RGB-D Co-Segmentation on Indoor Scene with Geometric Prior and Hypothesis Filtering

Lingxiao Hang, Zhiguo Cao(✉), Yang Xiao, and Hao Lu

National Key Lab of Science and Technology of Multispectral Information Processing, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China
{lxhang,zgcao,Yang_Xiao,poppinace}@hust.edu.cn

Abstract. Indoor scene parsing is crucial for applications like home surveillance systems. Although deep learning based models like FCNs [10] have achieved outstanding performance, they rely on huge amounts of hand-labeled training samples at pixel level, which are hard to obtain. To alleviate labeling burden and provide meaningful clues for indoor applications, it's promising to use unsupervised co-segmentation methods to segment out main furniture, such as bed and sofa. Following traditional bottom-up co-segmentation framework for RGB images, we focus on the task of co-segmenting main furniture of indoor scene and fully utilize the complementary information of RGB-D images. First, a simple but effective geometric prior is introduced, using bounding planes of indoor scene to better distinguish between foreground and background. A two-stage hypothesis filtering strategy is further integrated to refine both global and local object candidate generation. To evaluate our method, the *NYUD-COSEG* dataset is constructed, on which our method shows significantly higher accuracy compared with previous ones. We also prove and analyze the effectiveness of both bounding plane prior and hypothesis filtering strategy with extensive experiments.

Keywords: Indoor RGB-D co-segmentation
Geometric prior for indoor scene · Object hypothesis generation

1 Introduction

Indoor scene parsing has great significance for applications like home surveillance systems. Deep learning models such as Fully Convolutional Networks (FCNs) [10]

This work is jointly support by the National High-tech R&D Program of China (863 Program) (Grant No. 2015AA015904), the National Natural Science Foundation of China (Grant No. 61502187), the International Science & Technology Cooperation Program of Hubei Province, China (Grant No. 2017AHB051), the HUST Interdisciplinary Innovation Team Foundation (Grant No. 2016JCTD120).
The First Author of This Paper is a Student.

J.-H. Lai et al. (Eds.): PRCV 2018, LNCS 11256, pp. 168–179, 2018.
https://doi.org/10.1007/978-3-030-03398-9_15

has achieved great success. However, training these models heavily relies on labeling huge amounts of samples, which is time consuming and labor intensive. In contrast, the unsupervised co-segmentation methods can simultaneously partition multiple images that depict the same or similar object into foreground and background. It can considerably alleviate labeling burden by producing object masks without semantic labels, which can be used as ground truth for training of deep neural networks [17]. Besides, co-segmenting main indoor furniture (bed, table, etc.) can provide meaningful clues for room layout estimation and human action analysis.

Although RGB co-segmentation has been studied thoroughly, RGB-D indoor scene co-segmentation remains an untouched problem. We discover two challenges when directly applying previous RGB methods. First, foreground and background appearance models are initialized using intuitive priors such as assuming pixels around the image frame boundaries as background, which fails in complex indoor scene. Second, the cluttering and occlusion of indoor condition make it hard to generate high-quality object candidates depending on RGB only in the unsupervised manner of co-segmentation.

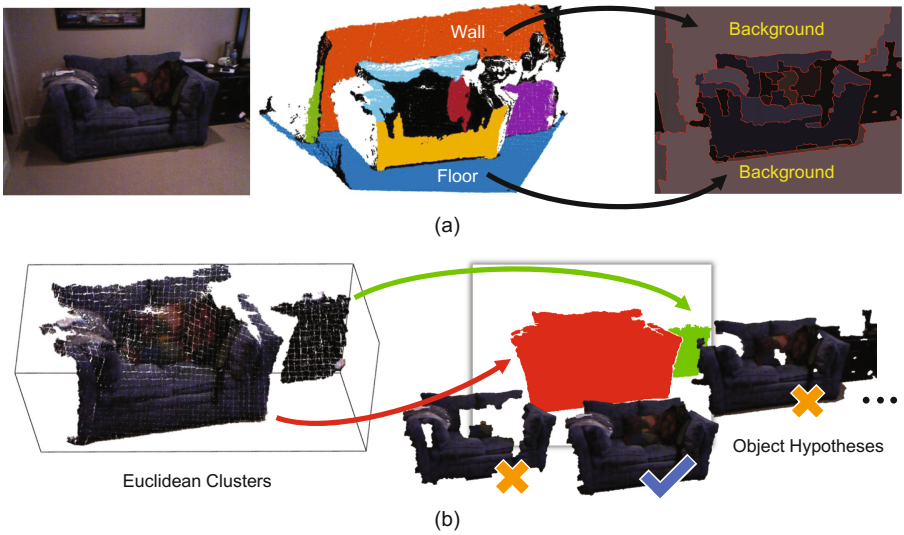

Fig. 1. Demonstration of our main contributions. (a) The geometric prior is used to reliably classify bounding planes of indoor scene, like wall and floor, as background. (b) The Euclidean clusters corresponding to foreground objects are leveraged to filter incomplete or overstretched object hypotheses.

To handle these challenges, we propose to integrate the geometric prior and hypothesis filtering strategy, shown in Fig. 1, into the traditional bottom-up co-segmentation pipeline. Our method fully utilizes the intrinsic properties of RGB-

D indoor scene to remedy the deficiencies of previous methods. The motivation of our method is detailed in the following two aspects.

First, our geometric prior addresses the problem of disentangling foreground from background. Existing unsupervised methods rely on boundary prior [1, 12], saliency prior [11,15] or even objectness prior [3,4] that requires training. All these priors are either ineffective or complicated in terms of indoor scene. On the one hand indoor objects commonly have intersection with image frame boundaries and show little contrast with background. On the other hand, the objectness methods are not specifically trained for indoor scene, which requires re-training on large labeled datasets. Instead, considering the abundant plane structures, shown in Fig. 1(a), our approach utilizes the unsupervised bounding plane prior that reliably specifies background regions. This simple but effective prior has no burden on manpower or computing resources.

Second, we improve object hypothesis generation for indoor images by going beyond simply combining connected segments in 2D RGB images. Since two neighboring segments that belong to different objects could be adjacent in 2D image plane but are spatially separated in 2.5D real world coordination. Inspired by this insight, our object hypothesis generation exploits a two-stage filtering strategy, using Euclidean clustering in 2.5D space to obtain separated point clusters. This improvement on hypothesis generation is able to increase the proportion of physically reasonable and high-quality proposals, which reduces the error during hypothesis clustering, especially for large objects.

To the best of our knowledge, this is the first paper addressing co-segmentation in RGB-D indoor scene. To evaluate our method, we re-organize the *NYUD v2* dataset [18] to establish a proper benchmark for co-segmentation of indoor scene. We demonstrate that our method can achieve state-of-the-art performance on our RGB-D indoor dataset. Our contributions are as follows:

- Our work provides the field of indoor RGB-D co-segmentation the first methodology focusing on large objects, which can help reduce the manual labeling effort for CNNs.
- A simple but effective bounding plane prior is first proposed to better distinguish foreground and background for RGB-D co-segmentation of complex indoor scene.
- A two-stage hypothesis generation filtering strategy is devised to overcome cluttering and occlusion problems of indoor scene, producing high-quality object proposals.

2 Related Work

Work Related to Unsupervised Co-Segmentation. Co-segmentation aims at jointly segmenting common foreground from a set of images. One setting is that only one common object is presented in each image. Color histogram was embedded as a global matching term into MRF-based segmentation model [14]. In [5] co-segmentation was formulated as a discriminative clustering problem with classifiers trained to separate foreground and background maximally. Yet

another more challenging setting is to extract multiple objects from a set of images, which is called the MFC (Multiple Foreground Co-segmentation). It was first addressed in [6] by building appearance models for objects of interest, followed by beam search to generate proposals. Recently RGB-D co-segmenting small props was tackled using integer quadratic programming [3]. Different from previous works, our method features RGB-D indoor scene.

Work Related to Co-Segmentation of Indoor Point Cloud Data. Another similar line of work aims at co-segmenting a full 3D scene at multiple times after changes of objects' poses due to human actions. Different tree structures [9,16] were used to store relations between object patches and present semantical results. However, the depth images we use are single viewed in 2.5D space, which suffer from the occlusion and cluttering problem eluded by their full 3D counterparts. Our proposed method is able to overcome these challenges by exploiting rich information of RGB-D image, without resorting to full viewed 3D data.

3 Bottom-Up RGB-D Indoor Co-Segmentation Pipeline

3.1 The Overall Framework for Bottom-Up Co-Segmentation

Our co-segmentation of main furniture for indoor images can be categorized as the MFC (Multiple Foreground Co-segmentation) problem. Given the input images $\mathcal{I} = \{I_1, ..., I_M\}$ of the same indoor scene, the goal is to jointly segment K different foreground objects $\mathcal{F} = \{F_1, ..., F_K\}$ from $\mathcal{I}$. As a result, each I_i is divided into non-overlapping regions with labels containing a subset of K foregrounds plus a background G_{I_i}. According to scenario knowledge, we define common foreground as major indoor furniture with certain functionality.

Traditional bottom-up pipeline for MFC co-segmentation [1] consists of three main steps, namely superpixel clustering, region matching and hypothesis generation. The first step merges locally consistent superpixels into compact segments. The second step refines segments in each image by imposing global consistency constraints, with the result that similar segments across images have the same label. The third step goes to a higher level that object candidates are generated by combining segments, which are later clustered to form final segmentation result.

With the motivation in Sect. 1, we made improvements to the first and the third step of the bottom-up pipeline, utilizing 2.5D depth information as a companion to RGB space so as to reduce ambiguity resulted from relying 2D color image only.

The pipeline of our method is shown in Fig. 2. For simplicity and clarity, we only show the co-segmentation pipeline of a single RGB-D image. Also, the second step in the traditional MFC of imposing consistency constraints across images is not shown, which directly follows [1].

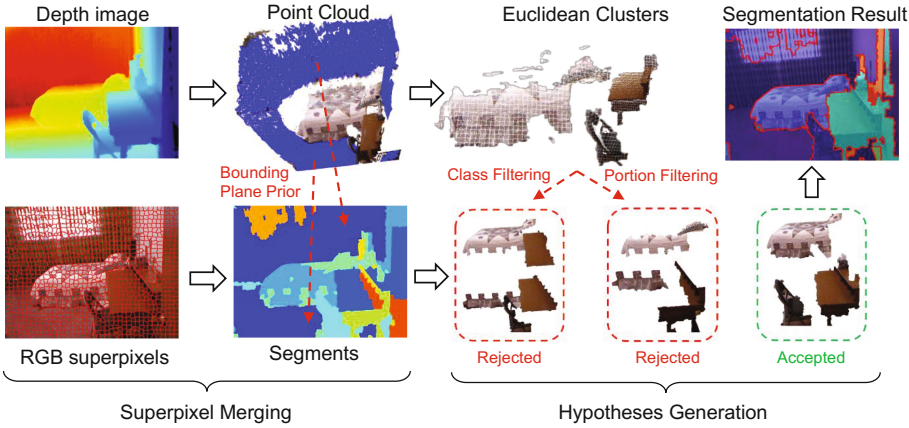

Fig. 2. The main technical pipeline of our RGB-D co-segmentation using the bounding plane prior and the two-stage hypothesis filtering. For simplicity, we exemplify the bottom-up segmentation with only one RGB-D image sample in M input images.

3.2 Superpixel Merging with Bounding Plane Prior

Given a depth image, we can use the pin-hole camera model to transform it into the 2.5D space, where each pixel p_i in 2D image has a 3D real-world coordinate $p_s(x, y, z)$.

For indoor scene, there are rich geometric structures and space relationships that can be very useful as guidance for unsupervised CV task, such as large planes, affordance of objects, etc. As can be apparently observed, the bounding planes, which correspond to walls, floors and ceilings in a real indoor scene, can be taken as a reliable prior for background regions. These bounding planes have two features to define the background. One feature is that these planes are the outer-most planes within the 2.5D space, whose only functionality is to enclose foreground objects within the room inside. The other feature is that dominant foreground objects in the scene always take up a certain amount of cubic space, whose consisting points will not lie on a sole plane.

Following [2], we first perform plane segmentation using 2.5D point cloud data. Iteratively using RANSAC to estimate plane parameters and Euclidean distances to assign points to planes and all planes in a image can be found, denoted by $\mathcal{P}_{I_i}$. Suppose the normal vector of each plane points towards the camera, the set of bounding planes $\mathcal{BP}_{I_i}$ for I_i is selected by its first feature, defined as:

$$\mathcal{BP}_{I_i} = \left\{ P_k \middle| P_k \in \mathcal{P}_{I_i}, \frac{1}{N}\sum_{s=1}^{N} \mathbb{1}\left\{D(p_s, P_k) < 0\right\} < \tau \right\}, i = \{1, ..., M\} \qquad (1)$$

where $D(p_s, P_k)$ is the Euclidean distance of point p_s to plane P_k and $\mathbb{1}\{\cdot\}$ is the indicator function. Referring to the first feature of bounding plane, the ratio of points on the outer side of the plane should be lower than a given threshold τ.

As the first main step in our bottom-up co-segmentation framework, merging locally similar superpixels into segments begins with the method of [8] to produce superpixels for each image respectively (the number of superpixels for each images is set to $N = 1200$). For superpixel merging, each superpixel S in $\mathcal{S}_{I_i}$ excluded by $\mathcal{BP}_{I_i}$ is assigned to an initial foreground segment R_c with probability given by a set of parametric functions v^c, $c = \{1, ..., C\}$. The parametric function $v^c : \mathcal{S}_{I_i} \rightarrow \mathbb{R}$ can be defined by the c-th foreground model, which in this paper is GMM (Gaussian Mixture Model). We use GMM with 32 Gaussian components to determine the color histogram $\mathbf{h}_S$ for each superpixel S. Thus, the probability of S belonging to the set of c-th foreground segments R_c is measured by the normalized χ^2 distance between $\mathbf{h}_S$ and $\mathbf{h}_{R_c}$. In terms of S included by $\mathcal{BP}_{I_i}$, we use $C + 1$ to denote background segment label and the probability is assumed to win over other segment label. The overall segment label probability for every superpixel is given by

$$P(R_c|S) = \begin{cases} \chi^2(\mathbf{h}_S, \mathbf{h}_{R_c}) & \text{if } S \notin \mathcal{BP}_{I_i} \\ 1 - \epsilon & \text{if } S \in \mathcal{BP}_{I_i}, c = C + 1 \\ \epsilon / C & \text{otherwise} \end{cases} \tag{2}$$

where ϵ is a quantity close to 0. After initializing the probability of assigning each superpixel S to segment R_c, we refine this merging result by GrabCut [13] using $P(R_c|S)$. Thus we can get the refined set of segments for each image, denoted as $\mathcal{R}_{I_i}$.

3.3 Two-Stage Hypothesis Filtering with Point Cloud Clustering

As the third main step of our bottom-up pipeline, hypothesis generation step combines arbitrary numbers of connected segments to form a pool of object candidates, which is crucial for the final foreground segmentation. Sensible hypotheses can accurately be clustered into K objects contained in the input images. We make the observation that final segmenting of objects is determined by two properties of object hypotheses, diversity and reliability. Diversity means that the hypothesis pool should involve all possible objects in the image without missing any. Reliability is the probability that a candidate belongs to a whole foreground object. Our goal is to find a pool with suffice diversity wherein each candidate is of maximal reliability.

Naively combining all possible connected segments in $\mathcal{R}_{I_i}$ to form object candidate reaches the maximum of pool diversity but the minimum of reliability. To make a trade-off, we propose a two-stage hypothesis filtering strategy to enlarge the proportion of reliable candidates while still retain the diversity.

Before filtering, we first provide a measurement tool for reliable candidate or in other words, *objectness*. While it is challenging for general purposed objectness prediction, in the case of RGB-D indoor scene it can be reduced to Euclidean clustering. In 2.5D point cloud, ignoring the bounding planes found in Sect. 3.2, we can find dominant clusters using Euclidean distance within a neighborhood tolerance and map them back to 2D image frame. These clusters, denoted as

$Q_k \in \mathcal{E}_{I_i}$ of image I_i, represent occupancy of dominant objects in the image, hence candidates who coincide with them are reliable.

Class Filtering. Spatially isolated point cloud clusters Q_k represent different objects respectively. We use class filtering to rid off hypotheses with coverage over two or more clusters. Let $\mathcal{H}_0$ denote hypothesis pool without filtering, $\mathcal{H}_1$ with class filtering, then the first selection step of candidates h can be expressed as

$$\mathcal{H}_1 = \left\{ h \middle| \sum_{k=1}^{\|\mathcal{E}_{I_i}\|} \mathbb{1}\left\{h \cap Q_k \neq \emptyset\right\} = 1, h \in \mathcal{H}_0 \right\} \tag{3}$$

The class filtering can refine the global segmentation result of foreground objects, largely alleviating the problem of segmenting out two or more objects that are in close proximity to each other as a single object.

Portion Filtering. Due to the inconsistent texture or piled clutter on the main furniture, it is likely to divide a whole object into locally consistent subsegments. To further improve the segmentation accuracy for main objects of indoor scene, we additionally impose portion filtering. Hypotheses that are overlapping with Q_k under a given threshold are discarded, leaving the most reliable candidate pool $\mathcal{H}_2$, which can be expressed as

$$\mathcal{H}_2 = \left\{ h \middle| \frac{area(h \cap Q_k)}{area(Q_k)} > \theta, h \in \mathcal{H}_1 \right\} \tag{4}$$

The portion filtering refines the segmentation for single main object, which in particular reduce the case where large objects are partially segmented.

4 Results and Discussion

4.1 *NYUD-COSEG* Dataset and Experimental Setup

Previous work on RGB-D co-segmentation such as [3] used the dataset of images captured under controlled lab environment or estimated depth images from general RGB image. No dataset of indoor scene suited for co-segmentation has been put forward. Based on the widely used RGB-D indoor dataset *NYUD v2* [18] for supervised learning algorithms, we propose a new dataset, *NYUD-COSEG*, with modification to the original *NYUD v2* dataset to extensively test our method and compare with other state of the art co-segmentation methods.

Since large furniture plays a more important role in scene layout estimation or applications involving daily human actions, we take classes like floor, wall and ceiling as background while furniture like bed, table and sofa as foreground. With this definition of object class of interest, we construct the *NYUD-COSEG* dataset by firstly grouping images captured in the same scene with aforementioned foreground classes. Each group contains 2 to 4 images and can be taken

as input for any co-segmentation algorithm. Next, the original ground truths are re-labeled. Trivial classes such as small props are removed. Small objects overlapped with large furniture are merged as the latter, exemplified by taking the pillow class as the bed class. The class-simplified ground truth is more sensible for evaluation of unsupervised methods.

After the organizing, the *NYUD-COSEG* dataset can be divided into of 3 main classes as `Bed`, `Table` and `Sofa`, each containing 104, 31 and 21 images respectively. It contains 62 classes in total (we consider all classes during evaluation).

We randomly choose 20% of images in the *NYUD-COSEG* dataset as validation set and apply grid search to find the optimal value for parameters. In our implementation, we set $C = 8$ in Eq. (2) and $\theta = 0.8$ in Eq. (4) as default.

4.2 Evaluation Metric and Comparison Study on NYUD-COSEG

The evaluation metric we adopt for co-segmentation algorithm on indoor scene is frequency weighted IOU (f.w.IOU). This choice takes into consideration that for room layout estimation and its applications, dominant objects (bed, sofa, etc.) of an image has more significance than less obvious ones (cup, books, etc.). On the contrary, metrics such as pixel accuracy, mean accuracy and mean IOU make no different treat on large and small objects, which is not practical for unsupervised co-segmentation algorithm comparison on indoor dataset. Let n_{ij} be the number of pixels of class i classified as class j, $t_i = \sum_j n_{ij}$ be the total number of pixels belonging to class i, and $t = \sum_i t_i$ be the number of all pixels. The f.w.IOU can be defined as $\frac{1}{t}\sum_i t_i n_{ii} / \left(t_i + \sum_j n_{ji} - n_{ii}\right)$.

We first make self-comparison among our proposed method and its several variants to verify the effectiveness of bounding plane prior and hypothesis filtering. We show the result of our method with center prior instead of bounding plane prior (BP^-), with class filtering only (PF^-), with portion filtering only (CF^-), without any filtering (F^{2-}), and our full version (Our), respectively. We then compare our method with two recent RGB co-segmentation of multiple foreground objects [1,7], with code available on the Internet.

Table 1 lists the f.w.IOU scores of each method on our *NYUD-COSEG* dataset. Some of the visual results are shown in Fig. 3. From both quantitative and qualitative results, we can make the following observations: (i) Our method and its variants have significantly higher f.w.IOU than other methods, with our full version exceeding previous RGB methods by at least 16% on average. The result confirms that the depth information has great potential in unsupervised co-segmentation. (ii) The bounding plane prior is the most decisive part in performance boosting, of which the absence causes the lowest average score among all variants. Correctly distinguishing between foreground and background is essential for further clustering and segmentation. (iii) The two-stage hypothesis filtering is also effective. Class filtering has more effect than portion filtering. The former avoids merging of different objects in the global image and the latter adds more detailed refinement to single objects.

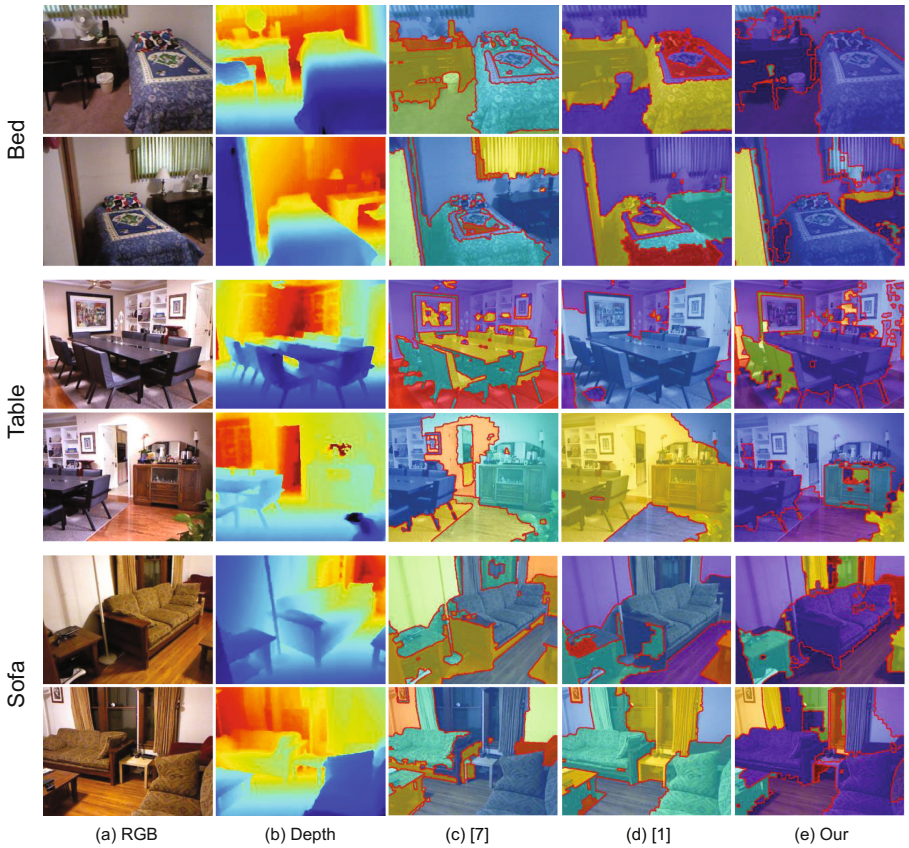

Fig. 3. Some qualitative co-segmentation results on our RGB-D indoor co-segmentation dataset *NYUD-COSEG*. From left to right: input RGB images, depth maps, results of [1,7], Our full version. (Common objects are shown in the same color with red separating boundaries.). (Color figure online)

4.3 Parameter Evaluation and Discussion

As mentioned in Sect. 4.1, our method contains two important parameters: cluster number C for superpixel merging and portion ratio θ for portion filtering. We fix one parameter to default and vary the other in a reasonable range to see how the f.w.IOU score will change accordingly, as shown in Fig. 4. The purple line indicating mean f.w.IOU score proves that our default values for the two parameters are optimal. Additionally, we find in Fig. 4(a) that too many clusters will not improve segmentation accuracy. Besides, in hypothesis generation step, the time costing is proportional to 2^C. As shown in Fig. 4(b) the accuracy varies mildly with respect to portion ratio θ, within 2.6%, although higher θ has the tendency to improve the result in view of mean f.w.IOU score.

Table 1. Comparison of f.w.IOU score of different methods (%) on *NYUD-COSEG* dataset. The highest is marked in bold.

Method	[7]	[1]	BP^-	CF^-	PF^-	F^{2-}	Our
Bed	43.56	46.37	53.08	62.59	65.85	61.18	**68.42**
Table	46.13	42.63	48.16	52.68	57.21	51.69	**58.31**
Sofa	46.66	53.82	59.96	54.69	62.40	56.69	**64.56**
Mean	45.45	47.61	53.73	56.65	61.82	56.52	**63.76**

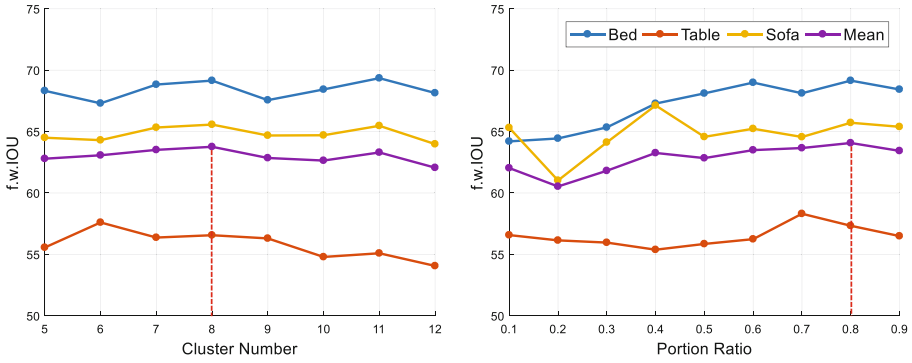

Fig. 4. The accuracy changing with respect to variation of two parameters of our co-segmentation method. (Color figure online)

5 Conclusion

In this paper the problem of RGB-D indoor co-segmentation of main furniture is considered. Previous methods use RGB images only. As indoor scene are typical of cluttering and occlusion, foreground merged with similar background and low quality object hypotheses are the two main factors that hinder the performance. We propose to handle these challenges using geometric and spatial information provided by depth channel. Bounding plane prior and a two-stage hypothesis filtering strategy are introduced and integrated into traditional bottom-up co-segmentation framework. To evaluate our method, the *NYUD-COSEG* dataset is constructed based on *NYUD v2*, with thorough experiments proving the effectiveness of our two improvements.

As the first work on the task of indoor co-segmentation, our method is limited in segmenting small objects like stuff on the table, which is most challenging in terms of unsupervised machine learning condition. In the future work we plane to extending our model by incorporating more supervising signals such as supporting relationship to discern small objects. Besides, the question of how to use probabilistic models to formulate our bounding plane prior and hypothesis

filtering is worth studying. We believe it will reduce the number of parameters needed to be set manually and thus can elevate the robustness of our method.

References

1. Chang, H.S., Wang, Y.C.F.: Optimizing the decomposition for multiple foreground cosegmentation. Comput. Vis. Image Underst. **141**, 18–27 (2015)
2. Deng, Z., Todorovic, S., Latecki, L.J.: Unsupervised object region proposals for RGB-D indoor scenes. Comput. Vis. Image Underst. **154**, 127–136 (2017)
3. Fu, H., Xu, D., Lin, S., Liu, J.: Object-based RGBD image co-segmentation with mutex constraint (2015)
4. Fu, H., Xu, D., Zhang, B., Lin, S.: Object-based multiple foreground video co-segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3166–3173 (2014)
5. Joulin, A., Bach, F., Ponce, J.: Discriminative clustering for image co-segmentation. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1943–1950. IEEE (2010)
6. Kim, G., Xing, E.P.: On multiple foreground cosegmentation. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 837–844. IEEE (2012)
7. Kim, G., Xing, E.P., Fei-Fei, L., Kanade, T.: Distributed cosegmentation via submodular optimization on anisotropic diffusion. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 169–176. IEEE (2011)
8. Levinshtein, A., Stere, A., Kutulakos, K.N., Fleet, D.J., Dickinson, S.J., Siddiqi, K.: TurboPixels: fast superpixels using geometric flows. IEEE Trans. Pattern Anal. Mach. Intell. **31**(12), 2290–2297 (2009)
9. Lin, Y.: Hierarchical co-segmentation of 3D point clouds for indoor scene. In: 2017 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 1–5. IEEE (2017)
10. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
11. Meng, F., Li, H., Liu, G., Ngan, K.N.: Object co-segmentation based on shortest path algorithm and saliency model. IEEE Trans. Multimed. **14**(5), 1429–1441 (2012)
12. Quan, R., Han, J., Zhang, D., Nie, F.: Object co-segmentation via graph optimized-flexible manifold ranking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 687–695 (2016)
13. Rother, C., Kolmogorov, V., Blake, A.: GrabCut: interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. (TOG) **23**, 309–314 (2004)
14. Rother, C., Minka, T., Blake, A., Kolmogorov, V.: Cosegmentation of image pairs by histogram matching-incorporating a global constraint into MRFs. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 993–1000. IEEE (2006)
15. Rubinstein, M., Joulin, A., Kopf, J., Liu, C.: Unsupervised joint object discovery and segmentation in internet images. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1939–1946. IEEE (2013)
16. Sharf, A., Huang, H., Liang, C., Zhang, J., Chen, B., Gong, M.: Mobility-trees for indoor scenes manipulation. In: Computer Graphics Forum, vol. 33, pp. 2–14. Wiley Online Library (2014)

17. Shen, T., Lin, G., Liu, L., Shen, C., Reid, I.: Weakly supervised semantic segmentation based on co-segmentation. arXiv preprint arXiv:1705.09052 (2017)
18. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from RGBD images. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7576, pp. 746–760. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33715-4_54