Indoor Scene Reconstruction using Primitive-driven Space Partitioning and Graph-cut

Sven Oesau, Florent Lafarge, Pierre Alliez

Indoor Scene Reconstruction using Primitive-driven Space Partitioning and Graph-cut

Sven Oesau, Florent Lafarge and Pierre Alliez

Inria Sophia Antipolis - Méditerranée, France

Abstract

We present a method for automatic reconstruction of permanent structures of indoor scenes, such as walls, floors and ceilings, from raw point clouds acquired by laser scanners. Our approach employs graph-cut to solve an inside/outside labeling of a space decomposition. To allow for an accurate reconstruction the space decomposition is aligned with permanent structures. A Hough Transform is applied for extracting the wall directions while allowing a flexible reconstruction of scenes. The graph-cut formulation takes into account data consistency through an inside/outside prediction for the cells of the space decomposition by stochastic ray casting, while favoring low geometric complexity of the model. Our experiments produces watertight reconstructed models of multi-level buildings and complex scenes.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling; Boundary representations; Curve, surface, solid, and object representations—

1. Introduction

In the recent years reconstruction of architectural scenes has got more and more attention. Compared to the reconstruction of outdoor environments, indoor scene modeling is still in a quite early stage. Indoor scenes impose different challenges than outdoor environments. Buildings 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 usually a high amount of clutter. Additional challenges are posed by varying point density and anisotropy.

Related work - Sanchez et al. [SZ12] proposed a method for modeling multi-level building interiors. They assume a Manhattan World scene for segmentation and employ RANSAC for reconstruction. Additionally they employ model fitting of a staircase model to unsegmented points. The method results in a polygon soup of permanent structures, but there is no structure information such as connected rooms. Budroni et al. [BB10] apply sweeping techniques to identify Manhattan World directions and to locate wall segments. The floor plane is decomposed by those detected segments and the ground plan is reconstructed based on the point density on the floor. The result is a watertight model extracted from the ground plan. Xiao et al. [XF12] introduce a *constructive solid geometry* (CSG) based method to reconstruct large-scale indoor environments for visualization purposes. They propose a greedy algorithm for creating a CSG model guided by an objective function to measure quality and control detail. Their method is not restricted to Manhattan World geometry, but their CSG method requires orthogonal or parallel structures as it fits rectangles.

Jenke et al. [JHS09] propose a greedy algorithm fitting cuboids to primitives detected by RANSAC. Due to the fitting of cuboids their approach requires rectangular geometry. The cuboid fitting requires at least five planar primitives, hence may fail in presence of clutter or missing data. Other work focus on semantization of indoor scenes [KMYG12,SXZ*12] that concentrate on understanding clutter in indoor environments.

Contributions - We propose a new method to reconstruct a volumetric model of the indoor space to overcome the aforementioned restrictions. Our method consists of three steps. First, *horizontal slicing* horizontal permanent structures at different altimetric levels are detected from the point cloud while separating the point cloud into *horizontal structure slices* and *wall slices*. Second, wall directions are extracted to create a *3D space partitioning* driven by a line procedure. Third, the *volumetric cells* are labeled into empty space, i.e., inside, and occupied space, respectively for permanent structure or outside.

- **Primitive-driven space partition:** We do not restrict our model to the Manhattan World geometry and instead are able to detect walls in any directions. We assume that floor and ceiling are horizontal and that walls are vertical. For the extraction of wall directions we use the Hough Transform. The Hough Accumulator inherently offers alignment of nearly collinear wall segments. In addition, it is suitable for applying further regularization such as parallelism or co-angularity.

- **Watertight model handling missing data:** A *3D space partitioning* into *volumetric cells* is created by using detected wall directions. By limiting the possible results with the *3D space partitioning* a consolidation of missing data based on the wall directions is enforced. The *volumetric cells* are labeled as either empty or occupied by energy minimization. This provides robustness to missing data and outliers.
- **Multi-level indoor scenes:** Our approach is able to reconstruct a single scene containing a whole building with multiple levels in one optimization, without a priori knowledge about the presence of levels.
- **Applicable to raw data:** Many datasets provide only raw information such as point locations. To be as general and applicable as possible our method relies only on dense point data and additional knowledge about vertical direction.

2. Proposed method

The proposed method is composed of three successive steps.

Horizontal Slicing - In this step we horizontally split the point cloud to separate horizontal structures, e.g., floor and ceiling components. For detection a histogram of the point density along the vertical direction is created, see Figure 1. To remove the clutter from the histogram we filter out points whose normals are not parallel to the vertical direction. If normals are not available, they are estimated through local principal component analysis (PCA).
The peaks in the histogram m_i are located by using mean shift. The point cloud is then separated into *horizontal structure-slices* enclosing the peaks and containing mostly important horizontal structures. The space between peaks is referred to as *wall-slices* containing vertical structures and clutter, see Figure 1.

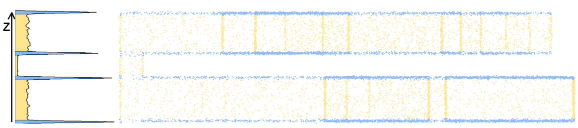

Figure 1: *Horizontal slicing applied to a synthetic scene.* ***Left:*** *Histogram along vertical direction. The point cloud is split by horizontal planes when the gradient strongly flattens within a small range around the detected peaks in the histogram. Ranges for horizontal structure-slices are depicted in blue, and yellow for wall-slices.* ***Right:*** *Side-view of point cloud colored by slice type.*

Primitive-driven space partitioning - The extraction of wall directions is done in 2D in the horizontal plane as it is more tolerant to missing data and eases the filtering of clutter. Each *wall-slice* is thus projected vertically into the horizontal plane. Since walls are vertically oriented they get projected into lines. To remove anisotropy and speed up the following steps a spatial sensitive down-sampling using an occupancy map with a uniform grid-size τ is applied.
Before employing a Hough Transformation for extracting wall directions, points on clutter have to be separated from points on walls. As walls are assumed to be piecewise linear a PCA is applied to the 2τ neighborhood of each point to identify linear parts of the boundary. Wall directions are obtained as lines from each boundary point by linear least-squares-fitting to the local neighboring boundary points. The lines are added to the Hough Accumulator as angle and signed distance to the projected bounding box center.
The representation of wall directions in parameter space allows an easy detection of parallel walls and offers the possibility to apply regularization to the scene. Mean-shift is used to extract lines from the Hough Accumulator. As walls in architecture are often collinear within and even across levels, we combine detected wall directions of all *wall-slices*. Before splitting the horizontal plane into a *2D cell decomposition*, very similar wall directions are clustered by prefering the one with higher support in the Hough Accumulator. The *3D space partitioning* is created by vertically stacking the *2D cell decomposition*, one for each *wall-slice*. Each *2D cell decomposition* is vertically extended to the heights of the enclosing peaks in the histogram of the associated *wall-slice*, see Figure 2.

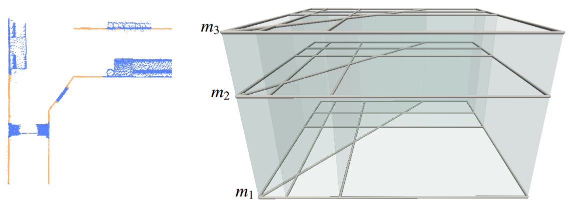

Figure 2: ***Left:*** *Wall-slice of real data set after border classification* $\frac{\lambda_2}{\lambda_1} < 0.1$ *using* $\lambda_1 > \lambda_2$ *the eigenvalues from PCA. Clutter (blue) like the couches and the trash bin have been sorted out. Walls (orange) have been correctly detected.* ***Right:*** *Turning the 2D cell decomposition into 3D space decomposition by stacking. Each wall-slice is represented by one 2D cell decomposition. The height of each 2D cell decomposition is extended to the enclosing peak heights* m_i.

Labeling - The binary labeling of the *volumetric cells* is formulated as an energy minimization and solved via *graph-cut* [BVZ01]. The graph is embedded into the *3D space partitioning* by associating a node to every *volumetric cell* and connecting nodes of vertically or horizontally adjacent *volumetric cells*. *Graph-cut* can be used to solve energy minimizations of the form:

$$\min_{l \in \{0,1\}^N} \sum_{i \in C} D_i(l_i) + \alpha \sum_{(i,j) \in E} V_{i,j}(l_i, l_j),$$

where C is the *volumetric cells* set and E is the set of adjacent cell pairs. D_i is the data term that provides a cost for each label l_i assigned to the cell i to favor data faithfulness. $V_{i,j}$ is referred to as the regularization term and provides a pairwise cost for connected cells. The α parameter is used to trade regularity for data faithfulness.

Regularization term - We choose the regularization term to favor low complexity. The cost γ is defined as the surface area. For scale normalization the area is divided by the horizontal area of the bounding box. As Xiao and Furukawa [XF12] observed however, approaches that penalize surface area tend to miss thin details like walls. To preserve thin structures we introduce a weight to lower the cost of different labels at adjacent cells where permanent structures are expected: A weight is added to the regularization term without eliminating the preference for simple models.

To determine the weights between vertically connected cells the points in the *horizontal structure-slices* are vertically projected onto the shared face. The weight $\omega_{i,j}$ is then defined via an occupancy map with size 2τ as the ratio of occupied segments to total segments.

For weighting edges between horizontally connected cells the shared face is vertically projected into an edge. The coverage of the edge by the boundary 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 #occupied bins / #total bins. This results into the following regularization term:

$$V_{i,j}(l_i,l_j) = \begin{cases} 0, & l_i = l_j \\ (1-\omega_{i,j})\cdot\gamma, & l_i \neq l_j \end{cases}$$

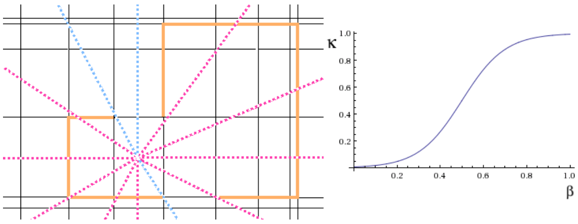

Figure 3: ***Left:*** *Schematic view of the ray-casting to predict the empty or occupied label for a cell. The cell decomposition is depicted in black and edges with $\omega_{i,j} > \frac{1}{2}$ are depicted in orange. Rays indicating empty, i.e. , odd number of intersections, are depicted in pink versus blue.* ***Right:*** *Mapping sigmoid function turning the ratio κ of rays indicating empty space to total number of rays into β.*

Data term - To provide a cost for each combination of cell and label, we must estimate for each cell whether it belongs to the empty or occupied space. The problem of identifying inside and outside is recurrent for surface reconstruction [LA13].

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 occupied space. For counting intersections between a ray and the point cloud one solution would be to define a surface from the point cloud. However, clutter and missing data are likely to interfere and falsify the number of intersections. We thus make use of the weights used for the regularization term as they are meant for indicating the presence of permanent structure. Rays are 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.

To improve stability, we rely on sending a higher number of rays. The ratio of rays κ with an even number of intersections to the total number of rays is mapped to β with a sigmoid function, see Figure 3. 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 ρ. ρ is defined as the volume of the cell that, for scale normalization, is divided by the volume of the bounding box. This leads to the final data term function:

$$D_i(l_i) = \begin{cases} (1-\beta)\cdot\rho, & l_i = 0 \\ \beta\cdot\rho, & l_i = 1 \end{cases}$$

3. Experiments

We evaluate our method on both a synthetic multi-level data set and on a real data set. The synthetic model has been created with a mesh-based editing software and turned into a point cloud by random sampling. Our algorithm has been implemented using C++ and the CGAL library [CG12]. For energy minimization we use the Graph-Cut library [BVZ01].

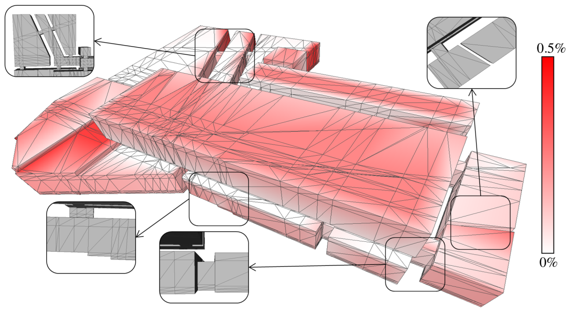

Figure 4: *Reconstructed 3D model from a synthetic data set. The colors depict the Hausdorff distance from the reconstruction to the ground truth. The error is uniformly distributed over the model and rather low,* 0.4975% *w.r.t. the bounding box diagonal.* ***Close-ups:*** *The two upper close-ups show reconstructed parts that feature non orthogonal wall directions. The upper left close up shows accurate reconstruction of thin walls. Sensitivity to variation in floor and ceiling height is illustrated by the two lower close-ups.*

Accuracy - To evaluate our method we created a synthetic multi-level data set that includes non-orthogonal wall directions as well as different heights of floor and ceiling on the same level. The result shows that different wall directions have been reconstructed accurately, see Figure 4.

For evaluating the accuracy of our method we compare the reconstructed 3D model with the ground-truth mesh that has been sampled to generate the point cloud. The symmetric Hausdorff distance is 0.5515% w.r.t. the bounding box diagonal, whereas the Hausdorff distance from the result to the ground truth is 0.4975%, see Figure 4.

We further apply our method to a real data set available at http://www-video.eecs.berkeley.edu/research/indoor/. The scan shows a hallway exhibiting non Manhattan World geometry including curved walls and archways. Several ways and rooms have been partially scanned without being entered so that the data set also features missing data. The scene contains many details such as doors and tilted windows, as well as clutter such as couches and a curtain. The result is shown by Figure 5.

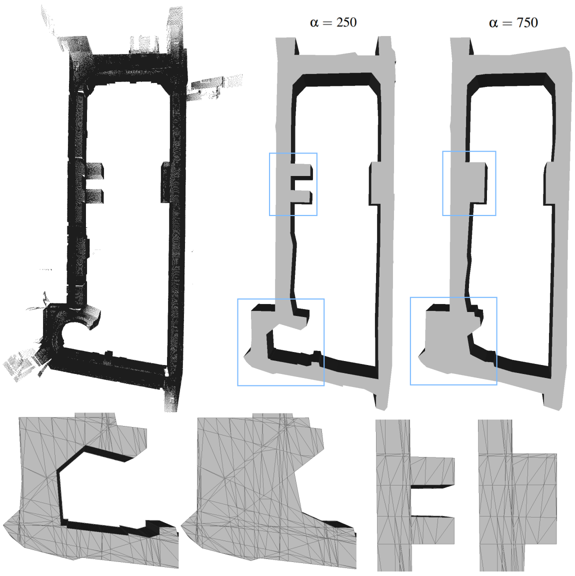

Figure 5: ***Upper left:*** *Original point cloud of a real data set.* ***Upper middle & right:*** *Two α values used to trade regularity for data faithfulness. Middle: $\alpha = 250$. Right: $\alpha = 750$ leading to high regularity.* ***Lower row:*** *With a high α value for regularization the circular area is closed and the two small doorways are merged. Although our method is devised to detect only linear wall segments it provides a rough reconstruction of the circular part.*

Performance - Running times and parameters are provided in Table 1. Time measurements were performed on an Intel Core i7 920 with 16 GBs RAM (the time spent to estimate the normals being omitted). The most time consuming parts of the algorithm are the PCA to identify the boundary points and the computation of the weights for the regularization term. kD-Trees have been used both for determining the neighborhoods during PCA and for computing the weights used for regularization.

Limitations - The wall segment extraction is not feature sensitive and therefore may miss small details such as windows. Additionally, our method is restricted to planar walls and requires horizontal floor and ceiling as well as vertical walls. This assumption may not be suitable to atypical or free-form architectural buildings. Figure 5 illustrates how we provide a rough piecewise linear approximation for curved walls. Finally, the method is highly dependent on the piecewise detection of wall segments as the possible solutions are restricted to the detected walls.

	synthetic	5th floor
#points	1.000.000	1.756.642
τ	2.5	0.08
Hough res: *angle* $\cdot$ *distance*	$1^\circ \cdot 0.25cm$	$1^\circ \cdot 0.8cm$
α	20	250
Spatial partitioning	4.1s	22s
Model extraction	9.9s	6.4s

Table 1: *Choosen parameters and running time for single threaded execution.*

4. Conclusion

We have presented a novel scene reconstruction method for interior environments. Our approach first detects permanent structures. We then create a robust 3D space decomposition well aligned with the detected permanent structures, and employ *graph-cut* to label the cells of the 3D space decomposition, in order to find a plausible solution that best fits the detected geometry. The use of the Hough Transform allows us to extract arbitrary wall directions, but the reconstruction of circular or curved wall segments is not satisfactory. Future work includes a more feature sensitive creation of the cell decomposition and the enforcement of global regularities in the model.

Acknowledgments: The authors wish to thank ASTRIUM for funding. Pierre Alliez is funded by an ERC Starting Grant 'Robust Geometry Processing', agreement 257474. We also want to thank the *Video and Processing Lab* of *UC Berkeley* for making the data set available.

References

[BB10] Budroni A., Boehm J.: Automated 3D Reconstruction of Interiors from Point Clouds. *International Journal of Architectural Computing 8* (2010), 55–73. 1

[BVZ01] Boykov Y., Veksler O., Zabih R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. PAMI 23*, 11 (Nov. 2001), 1222–1239. 2, 3

[CG12] CGAL, Computational Geometry Algorithms Library, 2012. http://www.cgal.org. 3

[JHS09] Jenke P., Huhle B., Strasser W.: Statistical reconstruction of indoor scenes. In *Proc. WSCG '09* (2009). 1

[KMYG12] Kim Y. M., Mitra N. J., Yan D.-M., Guibas L.: Acquiring 3D indoor environments with variability and repetition. *ACM Trans. Graph. 31*, 6 (2012). 1

[LA13] Lafarge F., Alliez P.: Surface reconstruction through point set structuring. In *Proc. of Eurographics* (2013). 3

[SXZ*12] Shao T., Xu W., Zhou K., Wang J., Li D., Guo B.: An interactive approach to semantic modeling of indoor scenes with an RGBD camera. *ACM Trans. Graph. 31*, 6 (Nov. 2012), 136:1–136:11. 1

[SZ12] Sanchez V., Zakhor A.: Planar 3D modeling of building interiors from point cloud data. In *Proc. of the ICIP (2012)* (2012). 1

[XF12] Xiao J., Furukawa Y.: Reconstructing the world's museums. In *Proc. of the 12th ECCV (2012)* (2012). 1, 3