# ROOM SEGMENTATION IN 3D POINT CLOUDS USING ANISOTROPIC POTENTIAL FIELDS

*Anonymous ICME submission*


## ABSTRACT

Automatic and robust partitioning of indoor 3D point clouds into rooms is a central requirement for emerging applications such as indoor navigation or facility management. Existing works are either based on the Manhattan-world assumption or rely on the availability of the scanner pose information. Instead, we follow the architectural definition of a room and consider it as an inner free space separated from other spaces through openings or partitions. For this we formulate an anisotropic potential field for 3D environments and illustrate how it can be used for room segmentation in the proposed segmentation pipeline. The experimental results confirm that our method outperforms state-of-the-art methods on a number of datasets including those that violate the Manhattan-world assumption.


***Index Terms***— Room segmentation, indoor reconstruction, point cloud, unsupervised clustering

## 1. INTRODUCTION

Indoor reconstruction is becoming an increasingly important topic because of the need for automatically generated semantic models of buildings from 3D data. Potential applications include architecture, civil engineering, facility management, indoor mapping and navigation. Prior to further processing it is important to partition the data into semantically meaningful parts, which is normally done by segmenting building point cloud data into rooms. This task is, however, made difficult by numerous factors, such as clutter, occlusion and large volume of data. Previous work has partially addressed these problems. In robotics, room segmentation has been mostly done in occupancy grid maps [1], [2], which are heavily influenced by clutter and occlusion present in the indoor environment. In computer vision and graphics, the segmentation methods work on 3D point cloud data directly, but they suffer from a number of limitations. The first limitation is the assumption of precise knowledge of the sensor poses [3], [4], [5]. Such information is highly specific to the used acquisition sensor, therefore it is difficult to combine data from multiple information sources (e.g. including RGBD and LIDAR scanners). It is also often the case that already available CAD models are combined with partial scans of the environment. The second limitation includes a strong assumption on Manhattan-world structure [6], [7], [5], [8]. Clearly, this is not true for a general indoor environment exhibiting curved walls and tilted ceilings.

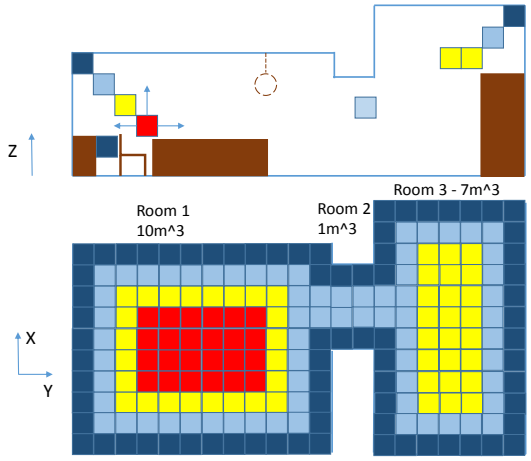



**Fig. 1**: Illustration of rooms as inner free spaces separated by smaller openings. Top: side view of an indoor environment with two rooms separated with a smaller room (corridor). Furniture is shown in brown, several free voxels with potential field (PF) values are also shown. Bottom: top-down view with proposed anisotropic PF maximum values along the vertical stack. Red color corresponds to high PF values and dark blue - to low.

To tackle this challenge, we review the definition of a room. Due to absence of a strict mathematical formulation, we refer to architectural context [9]. Here rooms are enclosures or divisions separated from other divisions by partitions. In other words, rooms are bigger (in volume) free spaces that are connected to each other through a smaller (in volume) free space, such as a door or an arch (see top part in Fig. 1). This formulation follows the human understanding of a room having a certain homogenous spatial volumetric signature within its boundaries. To address this formulation, we present a way to compute an anisotropic potential field for free space in 3D that is robust to clutter and occlusion. We further show how such representation can be used for room segmentation in a general indoor scene. Our framework is evaluated qualitatively and quantitatively with real and synthetic data of multiple buildings.

Contributions of this paper are as follows:

1. Framework to compute interior free space without assuming knowledge of scanner poses or the Manhattan-world structure of indoor environments.
2. 3D formulation of anisotropic potential field computation for free space that is robust to clutter and occlusion.
3. Room segmentation pipeline that makes no assumptions on the room layout.

## 2. RELATED WORK

Related work on room segmentation can be separated into several areas.

**Robotics**. Most of the previous work in the area of robotics deals with room segmentation in 2D data, in particular with occupancy grid maps captured with a robotic platform [1], [2], [10]. However, it has been shown that the presence of clutter and objects in the environment significantly deteriorates the performance of segmentation approaches, such as Voronoi maps or distance transform [1]. Furthermore, the approaches of [2] and [10] require information on the sensor trajectory as well as a significant amount of training data. This limits the performance of such approaches for general scenes.

**Computer vision and graphics**. In the area of computer vision and graphics, 3D room segmentation has been addressed by [6], [7], [5], [8], [3], [4], [11]. Several of these approaches assume geometric regularity and absence of occlusion [6], which is rarely the case for an indoor environment. Other methods assume a Manhattan-world structure of the building [7], [5], [8]. But in reality this assumption is often violated for buildings with tilted ceilings and curved walls. A number of approaches further require the scan poses for every point in the point cloud [3], [4], [5], [8]. In practice this information can be difficult to obtain, because the 3D data is often combined from multiple sources, such as RGBD and LIDAR scanners or CAD models. To the best of our knowledge, the previous work has not fully addressed the definition of a room as an enclosed free space within an indoor environment that possesses a certain spatial signature.

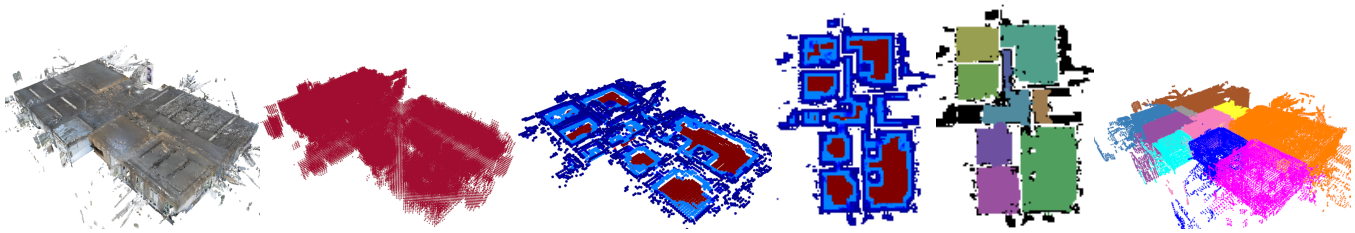
## 3. METHODOLOGY

An overview of the proposed method is given in Fig. 2. We operate on a 3D point cloud (a) and start with detecting interior free space (b). We further proceed with computing a 3D anisotropic potential field (PF) for free voxels. Afterwards, we perform maxima detection in the PF values of each vertical voxel stack (c) and store the maximum value into a 2D PF map (d). Given the PF image, we perform clustering using information about the PF values as well as the visibility between voxels (e). Finally, we map the labeled free space back to the 3D point cloud (f). As input data we use 3D point cloud data acquired using either RGBD sensors [5], [7] or LIDAR scanners [3]. We do not use RGB information for any of the algorithms as geometry is sufficient for room segmentation.

### 3.1. Interior free space classification

**Free vs. busy space classification**. As a room encompasses free space, we first need to recognize free space as compared to busy space occupied by objects and architectural parts. For this, we voxelize the entire space spanned by the bounding box of the point cloud data. Each voxel that contains at least one point will be labeled as busy, and free otherwise. Hence, the voxels corresponding to furniture and other indoor objects will initially be labeled as busy even though they represent the inner volume of the room. As our goal is to reconstruct the inner volume of rooms and compute its volumetric signature, we are interested in labeling such voxels as free. [12] has proposed an approach to classify voxels into free and busy using volumetric graphcut. Unfortunately, this method requires camera poses and such information is often not available. Therefore, we proceed differently and first apply binary 3D morphological operations along the vertical direction operating on the grid occupancy (see top part in Fig. 1). In particular, we identify isolated busy voxels surrounded by free voxels and subsequently label them as free. In essence, we detect the following pattern along the vertical direction ”busy”-”free”-”busy”-”free”-”busy”. The central busy voxels matching such pattern will be labeled as free. Clearly, such voxel operations on large-scale datasets can result in high computational complexity, therefore we choose a relatively large voxel size, e.g. with a side length of 18cm. In order to further reduce complexity, we precompute the 3D coordinates of every voxel and use a lookup table for neighbor search.

**Interior vs. exterior free space classification.** Now, for the identified free space we need to classify it into interior (inside the building) and exterior (outside the building). To identify interior space, [12] has proposed to check if the free space is enclosed by busy space using visibility and to further formulate it as a Markov Random Field (MRF) problem, which can be efficiently solved using graphcuts. Unfortunately, performing such visibility checks in 3D for large-scale point clouds would result in a prohibitively high computational complexity. Instead, we check if this free voxel is placed between two busy voxels (so called enclosing). To keep computational complexity low, we do not check for enclosing in all directions in 3D, but instead leverage the properties of large-scale point cloud datasets having limited and small number of dominant directions. Thus, we perform enclosing checks for every free voxel only along the main basis directions of the indoor environment. As we cannot assume orthogonal axis-aligned environments, we estimate the main directions in XYZ space using the Mixture of Manhattan Frames algorithm [13]. It is essentially a formulation of K-means clustering on a hypersphere that allows to estimate



(a) Input point cloud (b) Interior free space (c) Maxima of 3D PF (d) 2D PF (e) Labeled image (f) Labeled point cloud

**Fig. 2**: Overview of the method.

dominant directions. It has an advantage as compared to the principal component analysis, as it allows to estimate more than three principal directions. Thus, we can efficiently precompute directions along which we need to check for neighboring busy voxels instead of performing costly geometric checks in 3D space. Clearly, the level of occlusion is different in various parts of the indoor space due to the scanning procedure. It has been commonly observed that upper parts of the environment (e.g. ceiling and elevated parts) are less likely to be occluded during the scanning procedure as compared to the floor and lower parts of the environment [14]. Therefore, when accumulating evidence for a free voxel to be exterior, we choose different weights for various directions of enclosing. In particular, evidence is computed as follows:

$$E(v) = w_1 \cdot \mathbb{1}(z-) + w_2 \cdot \mathbb{1}(z+) + w_3 \cdot \mathbb{1}(z) + \\ + w_4 \cdot \mathbb{1}(dom_1) + w_5 \cdot \mathbb{1}(dom_2), \quad (1)$$

where $\mathbb{1}(z-)$ is 1 in case this free voxel $v$ has busy voxels below along the z-axis, $\mathbb{1}(z+)$ is 1 in case there are busy voxels above along the z-axis, $\mathbb{1}(z)$ is 1 in case there are busy voxels above and below, $\mathbb{1}(dom_1)$ is 1 in case there are busy voxels along both directions of the first dominant direction $dom_1$, which has been estimated using the Mixture of Manhattan Frames method and typically lies in the horizontal plane. Similarly, $dom_2$ refers to the second dominant direction that also lies in the horizontal plane. Here, we choose higher weight for the case of busy voxels located below as such voxels indicate higher probability of this space being interior: $w_1 = 0.43$. In contrast, the remaining weights are set to $w_2 = w_3 = w_4 = w_5 = 0.1425$. Please note that even though the number of dominant directions could be greater than two, we have observed that two dominant directions suffice for the majority of the considered datasets. We use $E(v)$ as the data term for a MRF formulation in combination with the smoothness term $E(l) = 0.6$. This value has been experimentally verified to obtain a proper regularization of the indoor voxels, while still accurately following the computed free space evidence. We further build a 6-neighborhood connected graph spanning all free voxels. We then compute the graphcut using the Boykov-Kolmogorov mincut algorithm [15] in order to find interior free voxels.

### 3.2. Anisotropic potential field computation

Once the interior free space has been detected (see (b) in Fig. 2), we can proceed with the room segmentation. So far many voxels within the room volume have been labeled as busy due to the presence of furniture and other objects. One could separate indoor objects from architectural elements of the building, but this remains a challenging problem in indoor reconstruction [3], [4], [16]. Instead, we leverage the observation that PF-based approaches for path planning and room segmentation have shown a good performance in the past for 2D scenarios [1]. The PF value of the free voxel is normally defined as its distance to the closest busy voxel. Unfortunately, a straightforward formulation of PF in 3D would result in significant variations in its values due to clutter and indoor objects, which have little in common with room boundaries (see table and chair in top part of Fig. 1). Therefore, we instead propose to perform nearest neighbor search in the half-space spanning the positive z-direction. This way, every voxel stores the L2-distance to the closest busy voxel lying in the half-space spanning positive z values, so called anisotropic potential field value.

Given the 3D PF map (see Fig. 2 (c)), one could formulate clustering as an MRF problem with the PF gradient as a data term, thus enforcing smoothness. However, we have observed that the maximum PF value along the vertical stack of voxels is sufficient to detect disconnected components of the rooms. This has the further advantages of low computational complexity and ability to provide a simple visualization (see lower part in Fig. 1). The resulting maximum values are now stored in a 2D image, which is used for further processing (see (d) in Fig. 2). In order to enhance robustness of the method, we further perform visibility checks between voxels and store 1 in case the other voxel is visible from this one, and 0 otherwise. The other voxel is visible if there are no busy voxels in the ray direction starting from the first voxel and terminating in the second voxel. Here, instead of performing visibility checks for every voxel, we only do so for the highest free voxel along the vertical stack. We have observed that due to varying ceiling profile the visibility of the highest voxel carries more information for space partitioning as compared to including the voxel with the maximum PF value.

### 3.3. Clustering

Now, given the 2D PF map, we need to identify discontinuities, as these indicate room boundaries. One could employ Voronoi graphs combined with merging heuristics [1], but we have observed that this would impose constraints on the room layout and shape. Common methods for segmentation such as spectral clustering, k-means and random walks suffer from various limitations, such as clusters having similar number of points or convex shapes. Similarly, graphcut algorithms suffer from erroneous merging of smaller rooms into the neighboring bigger ones. In contrast, density-based clustering algorithms (e.g. DBSCAN) have an advantage as they do not assume any specific cluster shape, but instead perform region growing based on density. A further advantage of such algorithms is that they can use a general distance metric, thus being able to incorporate other distance measures besides Euclidian space. However, DBSCAN is very sensitive to the chosen value of the neighborhood radius. Therefore, we instead choose its extension called HDBSCAN [17] that employs a new cluster stability measure in order to maximize the overall stability of selected clusters.

Prior to clustering, we perform local maxima detection in the 2D PF map in order to first find rooms of larger size. To derive a threshold, we build a histogram of intensity values and detect the density peak with largest value. Now we define the distance matrix for the voxels as follows:

$$D = D_{vis} \cdot w_{vis} + D_{eucl} \cdot w_{eucl} + D_{PF} \cdot w_{PF}, \quad (2)$$

where $D_{eucl}$ is the euclidian distance between two voxel coordinates. $D_{PF}$ is the difference of the PF values of two voxels. $D_{vis}$ is the distance between voxels $p$ and $q$ based on visibility computed as normalized hamming distance between their visibility vectors [5]:

$$D_{vis}(p,q) = \frac{dist(V(p), V(q))_{hamming}}{\sum_i V_i(p) + \sum_i V_i(q)}, \quad (3)$$

where $V(p)$ is the visibility vector of voxel $p$, such that $V_i(p) = 1$, if voxel $i$ is visible from voxel $p$, and 0 otherwise. In order to take into account information on the PF difference as well as visibility change within different parts of the environment we choose the parameters as follows: $w_{vis} = 0.3$, $w_{eucl} = 0.6$, $w_{PF} = 0.1$.

After clustering, there are a number of unlabeled points, which remain after the thresholding operation. To label them, we threshold the remaining unlabeled points to obtain local maxima. The threshold is the lowest detected density maximum in the intensity histogram. Afterwards, we perform a number of merging operations. For this, we use the previously introduced distance in Eq. (2) to select the closest cluster for merging for every point. In case the distance to the closest cluster is too high, such points are assigned to a new cluster.

### 3.4. Mapping of free space labeling to busy space

Given the labeled 2D map (see Fig. 2), we need to propagate the labeling onto 3D busy voxels. For this, we start with the labeled voxel and propagate labeling onto unlabeled voxels in the stack along the vertical direction. Afterwards, for each busy voxel we find its nearest 10 neighboring free voxels with labels. The most often occuring label of the labeled voxels will indicate the labeling of the considered busy voxel.

## 4. EXPERIMENTAL RESULTS

### 4.1. Evaluation

We evaluate our approach using several datasets. In particular, we use the large-scale dataset of [7] with labeled rooms spanning 4 buildings and counting in total 175 rooms. For qualitative evaluation we further include the unlabeled dataset of [5]. Finally, we also verify the performance on the unlabeled laser scanner dataset that violates the Manhattan-world assumption as described in [3], i.e. exhibiting tilted ceilings and curved walls. We do not vary any of the parameters of our algorithm across the scenes.

The experimental results using our method along with the method of [7] are given in Fig. 3. The quantitative evaluation for this dataset is given in Table 1. Here due to unavailability of the source code or the labeled data of the algorithm from [7] we cannot use the Adjusted Rand Index (ARI) metric as employed in [7], because this would result in the inconsistent evaluation. In particular, because ARI operates on points, it is biased towards larger rooms, e.g. it inadequately measures incorrect labeling of smaller rooms. Therefore, we instead count how many rooms have been erroneously labeled (shown as red ellipses in Fig. 3). Such metric also allows for a meaningful evaluation and can adequately represent algorithm segmentation performance. One can observe in Fig. 3 that the approach in [7] does not perform well for the rooms that are not aligned with main walls of the building. See an example of this in the top part of Area 1, top part of Area 2 and right part of Area 3. Furthermore, smaller rooms are often erroneously merged into the neighboring bigger ones - see top left and bottom parts in Area 1 and middle part of Area 2. In contrast, our approach does not assume the Manhattan-world structure therefore is able to label such rooms correctly. It still incorrectly labels several rooms due to irregularities in the PF map. We want to point out that our evaluation is rather conservative, because the dataset labeling [7] is inconsistent across different buildings: in some buildings the corridor is labeled in parts, while in others it is labeled as a whole.

We also show the results of room segmentation for a number of buildings that exhibit tilted ceilings and curved walls, thus violating the Manhattan-world assumption - see Fig. 4. One can see that our approach is able to peform on-par with the method of [3], even though we make significantly less assumptions (e.g. no scanner poses or do not rely on presence

**Table 1**: Room segmentation results on the dataset of [7]. Numbers in two right columns show the number of incorrectly labeled rooms. Lower values indicate better segmentation performance.

| Area | Number of rooms | Our | Armeni [7] |
|---|---|---|---|
| 1 | 44 | **2** | 8 |
| 2 | 40 | **10** | 12 |
| 3 | 23 | **5** | 7 |
| 5 | 68 | **7** | 13 |
| Total | 175 | **24** | 40 |

of planar regions).

We further show results for room segmentation on the dataset of [5] in Fig. 5. One can observe that our approach outperforms [5] in several places: we detect the room in the top-right part of Office 1 and in the right part of Office 2. This is in spite of the fact that we do not use the information about scanner poses as compared to [5]. Furthermore, we are able to segment parts of the outdoor space. The inability of [5] to segment these correctly is due to the merging heuristics of free space voxels which result in erroneous merging of two rooms. We want to note that our approach does not perform very well in the parts of the point cloud, where data is very sparse. such as the left part of Apartment 1 or left part of Office 2 in Fig. 5. In such cases, the algorithm of [5] heavily relies on the scanner pose information to discard such voxels prior to segmentation.

### 4.2. Discussion

We have observed that for most environments the PF information is the most important feature for clustering. Nonetheless, in certain cases, such as transitions between corridors (like in the middle part of Area 1 in Fig. 3), it is important to include visibility so as to detect changes of space signature. We further acknowledge that in some cases, as in the case of detecting a long rectangular-shaped corridor, it can be disadvantageous to use visibility for clustering. In contrast, PF values become more important for such situations. Most importantly, we believe that PF maps even without segmentation results can provide a good illustration of the room layout which can be helpful for visual inspection by the human.

**Limitations**. Our segmentation approach does not yet support multi-storey buildings. We would like to note that most parts of the proposed pipeline as well as the proposed PF formulation apply to general 3D scenes. The extension to multi-storey buildings would foresee replacing the stack maxima operation with multimodal maxima detection and performing clustering in 3D space instead of 2D. Another important limitation is the moderate performance of our approach on very sparse point cloud data with multiple holes. This can be mitigated by extending the criteria of the interior space.

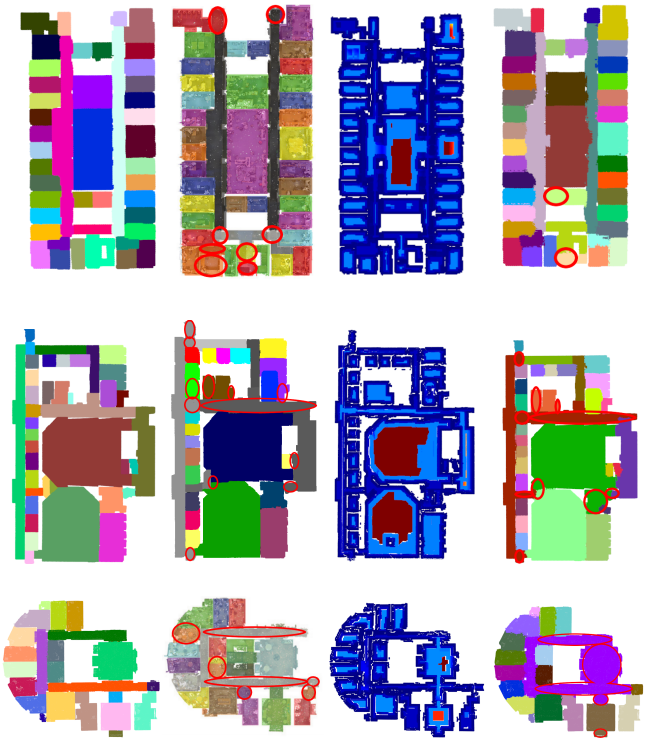


**Fig. 3**: Results for the large-scale dataset of [7]. From left to right: ground truth, results of [7], our PF map, our labeling result. Top row: Area 1, middle row: Area 2, bottom row: Area 3. Here with red ellipses we denote erroneously labeled rooms. See supplementary material for bigger figures.

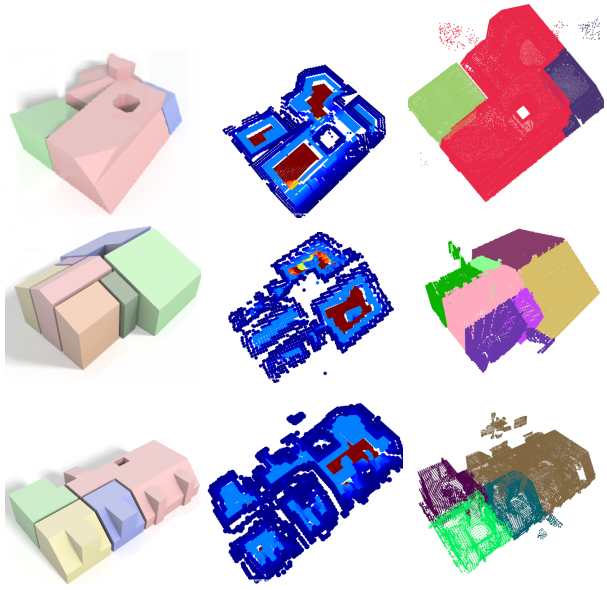


**Fig. 4**: Results for the unlabeled dataset violating the Manhattan-world assumption [3]. Top row: Modern, middle row: Cottage, bottom row: Penthouse. Left: reconstruction result of [3], middle: PF map, right: our result.

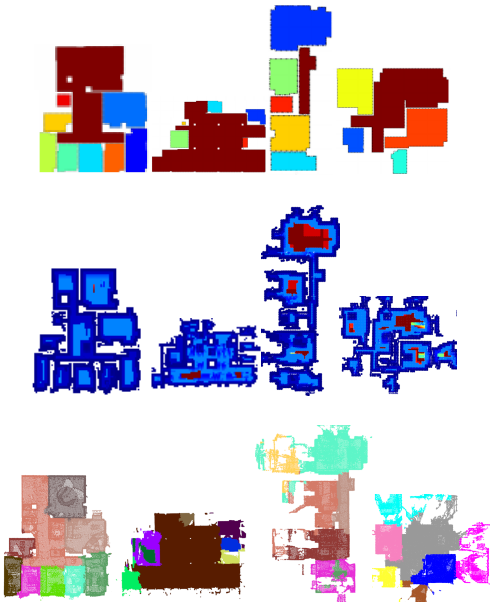


**Fig. 5**: Results for the unlabeled dataset of [5]. From left to right: Office 1, Office 2, Apartment 1, Apartment 2. Top row: results of [5], middle row: our PF map, bottom row: our labeling result. See supplementary material for bigger figures.

## 5. CONCLUSION

We have presented a novel framework to compute interior free space of indoor environments without assuming knowledge of scanner poses or the Manhattan-world structure. Operating on interior free space, we formulate a new anisotropic potential field for 3D environments that is robust to indoor clutter and occlusion. We then show how it can be used for general room segmentation without assuming any specific room layout. Our approach outperforms state-of-the-art methods for a number of datasets. Additionally, it is also applicable to new data modalities, such as point clouds resulting from a combination of CAD-models and depth scanner data.

## 6. REFERENCES


[1] R. Bormann, F. Jordan, W. Li, J. Hampp, and M. Hägele, "Room segmentation: Survey, implementation, and analysis," in *ICRA*, 2016, pp. 1019–1026.

[2] S. Friedman, H. Pasula, and D. Fox, "Voronoi random fields: Extracting topological structure of indoor environments via place labeling.," in *IJCAI*, 2007, vol. 7, pp. 2109–2114.

[3] C. Mura, O. Mattausch, and R. Pajarola, "Piecewise-planar reconstruction of multi-room interiors with arbitrary wall arrangements," *Computer Graphics Forum*, 2016.

[4] E. Turner, P. Cheng, and A. Zakhor, "Fast, automated, scalable generation of textured 3d models of indoor environments," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 3, pp. 409–421, 2015.

[5] S. Ikehata, H. Yang, and Y. Furukawa, "Structured indoor modeling," in *ICCV*, 2015, pp. 1323–1331.

[6] J. Xiao and Y. Furukawa, "Reconstructing the world's museums," *International Journal of Computer Vision*, vol. 110, no. 3, pp. 243–258, 2014.

[7] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese, "3d semantic parsing of large-scale indoor spaces," in *CVPR*, 2016.

[8] C. Mura, O. Mattausch, A. Jaspe Villanueva, E. Gobbetti, and R. Pajarola, "Automatic room detection and reconstruction in cluttered indoor environments with complex room layouts," Spring Conference on Computer Graphics, 2015.

[9] C.M. Harris, *Dictionary of architecture & construction*, McGraw-Hill, 2006.

[10] E. Brunskill, T. Kollar, and N. Roy, "Topological mapping using spectral clustering and classification," in *IROS*, 2007, pp. 3491–3496.

[11] E. Turner and A. Zakhor, "Floor plan generation and room labeling of indoor environments from laser range data," in *VISIGRAPP*, 2014, pp. 22–33.

[12] Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski, "Reconstructing building interiors from images," in *ICCV*, 2009, pp. 80–87.

[13] J. Straub, G. Rosman, O. Freifeld, J. J. Leonard, and J. W. Fisher III, "A mixture of manhattan frames: Beyond the manhattan world," in *CVPR*, 2014.

[14] O. Mattausch, D. Panozzo, C. Mura, O. Sorkine-Hornung, and R. Pajarola, "Object detection and classification from large-scale cluttered indoor scans," *Computer Graphics Forum*, vol. 33, no. 2, pp. 11–21, 2014.

[15] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *PAMI*, vol. 26, no. 9, pp. 1124–1137, 2004.

[16] H. Zhang, X. Chen, Y. Zhang, J. Li, Q. Li, and X. Wang, "Cuboids detection in rgb-d images via maximum weighted clique," in *2015 ICME*, 2015, pp. 1–6.

[17] R. Campello, D. Moulavi, A. Zimek, and J. Sander, "Hierarchical density estimates for data clustering, visualization, and outlier detection," *ACM Trans. Knowl. Discov. Data*, vol. 10, no. 1, 2015.