GRAPH-BASED RGB-D IMAGE SEGMENTATION USING COLOR-DIRECTIONAL-REGION MERGING

Xiong Pan[1], *Zejun Zhang*[1,⋆], *Yizhang Liu*[1], *Changcai Yang*[1], *Qiufeng Chen*[1], *Li Cheng*[2], *Jiaxiang Lin*[1], *Riqing Chen*[1]

{[1]the Digital Fujian Institute of Big Data for Agriculture and Forestry, College of Computer and Information Sciences; [2]Jinshan colledge}, Fujian Agriculture and Forestry University, Fuzhou, China.

ABSTRACT

Color and depth information provided simultaneously in RGB-D images can be used to segment scenes into disjoint regions. In this paper, a graph-based segmentation method for RGB-D image is proposed, in which an adaptive data-driven combination of color- and normal-variation is presented to construct dissimilarity between two adjacent pixels and a novel region merging threshold exploiting normal information in adjacent regions is proposed to control the proceeding of the region merging. We evaluate our method on the NYU-v2 depth database and compare it with several published RGB-D partition methods. The experimental results show that our method is comparable with the state-of-the-art methods and provides more details of structures in the scene.

Index Terms— RGB-D image segmentation, graph-based segmentation, data-driven weight, color-normal-region merging

1. INTRODUCTION

Image segmentation is an important task in image and video processing and has many applications, such as video surveillance in public spaces [1], object recognition [2, 3] and robotic search [4]. The goal of image segmentation is to provide structure information about the scenes covered by the cameras for advanced task, such as recognition, classification and 3D reconstruction. For RGB format images, the features using RGB channels is considered for grouping pixels perceptually into several disjoint regions. With the release of depth-cameras, such as Microsoft Kinect equipment [5], X-tion PRO[6] and Intels Realsense 3D cameras [6], it becomes easy for one to obtain simultaneous color and depth data of scenes, called as RGB-D image. Based on the RGB-D format image data, the quality of segmenting scenes can by improved via exploiting color and depth information simultaneously.

Recently, many kinds of algorithms, cluster-based methods [7], graph-based methods [8, 9, 10] and superpixel merging methods [11], have been presented for RGB-D images segmentation. The foundation of the graph-based methods is the region adjacency graph (RAG) established from original image, in which the key point is to design the dissimilarity (or similarity) between two adjacent regions (or pixels). One of methods of developing the dissimilarity between two adjacent pixels in RGB-D image is to combine color and depth information with many weights. Silberman et. al. [12] used 3D normal and color information at each pixel to obtain the dissimilarity of two adjacent pixels. In literature [11], a weighted geodesic driven metric has been proposed to measure the dissimilarity of two adjacent pixels and the superpixel structure of a RGB-D image was yielded based on it. Under the statistical clustering framework, Hanat et.al. [7] developed a joint color-spatial-directional based dissimilarity and clustered vertexes in the RAG (i.e., pixels) into several clusters (i.e., regions) with an iterative clustering method followed by a subsequent region merging process to generate final segmentation results. However, this method tends to produce the segmentation results with jagged edges, which may be resulted from the inaccuracy of estimating of statistical parameters in multivariate Gaussian distribution.

An interesting work is to extend the existing graph-based RGB image segmentation method to a new method by introducing depth (or geometrical) information into the existing model. Literature [8] incorporated depth data into a RGB-based dissimilarity, which appeared in the method proposed in [13], with a fixed weight fusion strategy, producing the RGB-D image segmentation method. However, this simple combination of the color and depth data tends to generate inaccurate segmentation results because of existence of noise in an image.

In this paper, a novel graph-based RGB-D image segmentation algorithm is proposed for indoor scene. An adaptive data-driven weight strategy is presented to combine color and normal information to construct dissimilarity between

⋆Corresponding author: zjzhang_fafu@163.com. This work was supported in part by the National Natural Science Foundation of China (No. 61501120, No. 61702101, No. 61701117), the Natural Science Fund of Fujian Province (No. 2016J01753, No.2017J01736), the construction fund for Digital Fujian Big Data for Agriculture and Forestry (No. KJG18019A), the special fund for scientific and technological innovation of Fujian Agriculture and Forestry University (No. CXZX2016026, No. CXZX2016031), and China ASEAN Maritime Cooperation Fund Project (No. 2020399)

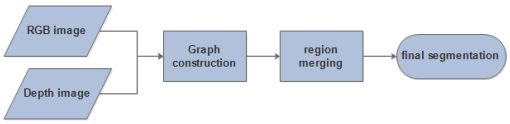

Fig. 1. Outline of the proposed segmentation method.

two adjacent pixels, which is robust against noise in an image. The weight at each pixel is calculated by using bi-semi-Gaussian functions with multi-direction. In order to control region merging process accurately, a novel merging threshold is developed by utilizing normal information in sub-regions. Based on our data-driven dissimilarity of two adjacent pixels and adaptive region merging threshold, a region merging criterion, the same as one used in [13], is proposed, which is an ordered region merging strategy and in which each adjacent pixel-pair will be accessed no more than once.

The rest of this paper is organized as follows: Section 2 presents the proposed method. The experimental results are shown in Section 3, which includes the comparison with the state-of-the-art methods and discussion. Finally, Section 4 draws conclusions and discusses future perspectives.

2. METHODOLOGY

In this section, we present our RGB-D image segmentation algorithm for indoor scene. Firstly, the outline of the proposed method is shown in Subsection 2.1. Secondly, Subsection 2.2 describes the graph-based representation of RGB-D image, which is the foundation of our segmentation method. Finally, we present the proposed region merging processing in Subsection 2.3.

2.1. Outline of the proposed method

The proposed segmentation method is based on region merging and its outline is shown in Fig. 1, which consists of two blocks. The first block is to construct the weighted graph to represent an image by combining RGB and normal information with data-driven weight. The second one generates the final segmentation with region merging.

2.2. Weighted graph

Let $I(x,y) : \Omega \rightarrow \mathbb{R}^4$ represents an RGB-D image, where $\Omega = \{(x,y) \in \mathbb{Z}^2 : x \in \{1,2,\cdots,N_x\}, y \in \{1,2,\cdots,N_y\}\}$ indicates the discrete rectangular grid, N_x and N_y are the numbers of row and column of the image, respectively. $\mathbb{R}^4$ is 4-demension real space, which denotes three color channels and depth channel.

Let $\Re = \{\Omega_k, k = 1,2,\cdots,K\}$ denotes a segmentation result of an image. Each region Ω_k is the set of the pixels in a region. In this paper, an weighted undirected graph, known as Region Adjacency Graph (RAG), $RAG(V,E,\boldsymbol{w})$, is used to represent segmentation results. The node set $V = \{v_k, k = 1,2,\cdots,K\}$ indicates the K regions of the segmentation region, and the set E is a subset of $V \otimes V$, where $e_m = (v_k, v_p) \in E$ denotes that the regions Ω_k and Ω_p are adjacent. The weight, $w(e_{kp}) \in \boldsymbol{w}$, of an edge $e_{kp} = (v_k, v_p) \in E$ is dissimilar between two regions Ω_k and Ω_p, which is calculated by the dissimilarity measure Eq.(1) described in Section 2.2.1.

2.2.1. Dissimilarity of two adjacent pixels

The dissimilarity of two adjacent pixels, p_i and p_j, consists of two terms, color- and normal-variation between these two pixels, and they are linearly combined using adaptive data-driven weight. The dissimilarity of p_i and p_j is given by

$$\kappa(p_i,p_j) = \frac{w(p_i,p_j)\cdot c(p_i,p_j) + v(p_i,p_j)\cdot f(p_i,p_j)}{w(p_i,p_j) + v(p_i,p_j)} \quad (1)$$

where $c(p_i,p_j)$ and $f(p_i,p_j)$ are color- and normal-variation between p_i and p_j, respectively. $w(p_i,p_j)$ and $v(p_i,p_j)$ are data-driven weight calculated through using color information of pixels around these two pixels. The color variation $c(p_i,p_j)$ is from [13]

$$c(p_i,p_j) = \sqrt{r(p_i,p_j)^2 + g(p_i,p_j)^2 + b(p_i,p_j)^2} \quad (2)$$

and the normal variation is defined as

$$f(p_i,p_j) = 1 - \cos(\langle \vec{n}_{p_i}, \vec{n}_{p_j} \rangle) \quad (3)$$

where $\vec{n}_{p_i}$ and $\vec{n}_{p_j}$ denote the normal vectors at pixel p_i and p_j, respectively. The weight $w(p_i,p_j)$ and $v(p_i,p_j)$ are defined as

$$w(p_i,p_j) = \max_{\theta}\{\nabla(p_i,p_j;\theta)\}, v(p_i,p_j) = \exp\{w(p_i,p_j) - \alpha\} \quad (4)$$

and $\nabla(p_i,p_j;\theta) = |m_{g_+}(p_i,\theta) - m_{g_-}(p_j,\theta)|$, where α is a constant, and $m_{g_+}(p_i,\theta)$ and $m_{g_-}(p_j,\theta)$ are Gaussian-weighted mean of rgb channels of pixels in two bi-semi-circle regions at direction θ, as shown in Fig. 2. In Fig. 2, g_+ and g_- are Gaussian weight of pixels in two semi-circle regions, respectively, and radius of each semi-circle is proportional to the standard deviation σ of Gaussian function g_+(or g_-).

2.2.2. Properties of the proposed dissimilarity

The data-driven weighted combination of color- and normal-variation in Eq.(1) is one of the major contributions of this paper. Comparing with literatures [13, 14], in which color and depth information are combined with fixed weight, our proposed data-driven weight ensures that the adjacent pixel-pairs straddle (or near) the border between two regions have larger weights and the others possess smaller weights, which makes the subsequent region merging process merges the two

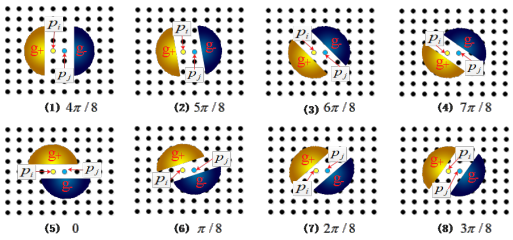

Fig. 2. Bi-semi-Gaussian functions. 0, $4\pi/8$,. . ., and $7\pi/8$ are angles of two bi-semi-circle Gaussian functions, g_+ and g_-.

adjacent pixels (or regions) from an identical homogeneous region as soon as possible. Moreover, as the weight average of pixels value using Gaussian function, the proposed adaptive weight strategy is inherent in suppress the noise around the adjacent pixel-pairs.

2.3. Region merging

The goal of graph-based region merging is to partition a graph into several disjoint sub-graphs with a criterion that determines the quality of partition results and the efficiency of the method. Inspired by [13], the proposed criterion consists of three aspects: representation of region, order of region merging and merging threshold.

The same as approach used in [13], in this paper, a region is represented by a Minimum Spanning Tree (MST) produced by its corresponding subgraph delimiting the region, and the segmentation results with N regions are denoted by a Minimum Spanning Forest (MSF) with N MSTs. With respect to the order of region merging, sequential merging strategy is utilized following a sorting procedure based on the proposed dissimilarity measure Eq.(1). Here, the sorting procedure reorder the set E of the RAG in ascending order of dissimilarity in $\boldsymbol{w}$ of the RAG. The reordered set E is named as E_{order}, $E_{order} = \{e_1, e_2, \ldots, e_n\}$, where dissimilarities of edge $e_i, i = 1, 2, \cdots, n$, satisfy $w(e_1) \leq w(e_2) \leq \ldots \leq w(e_n)$.

2.3.1. *Merging threshold*

For a partitioning result $\Re$, $\Re = \{\Omega_k, k = 1, 2, \cdots, K\}$, with K regions, their corresponding MSTs are $MST(\Omega_1)$, $MST(\Omega_2), \ldots,$ and $MST(\Omega_K)$. When considering the k-th edge, $e_k \in E_{order}$, which links two pixels p_l and p_m, and assuming that the pixels p_l and p_m belong to region Ω_i and Ω_j, respectively, the threshold value $MInt(\Omega_i, \Omega_j)$ determines whether these two regions, Ω_i and Ω_j, will be merged, and the threshold value is defined as,

$$\begin{aligned} MInt(\Omega_i, \Omega_j) &= \min\{Int(\Omega_i) + T(\Omega_i; \Omega_j), Int(\Omega_j) + \\ &\qquad T(\Omega_j; \Omega_i)\} \\ Int(\Omega_k) &= \max_{e \in MST(\Omega_k)} \{w(e)\}, k = i, j \end{aligned} \tag{5}$$

where $T(\Omega_i; \Omega_j)$ is computed by

$$\begin{aligned} T(\Omega_i; \Omega_j) &= [\varepsilon - \varphi(\Omega_i, \Omega_j)]/N_i \\ T(\Omega_j; \Omega_i) &= [\varepsilon - \varphi(\Omega_i, \Omega_j)]/N_j \end{aligned} \tag{6}$$

where ε is a constant; N_i and N_j are the number of pixels in the regions Ω_i and Ω_j, respectively; and $\varphi(\Omega_i, \Omega_j)$ is defined as, $\varphi(\Omega_i; \Omega_j) = 1 - \cos\langle \vec{n}_{\Omega_i}, \vec{n}_{\Omega_j} \rangle$, in which $\vec{n}_{\Omega_i}$ and $\vec{n}_{\Omega_i}$ are average normal in regions Ω_i and Ω_j, respectively.

2.3.2. *Merging algorithm*

Employing the reordered set E_{order}, the region- (or pixel-) pairs will be merged orderly, and each edge in the set E of the RAG, $RAG(V, E, \boldsymbol{w})$, is accessed by the region merging process no more than once. Without iterative searching and updating of the set $\boldsymbol{w}$ in the RAG, so this proposed merging algorithm is very efficient. Pseudo-code of the proposed region merging algorithm is shown in **Algorithm 1**.

Algorithm 1 Color-Directional-Region Merging Method

Input: RGB image, Depth image, the standard deviation σ of the Gaussian function, the constant α and ε.
Output: Final segmentation results represented by an MSF.

Construct $RAG(V, E, \boldsymbol{w})$ using Eq.(1)-(4)
Sort the set E in ascending order of dissimilarity in $\boldsymbol{w}$, and generate the reordered edge set E_{order}.
For $k = 1$ to $|E_{order}|$ do
    Find two adjacent regions Ω_i and Ω_j to make the vertices v_i and v_j liked by the k-th edge e_k satisfy that $v_i \in \Omega_i$ and $v_j \in \Omega_j$
  If $\Omega_i \cap \Omega_j = \emptyset$do
    Compute the merging threshold $MInt(\Omega_i, \Omega_j)$ using Eq.(5)-(6)
    If $w(e_k) \leq MInt(\Omega_i, \Omega_j)$ do
      Merging the region Ω_i and Ω_j into a larger one
    End
  End
End
Return final segmentation results.

3. RESULTS AND DISCUSSION

3.1. Dataset and evaluating benchmarks

In this paper, the NYUv2 dataset [12], which composed of 1449 indoor images with RGB and depth data, is used to evaluate the proposed algorithm comparing with three published methods, including GB-RGBD [13], PIS [15] and JCSD-RM [7]. Variation-of Information(VI), Probability Rand-Index(RI), Ground-Truth-Region Covering(GTRC)and precision-recall based Boundary F-Measure(BFM) [16] are used to compare the results generated by different methods and ground truth labels in [6], quantitatively.

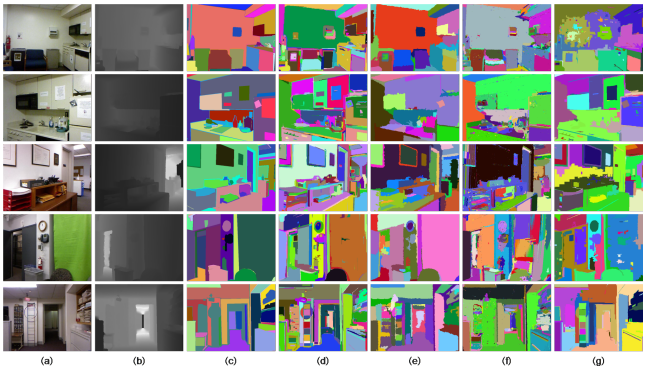

Fig. 3. Segmentation results on NYUv2 dataset. (a)Original color image, (b)depth image, (c)ground Truth, (d)our method, (e)JCSD-RM, (f)PIS, and (g)GB-RGBD.

Table 1. Comparing with the state-of-the-art methods.

	GTRC			PRI		VI		P	R	BFM
	ODS	OIS	Best	ODS	OIS	ODS	OIS			
PIS	0.43	-	-	0.88	-	3.16	-	0.37	0.69	0.48
GB-RGBD	0.44	0.49	0.56	0.87	0.89	2.46	2.34	0.40	0.60	0.48
JCSD-RM	**0.55**	-	-	**0.91**	-	**2.12**	-	**0.56**	0.43	0.49
OUR METHOD	0.50	**0.55**	**0.61**	0.90	**0.91**	2.42	**2.25**	0.38	**0.76**	**0.51**

ODS: a universal fixed scale OIS: a fixed scale per image
Best: from any level of the hierarchy or collection Ground truth: extracted from [5]

Table 2. Comparing with the proposed and JCSD method.

	GTRC	PRI	VI	BFM
JCSD	0.46	0.87	2.68	0.46
OUR METHOD	**0.50**	**0.90**	**2.42**	**0.51**

3.2. Experiments and results

In all of the experiments, two parameters α and ε in Eq.(4) and (6) are set to 1 and [0.5, 3.0], respectively. The segmentation results of the GB-RGBD and the PIS are obtained by the authors′ source code [15], and the results of the JCSD-RM are available in its web site. Table 1 shows the benchmarks on the NYUv2. The best results are indicated by **bold face**.

Fig. 3 shows several segmentation results produced by different methods. Visually in Fig. 3, we can see that our proposed method obtains better results than the PIS and the GB-RGBD method, and comparing with the JCSD-RM method, our method also provides regions with smoother boundaries, which benefited from the utilization of the bi-semi-Gaussian function in the proposed dissimilarity.

3.3. Discussion

From Table 1, we can see that the performance of the proposed and the JCSD-RM method is comparable. Moreover, our method is better in boundary localization indicated by the BFM index, which is due to the quality of the proposed method to generate smoother boundaries than the JCSD-RM method. By the way, the difference between the ground-truth used this paper and one utilized in literature [7] is the reason why the quantitative indexes of the JCSD-RM method shown in Table 1 is not the same as ones in [7].

It is worth noticed that the methodology of the proposed method is analogous to the one in literature [7] that consists of two stages: initial partition and region merging. From Fig. 3, comparing segmentation results with ones provided by the JCSD-RM method, one can consider our segmentation results as an initial partition. From the view of initial partition, Table 2 gives the comparison of quantitative indexes of initial segmentation results generated by our and the JCSD method. Comparing the quantitative indexes of the JCSD method in Table 2 with the ones of the JCSD-RM method in Table 1, we are confident that our method can be improved further through subsequent region merging process using more elaborate criterion.

4. CONCLUSION

In this paper, we proposed a graph-based indoor scene segmentation method using RGB-D images. Multi-directional rotated bi-semi-Gaussian functions were used to calculate the dissimilarity between two adjacent pixels. Based on it, a data-driven combination method of color- and normal-information was presented to indicate the dissimilarity between two adjacent pixels and an undirected and weighted RAG was constructed. In the region merging stage, a novel method computing merging threshold was presented by utilizing normal information in two adjacent regions.

As discussed in this paper, our method can also be considered as an initial partition method for RGB-D image segmentation, and our segmentation results will be improved by using more elaborate region merging criterion. So, our further work will be focus on subsequent region merging exploiting sub-region information including color, shape and geometrical properties.

5. REFERENCES

[1] Z. Xu, C. Hu, and L. Mei, "Video structured description technology based intelligence analysis of surveillance videos for public security applications," *Multimedia Tools and Applications*, vol. 75, no. 19, pp. 12155–12172, 2016.

[2] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3367–3375.

[3] H. Wang, D. Huang, K. Jia, and Y. Wang, "Hierarchical image segmentation ensemble for objectness in rgb-d images," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 93–103, 2017.

[4] M. Senanayake, I. Senthooran, J. C. Barca, H. Chung, J. Kamruzzaman, and M. Murshed, "Search and tracking algorithms for swarms of robots: A survey," *Robotics and Autonomous Systems*, vol. 75, pp. 422–434, 2016.

[5] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1318–1334, 2013.

[6] S. Song, S. P. Lichtenberg, and J. Xiao, "Sun rgb-d: A rgb-d scene understanding benchmark suite," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 567–576.

[7] M. A. Hasnat, O. Alata, and A. Trmeau, "Joint color-spatial-directional clustering and region merging (jcsd-rm) for unsupervised rgb-d image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 11, pp. 2255–2268, 2016.

[8] S. Hickson, S. Birchfield, I. Essa, and H. Christensen, "Efficient hierarchical graph-based segmentation of rgb-d videos," in *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 344–351.

[9] J. Strom, A. Richardson, and E. Olson, "Graph-based segmentation for colored 3d laser point clouds," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 2131–2136.

[10] J. Yang, Z. Gan, K. Li, and C. Hou, "Graph-based segmentation for rgb-d data using 3-d geometry enhanced superpixels," *IEEE Transactions on Cybernetics*, vol. 45, no. 5, pp. 927–940, 2015.

[11] X. Pan, Y. Zhou, F. Li, and C. Zhang, "Superpixels of rgb-d images for indoor scenes based on weighted geodesic driven metric," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 10, pp. 2342–2356, 2017.

[12] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *European Conference on Computer Vision*, 2012, pp. 746–760.

[13] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[14] X. Ren, L. Bo, and D. Fox, "Rgb-(d) scene labeling: Features and algorithms," in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2759–2766.

[15] C. J. Taylor and A. Cowley, "Parsing indoor scenes using rgb-d imagery," in *Robotics: Science and Systems*, 2013, vol. 8, pp. 401–408.

[16] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.