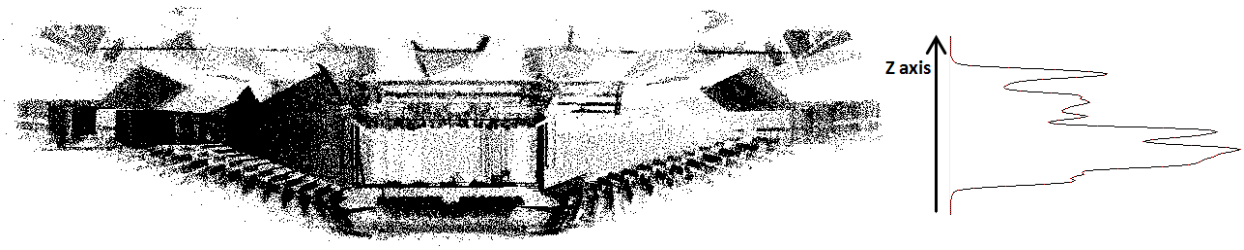

Figure 21: **Failure case with stepped floors.** Side-view of the auditorium at the Delft University of Technology. The vertical distribution is shown right. Due to the clutter the lowest horizontal part of the floor is not distinguishable from the stepped floor and clutter.

References

Adan, A., Huber, D., 2011. 3D reconstruction of interior wall surfaces under occlusion and clutter, in: International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, IEEE Computer Society, Los Alamitos, CA, USA. pp. 275–281.

Agarwal, P.K., Sharir, M., 1998. Arrangements and their applications, in: Handbook of Computational Geometry, Elsevier Science Publishers B.V. North-Holland. pp. 49–119.

Boulc'h, A., Marlet, R., 2012. Fast and robust normal estimation for point clouds with sharp features. Computer Graphics Forum 31, 1765–1774.

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 1222–1239.

Budroni, A., Boehm, J., . Toward automatic reconstruction of interiors from laser data, in: Proceedings of the 3rd ISPRS International Workshop 3D-ARCH 2009, IAPRS, Vol. 38, ISPRS.

Budroni, A., Boehm, J., 2010. Automated 3D reconstruction of interiors from point clouds. International Journal of Architectural Computing 8, 55–73.

Cgal, 2012. Computational Geometry Algorithms Library. Http://www.cgal.org.

Cheng, Y., 1995. Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 17, 790–799.

Davies, E.R., 2005. Computer and Machine Vision: Theory, Algorithms, Practicalities. Morgan Kaufmann Publishers Inc.

Edelsbrunner, H., Kirkpatrick, D., Seidel, R., 2006. On the shape of a set of points in the plane. IEEE Transactions on Information Theory 29, 551–559.

Huang, H., Wu, S., Gong, M., Cohen-Or, D., Ascher, U., Zhang, H., 2013. Edge-aware point set resampling. ACM Transactions on Graphics 32, 9:1–9:12.

Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., Fitzgibbon, A., 2011. Kinectfusion: Real-time 3D reconstruction and interaction using a moving depth camera, in: ACM Symposium on User Interface Software and Technology, ACM, New York, NY, USA. pp. 559–568.

Jenke, P., Huhle, B., Strasser, W., 2009. Statistical reconstruction of indoor scenes, in: International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, pp. 17–24.

Khoshelham, K., Elberink, S.O., 2012. Accuracy and resolution of kinect depth data for indoor mapping applications. Sensors 12, 1437–1454.

Kim, Y.M., Mitra, N.J., Yan, D.M., Guibas, L., 2012. Acquiring 3D indoor environments with variability and repetition. ACM Transactions on Graphics 31, 138:1–138:11.

Kolmogorov, V., Zabih, R., 2004. What energy functions can be minimized via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 65–81.

Lafarge, F., Alliez, P., 2013. Surface reconstruction through point set structuring. Computer Graphics Forum 32, 225–234.

Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohli, P., Shotton, J., Hodges, S., Fitzgibbon, A., 2011. Kinectfusion: Real-time dense surface mapping and tracking, in: Symposium on Mixed and Augmented Reality, IEEE Computer Society, Washington, DC, USA. pp. 127–136.

Oesau, S., Lafarge, F., Alliez, P., 2013. Indoor scene reconstruction using primitive-driven space partitioning and graph-cut, in: Eurographics Workshop on Urban Data Modelling and Visualisation, Eurographics Association. pp. 9–12.

Okorn, B.E., Xiong, X., Akinci, B., Huber, D., 2010. Toward automated modeling of floor plans, in: Symposium on 3D Data Processing, Visualization and Transmission.

Sanchez, V., Zakhor, A., 2012. Planar 3D modeling of building interiors from point cloud data, in: International Conference on Image Processing, IEEE Signal Processing Society, Piscataway, NJ, USA. pp.

1777–1780.

Schnabel, R., Wahl, R., Klein, R., 2007. Efficient RANSAC for point-cloud shape detection. Computer Graphics Forum 26, 214–226.

Shao, T., Xu, W., Zhou, K., Wang, J., Li, D., Guo, B., 2012. An interactive approach to semantic modeling of indoor scenes with an RGBD camera. ACM Transactions on Graphics 31, 136:1–136:11.

Smith, S.M., Brady, J.M., 1997. Susan - a new approach to low level image processing. International Journal of Computer Vision 23, 45–78.

Xiao, J., Furukawa, Y., 2012. Reconstructing the world's museums, in: European Conference on Computer Vision, Springer-Verlag. pp. 668–681.