Elastic Edge Boxes for Object Proposal on RGB-D Images

Jing Liu[1,2], Tongwei Ren[1,2(✉)], and Jia Bei[1,2]

[1] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
rentw@nju.edu.cn, {ljing12,beijia}@software.nju.edu.cn
[2] Software Institute, Nanjing University, Nanjing, China

Abstract. Object proposal is utilized as a fundamental preprocessing of various multimedia applications by detecting the candidate regions of objects in images. In this paper, we propose a novel object proposal method, named *elastic edge boxes*, integrating window scoring and grouping strategies and utilizing both color and depth cues in RGB-D images. We first efficiently generate the initial bounding boxes by edge boxes, and then adjust them by grouping the super-pixels within elastic range. In bounding boxes adjustment, the effectiveness of depth cue is explored as well as color cue to handle complex scenes and provide accurate box boundaries. To validate the performance, we construct a new RGB-D image dataset for object proposal with the largest size and balanced object number distribution. The experimental results show that our method can effectively and efficiently generate the bounding boxes with accurate locations and it outperforms the state-of-the-art methods considering both accuracy and efficiency.

Keywords: Elastic edge boxes · Object proposal · RGB-D image

1 Introduction

Object proposal aims to detect candidate regions possibly containing class-independent objects in an image [1], which is widely used as a fundamental of various multimedia applications, such as scene analysis [2], image annotation [3] and retrieval [4], object recognition [5] and matching [6], visual tracking [7], and social media mining [8].

Typically served as a preprocessing procedure, object proposal is usually needed to satisfy the following requirements: First, object proposal should cover all or most objects in images with a limited number of candidate regions. In this way, the candidate regions can provide a majority of image content and reduce the further processing cost. Second, the candidate regions, usually bounding boxes, should accurately cover the objects. As shown in [9], improving intersection over union (IoU) of candidate regions and objects is as important as increasing recall of objects in several applications, such as object detection. Finally, the

Q. Tian et al. (Eds.): MMM 2016, Part I, LNCS 9516, pp. 199–211, 2016.
DOI: 10.1007/978-3-319-27671-7_17

processing of object proposal should be efficient, which will benefit its usage in realtime or large-scale applications.

The existing methods address object proposal problem with two typical strategies: window scoring and grouping [9]. The methods using window scoring strategy typically score quantities of candidate bounding boxes according to some features which measure the likelihood of a box containing an object. They usually have high efficiency but fail to detect most of objects under high IoU. And the methods using grouping strategy generally initialize a number of segments and then merge the similar segments to produce final results. They can obtain the accurate bounding boxes especially under high IoU, but they are usually time consuming.

Generally speaking, the current object proposal methods suffer two problems. In one aspect, both window scoring strategy and grouping strategy have their drawbacks either in accuracy or efficiency, which limits their usage in many applications. An interesting idea is to combine these two strategies together to obtain both high accuracy and efficiency, but the related research is still in embryonic stage [10]. In the other aspect, depth information has been proved to be effective in discriminating objects from complex scenes [11], but most current methods merely focus on color cue and ignore depth in object proposal.

In this paper, we propose a novel object proposal method, named *elastic edge boxes*, by integrating window scoring and grouping strategies and exploring both color and depth cues in RGB-D images. Figure 1 shows an overview of the proposed method. To each RGB-D image (Fig. 1(a)), we first utilize window scoring strategy to identify the potential object locations with boxes according to edge cue (Fig. 1(b)). Then, we represent the RGB-D image with super-pixels and select the undetermined super-pixels for each box (Fig. 1(c)). Finally, we adjust the boundary of each box by applying grouping strategy on the undetermined super-pixels

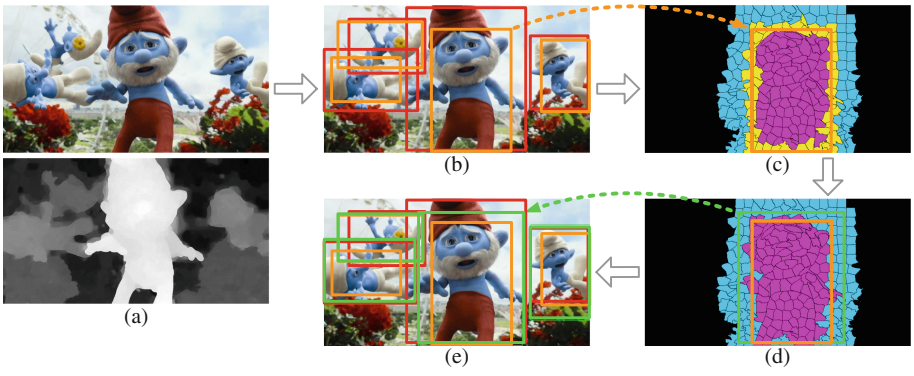

Fig. 1. An overview of the proposed method. (a) RGB-D image. (b) Initial bounding boxes (orange boxes) by window scoring strategy and ground truths (red boxes). (c) Super-pixel representation. (d) Box boundary adjustment by grouping strategy. (e) Final bounding boxes (green and orange boxes) and ground truths (red boxes) (Color figure online).

(Fig. 1(d)) and generate the final object proposal result (Fig. 1(e)). To the best of our knowledge, it is the first object proposal method integrating window scoring and grouping strategies for RGB-D images. To validate the performance of the proposed method, we construct an RGB-D image dataset, named *NJU1500*, on the base of stereo objectness dataset. The experiments show that our method can generate the bounding boxes with accurate locations under both low and high accuracy, and it outperforms the existing methods considering both accuracy and efficiency.

Our major contribution can be summarized as:

- We propose a novel object proposal method for RGB-D images, which can obviously improve the recall of objects with high IoU boxes.
- We provide a new RGB-D image dataset for object proposal, which can be used as a benchmark for the future research.

The rest of the paper is organized as follows. Section 2 provides a brief review of the related work. Section 3 describes the details of the proposed method. Section 4 shows the performance evaluation of the proposed method. Finally, the paper is concluded in Sect. 5.

2 Related Work

The strategies of the existing object proposal methods can be roughly classified into three categories: window scoring, grouping and integration of them.

Window Scoring. Window scoring based methods generate a pool of candidate windows and score the windows by their probabilities of containing an object with objectness measurements. Alexe *et al.* [1] first propose an objectness measurement based on a variety of appearance and geometry properties. Cheng *et al.* [12] utilize binarized normed gradient by training a linear classifier over edge features. Zitnick *et al.* [13] use edge cue to guide window refinement, which can be specially optimized for different IoU thresholds. Xu *et al.* [11] explore the effectiveness of depth cue in handling complex scenes. Overall, window scoring based methods can efficiently generate the bounding boxes as proposal results, but their performance under high IoU is usually limited.

Grouping. Grouping based methods over-segment the images into tiny parts, such as super-pixels, and merge the segments to generate the candidates of objects. Carreira *et al.* [14] use constrained parametric mincuts in merging by several different seeds and multiple features, and Humayun *et al.* [15] improve it by applying multiple graph cut segmentations and using edge detectors. Uijlings *et al.* [16] propose selective search method to merge super-pixels greedily. Rantalankila *et al.* [17] propose a similar merging strategy with different features in similarity measurement. Xiao *et al.* [18] extend selective search by specializing merging in high-complexity scenarios, and Wang *et al.* [19] improve it with multi-branch hierarchical segmentation. Manen *et al.* [20] use randomised super-pixel connectivity graph during merging. Long *et al.* [21] utilize bottom-up merging

to generate initial object candidates, and train a supervised descent model to greedily adjust the boxes. Arbelaez *et al.* [22] perform hierarchical segmentation and multiscale combinatorial grouping. Krähenbühl *et al.* [23] judiciously place object-like seeds and identify critical level sets in geodesic distance transforms as object proposal results. Overall, grouping based methods can generate accurate bounding boxes as well as object boundaries, especially under high IoU, but they are usually inefficient due to bottom-up merging.

Integration of Window Scoring and Grouping. It is interesting to integrate window scoring and grouping strategies together, for example, utilizing the object proposal result by window scoring strategy as the input of further grouping. Chen *et al.* [10] first apply this strategy in object proposal to achieve accurate bounding boxes while retaining high efficiency, but their method only focuses on RGB images and completely ignores depth cue.

3 Elastic Edge Boxes

3.1 Initial Bounding Boxes Generation

We first generate the initial bounding boxes by window scoring strategy. In the proposed approach, we utilize edge boxes method [13], which can efficiently detect the approximate locations of most objects by exploiting edge cue.

Edge boxes first generate candidate objects utilizing sliding window approach, and then scoring the boxes according to the number of contours completely inside each box which is highly indicative of the possibility of the box including an object. Score of box b_k is defined using:

$$score(b_k) = \frac{\sum_i \rho_k(e_i)\hat{m}_i}{2(w_k + h_k)^\eta} - \frac{\sum_{p \in b^{ct}} m_p}{2(w_k^{ct} + h_k^{ct})^\eta}, \tag{1}$$

where w_k and h_k are width and height of the box b_k; b_k^{ct} is a box centered in b_k with size $w_k^{ct} \times h_k^{ct}$ which equal $w_k/2$ and $h_k/2$ respectively; $\eta = 1.5$ is a parameter to offset the bias of larger windows generally containing more edges; m_p represents the edge magnitude of each pixel and $\hat{m}_i$ is obtained by summing up edge magnitude of each pixel in the ith edge group e_i enclosed by box b_k; ρ_k equals zero if e_i overlaps b_k's boundaries. Finally, non-maximal suppression is performed for the boxes to decrease the candidate number.

Though its performance under high IoU is not satisfactory, edge boxes can achieve high recall under low IoU. It means that edge boxes can detect the approximate location of objects but cannot provide bounding boxes with high accuracy. Hence, we adjust the initial bounding boxes to provide more accurate object proposal results.

3.2 Elastic Range Determination

Based on the initial object proposal result, we further determine the elastic range for each bounding box. Obviously, too small elastic range will limit the

adjustment and prevent from providing accurate bounding boxes, while too large elastic range may cause high computational cost and reduce the effect of initial proposal results. We represent images with super-pixels and utilize super-pixel as the basic operation unit in bounding boxes adjustment because super-pixel can well describe object boundaries, increase the robustness to depth map inaccuracy, and reduce the computational cost.

We represent an image as a set of super-pixels $S = \{s_1, \ldots, s_N\}$ which are generated by [24]. Given an initial bounding box b_k, we define $S_{in}^{b_k}$ as a set of super-pixels which are completely inside b_k (cyan ones in Fig. 2(c)), $S_{out}^{b_k}$ as a set of super-pixels which are completely outside b_k, and $S_e^{b_k}$ as a set of the rest super-pixels which are crossed by b_k (yellow ones in Fig. 2(c)). In our method, $S_e^{b_k}$ is used as elastic range.

For the number of super-pixels in $S_{in}^{b_k}$ and $S_{out}^{b_k}$ are usually unbalanced, we select a subset $\hat{S}_{out}^{b_k}$ of $S_{out}^{b_k}$ (blue ones in Fig. 2(c)) to avoid bias in bounding box adjustment. We sort the super-pixels in $S_{out}^{b_k}$ in ascending order according to their weights of minimum center distances to the super-pixels $S_{in}^{b_k}$. To each super-pixel s_i in $S_{out}^{b_k}$, its weight is calculated as:

$$\omega(s_i) = \arg \min_{s_j \in S_{in}^{b_k}} \big(dis(s_i, s_j)\big), \tag{2}$$

where $dis(,)$ denotes the distance between the centers of two super-pixels.

Then we select the super-pixels in $\hat{S}_{out}^{b_k}$ according to their weights, and the number of selected super-pixels is required similar to the number of super-pixels in $S_{in}^{b_k}$:

$$\frac{1}{\lambda} \mid S_{in}^{b_i} \mid \leq \mid \hat{S}_{out}^{b_i} \mid \leq \lambda \mid S_{in}^{b_i} \mid \tag{3}$$

where λ is the parameter which equals to 1.25 in our experiments.

3.3 Bounding Box Adjustment

As shown in Fig. 2(d), to each super-pixel in elastic range $S_e^{b_k}$, we calculate its similarities to all the super-pixels in $S_{in}^{b_k}$ and $\hat{S}_{out}^{b_k}$, and determine whether it should be included in the bounding box. Here, we utilize both color channel and depth channel of an RGB-D image, and define four decision parameters φ_{in}^c, φ_{in}^d, φ_{out}^c and φ_{out}^d as follows:

$$\varphi_{in}^c = \sum_{s_j \in S_{in}^{b_k}} sim^c(s_i, s_j), \quad \varphi_{in}^d = \sum_{s_j \in S_{in}^{b_k}} sim^d(s_i, s_j), \tag{4}$$

$$\varphi_{out}^c = \sum_{s_l \in \hat{S}_{out}^{b_k}} sim^c(s_i, s_l), \quad \varphi_{out}^d = \sum_{s_l \in \hat{S}_{out}^{b_k}} sim^d(s_i, s_l), \tag{5}$$

where $sim^c(,)$ denotes the average color similarity of two super-pixels in HSV space, and $sim^d(,)$ denotes the depth similarity of two super-pixels.

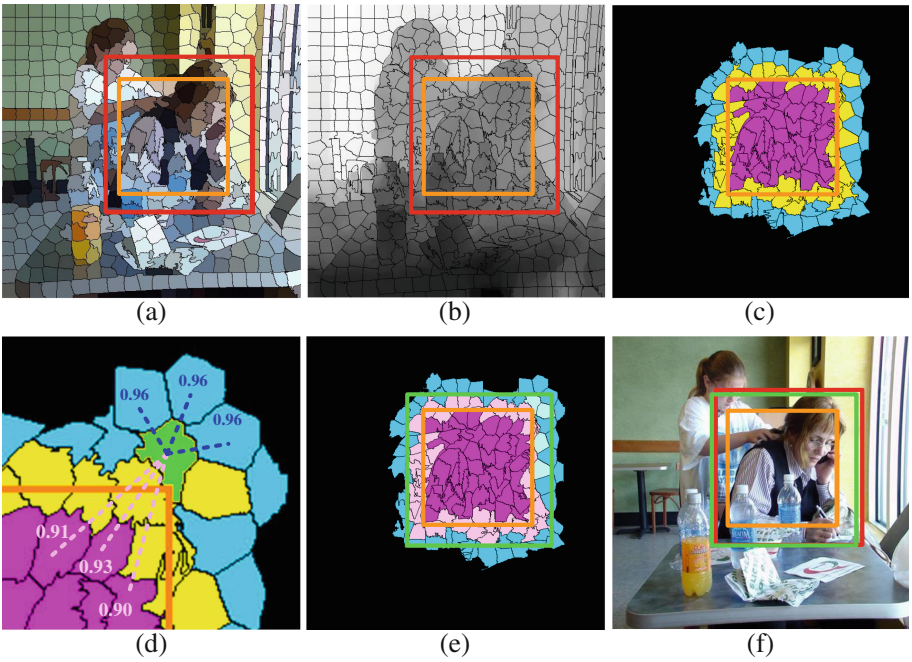

Fig. 2. Bounding box adjustment. (a) and (b) Color and depth channels of RGB-D image with ground truth (red box) and initial bounding box (orange box). (c) Elastic range (yellow super-pixels). (d) Details of decision in bounding box adjustment. (e) Adjusted bounding box (green box). (f) Final bounding boxes (green and orange boxes) and ground truth (red box) (Color figure online).

Based on the four parameters, we extend $S_{in}^{b_k}$ by adding the super-pixels satisfying the following requirements:

$$S_{in}^{b_k}{}^{*} = S_{in}^{b_k} \cup \left\{ s_i \in S_e^{b_i} \mid \varphi_{in}^{c} > \varphi_{out}^{c} \; and \; \varphi_{in}^{d} > \varphi_{out}^{d} \right\}.$$

Based on the extended super-pixel set $S_{in}^{b_k}{}^{*}$, we generate a new bounding box $\tilde{b}_i$ (green box in Fig. 2(e)). By adjusting each bounding box, we can obtain the final object proposal result as follows:

$$B^{*} = B \cup \left\{ \tilde{b}_i \mid \forall b_i \in B \; and \; b_i \neq \tilde{b}_i \right\}, \tag{6}$$

where B is the initial object proposal result.

4 Experiments

4.1 Dataset Construction

To validate the performance of our method, we construct a new RGB-D image dataset, named *NJU1500*, by extending the stereo objectness dataset [11].

Stereo objectness dataset provided in [11] is an RGB-D image dataset including 1,032 stereo images. As far as we know, it is the only RGB-D image dataset for object proposal. However, with the analysis of stereo objectness dataset, we find that it has some obvious drawbacks. We divide the images in stereo objectness dataset into six groups according to their object numbers, including 1, 2, 3, 4, 5, and 5+ (more than five), and the image numbers of the groups are 90, 417, 251, 133, 66, 75, respectively. It is easy to find that the distribution of object number among images is not balanced, which may lead to the bias in evaluation. Moreover, the numbers of objects contained in nearly half of the images are no more than 2, and the average number of objects per image is only 2.98, which makes object proposal task on it less challenging.

To overcome the drawbacks of stereo objectness dataset, we construct a new RGB-D image dataset named *NJU1500* based on it. To keep the balance of object number distribution among images and increase the average number of objects per image, we remove all the images with one object and a part of images with two objects, and supplement the images containing more than two objects. The selection of images with two objects only depends on their identifier in stereo objectness dataset, and the images of large identifiers are removed. 825 images are retained from the 1,032 images of stereo objectness dataset, and 675 images are supplemented. The supplemented images are collected from several 3D movies and videos, and the depth maps are calculated with Sun's optical flow method [25]. Similar to stereo objectness dataset construction, we annotate the ground truths of object locations according to PASCAL VOC2007 annotation guidelines. Five participants, including three males and two females, are invited to annotate the object bounding boxes for each supplementary image. The final constructed dataset includes five groups with 300 images per group, and the average number of objects per image increases from 2.98 to 4.22.

4.2 Performance Evaluation

We validate the performance of our method on *NJU1500* dataset. All the experiments are carried out on a computer with Intel i5 2.8 GHz CPU and 8 GB memory.

Figure 3 shows some examples of object proposal results generated by our method. The best bounding boxes to each ground truth within top 1,500 of each image are marked with green bounding boxes, and the IoU values of the bounding boxes to their corresponding ground truths are indicated with yellow numbers. We can find that almost all the objects are detected by our method with high IoU values, including obscure objects (such as the sword in the fourth image of the first row), small objects (such as the blue and red dustbins in the fifth image of the second row) and occluded objects (such as Papa Smurf in the third image of the third row).

We compare our method with the state-of-the-art methods, including binarized normed gradients (BING) [12], edge boxes (EB) [13], objectness (OBJ) [1], geodesic object proposal (GOP) [23], multiscale combinatorial grouping (MCG)

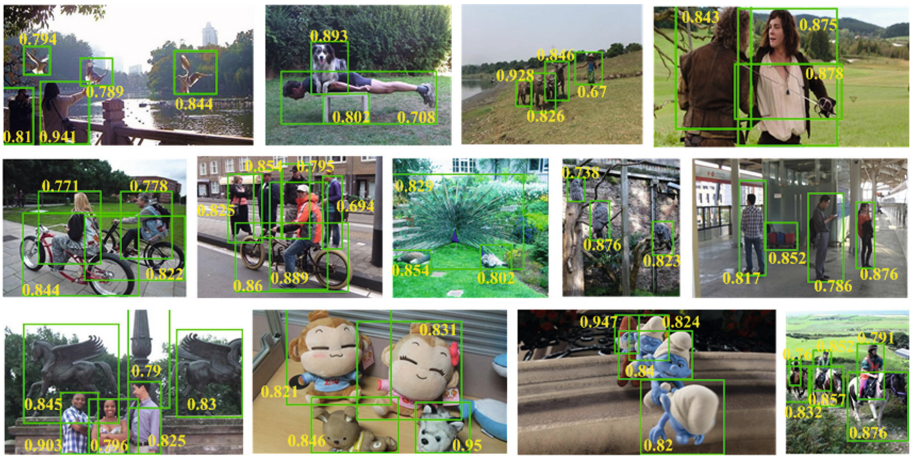

Fig. 3. Examples of object proposal results generated by our method (Color figure online).

[22], and selective search (SS) [25]. These methods of window scoring strategy or grouping strategy have excellent performance in object proposal [9]. In addition, we compare our method with two latest and somewhat similar methods, adaptive integration of depth and color (AIDC) [11] and multi-thresholding straddling expansion of edge boxes (M-EB) [10]. The former is also proposed for object proposal on RGB-D images, and the latter also utilizes integration proposal strategy for RGB images.

Accuracy. We validate the proposal accuracy with three criteria. Figure 4(a) and (b) show the recall vs. proposal number curves of all the methods when IoU equals 0.5 and 0.8, respectively. We can find that our method has similar proposal accuracy to the existing best methods when IoU equals 0.5, and it outperforms all the methods when IoU equals 0.8. It means that our method can handle object proposal requirements under both low and high IoU. Figure 4(c) shows the average recall (AR) vs. proposal number curve [9]. It is found that our method outperforms all the other methods except MCG, which is more than 10 times slower than our method as shown in Table 2. Figure 4(d) shows the recall vs. IoU curve. We can find that our method outperform the other methods when IoU is in range of [0.5, 0.8] and it is only worse than GOP, MCG and SS when IoU is larger than 0.8. However, all these three methods use grouping strategy and have low efficiency.

Table 1 provides more details of comparison results, in which "prop" denotes the proposal number, "0.5-DR" and "0.8-DR" denote the recall when IoU equals 0.5 and 0.8, and "AR" denotes average recall. It shows that our method is only slightly worse than EB in average recall when proposal number equals 500 but it has the best recall and average recall with different proposal numbers under all the other conditions.

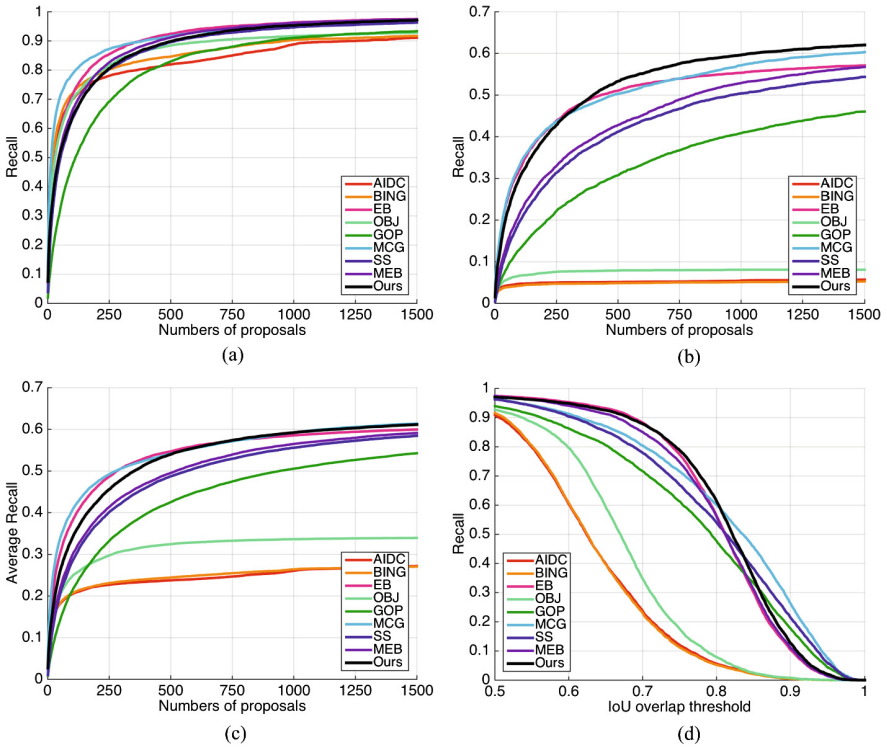

Fig. 4. Comparison of our method with the state-of-the-art methods. (a) and (b) Recall vs. proposal number curves when IoU equals 0.5 and 0.8. (c) Average recall vs. proposal number curve. (d) Recall vs. IoU curve with 1,500 bounding boxes.

Table 1. Comparison of our method and the state-of-the-art methods with different proposal numbers.

Method	Type	#prop=500			#prop=1000			#prop=1500		
		0.5-DR	0.8-DR	AR	0.5-DR	0.8-DR	AR	0.5-DR	0.8-DR	AR
AIDC	scoring	0.82	0.05	0.24	0.89	0.05	0.26	0.91	0.06	0.27
BING	scoring	0.85	0.05	0.24	0.90	0.05	0.26	0.92	0.05	0.27
EB	scoring	**0.92**	0.51	**0.55**	**0.96**	0.55	**0.59**	**0.98**	0.57	0.60
OBJ	scoring	0.90	0.08	0.32	0.92	0.08	0.34	0.93	0.08	0.34
GOP	grouping	0.83	0.30	0.43	0.91	0.41	0.51	0.93	0.46	0.54
MCG	grouping	**0.92**	0.50	0.54	0.95	0.57	**0.59**	0.97	0.60	**0.61**
SS	grouping	0.90	0.41	0.49	0.95	0.50	0.56	0.96	0.54	0.59
M-EB	integration	0.91	0.43	0.50	**0.96**	0.53	0.57	0.97	0.57	0.60
Ours	integration	**0.92**	**0.53**	0.54	**0.96**	**0.60**	**0.59**	**0.98**	**0.62**	**0.61**

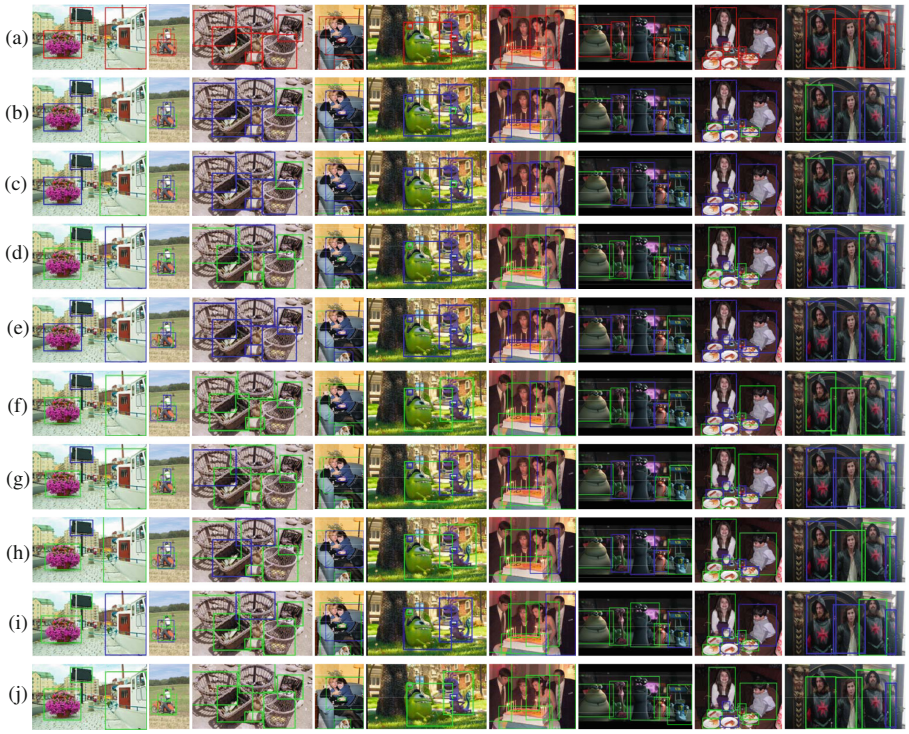

Fig. 5. Examples of object proposal results with different methods. (a) Original image with ground truth. (b)-(j) Object proposal results of AIDC [11], BING [12], EB [13], OBJ [1], GOP [23], MCG [22], SS, M-EB [10] and our method (Color figure online).

Figure 5 shows some examples of object proposal results generated by different methods. In the examples, red boxes indicate the ground truths, green boxes indicate the bounding boxes generated by different methods, and blue boxes indicate the missed ground truths when IoU equals 0.8. Though the structure of some images is complex and some objects are inconspicuous, our method can detect almost all the objects with high IoU.

Speed. We also compare the efficiency of all the methods. Table 2 presents the running time of all methods. Though some methods require much less time than other methods processing an image, such as BING and AIDC, they are obviously worse than our method in proposal accuracy (Fig. 4). The three methods, which have better proposal accuracy than our method when IoU is larger than 0.8, are obviously worse than our method in efficiency. The most efficient method among them, that is SS, is about 10 % slower than our method, and MCG is even 10 times slower than our method.

Table 2. Comparison of our method and the state-of-the-art methods in running time.

Method	Type	Language	Time (s)
AIDC	window	C++	0.07
BING	window	C++	0.06
EB	window	C++ & Matlab	0.69
OBJ	window	C++ & Matlab	4.13
GOP	grouping	C++ & Matlab	7.25
MCG	grouping	C++ & Matlab	60.12
SS	grouping	C++ & Matlab	6.39
M-EB	integration	C++ & Matlab	0.99
Ours	integration	C++ & Matlab	5.78

5 Conclusions

In this paper, we propose an object proposal method for RGB-D images by integrating window scoring and grouping strategies. The method generates the initial bounding boxes by an efficient edge-based window scoring method, and adjusts the bounding boxes by grouping the super-pixels in elastic range, which improves proposal accuracy while retaining high efficiency. Moreover, the effectiveness of depth cue is explored as well as color cue, which benefits to handle the images with complex scenes. The experiments show that our method can effectively and efficiently generate the bounding boxes with high IoU, and it outperforms state-of-the-art object proposal methods considering both accuracy and efficiency.

Acknowledgments. This work is supported by the National Science Foundation of China (61321491, 61202320), Research Project of Excellent State Key Laboratory (61223003), National Undergraduate Innovation Project (G1410284074) and Collaborative Innovation Center of Novel Software Technology and Industrialization.

References

1. Alexe, B., Deselaers, T., Ferrari, V.: Measuring the objectness of image windows. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2189–2202 (2012)
2. Li, T., Chang, H., Wang, M., Ni, B., Hong, R., Yan, S.: Crowded scene analysis: a survey. IEEE Trans. Circ. Syst. Video Technol. **25**(3), 367–386 (2015)
3. Sang, J., Changsheng, X., Liu, J.: User-aware image tag refinement via ternary semantic analysis. IEEE Trans. Multimedia **14**(3–2), 883–895 (2012)
4. Xu, X., Geng, W., Ju, R., Yang, Y., Ren, T., Wu, G.: OBSIR: object-based stereo image retrieval. In: IEEE International Conference on Multimedia and Expo, pp. 1–6 (2014)
5. Bao, B.-K., Liu, G., Hong, R., Yan, S., Changsheng, X.: General subspace learning with corrupted training data via graph embedding. IEEE Trans. Image Process. **22**(11), 4380–4393 (2013)

6. Hong, C., Zhu, J., Yu, J., Cheng, J., Chen, X.: Realtime and robust object matching with a large number of templates. Multimedia Tools Appl., 1–22 (2014)
7. Ren, T., Qiu, Z., Liu, Y., Tong, Y., Bei, J.: Soft-assigned bag of features for object tracking. Multimedia Syst. **21**(2), 189–205 (2015)
8. Tang, J., Tao, D., Qi, G.-J., Huet, B.: Social media mining and knowledge discovery. Multimedia Syst. **20**(6), 633–634 (2014)
9. Hosang, J., Benenson, R., Dollár, P., Schiele, B.: What makes for effective detection proposals? arXiv preprint. arXiv:1502.05082 (2015)
10. Chen, X., Ma, H., Wang, X., Zhao, Z.: Improving object proposals with multi-thresholding straddling expansion. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)
11. Xu, X., Ge, L., Ren, T., Wu, G.: Adaptive integration of depth and color for objectness estimation. In: IEEE International Conference on Multimedia and Expo (2015)
12. Cheng, M.-M., Zhang, Z., Lin, W.-Y., Torr, P.: BING: Binarized normed gradients for objectness estimation at 300fps. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3286–3293 (2014)
13. Zitnick, C.L., Dollár, P.: Edge boxes: locating object proposals from edges. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part V. LNCS, vol. 8693, pp. 391–405. Springer, Heidelberg (2014)
14. Carreira, J., Sminchisescu, C.: Constrained parametric min-cuts for automatic object segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3241–3248 (2010)
15. Humayun, A., Li, F., Rehg, J.M.: Rigor: reusing inference in graph cuts for generating object regions. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 336–343 (2014)
16. Uijlings, J.R.R., van de Sande, K.E.A., Gevers, T., Smeulders, A.W.M.: Selective search for object recognition. Int. J. Comput. Vis. **104**(2), 154–171 (2013)
17. Rantalankila, P., Kannala, J., Rahtu, E.: Generating object segmentation proposals using global and local search. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2417–2424 (2014)
18. Xiao, Y., Lu, C., Tsougenis, E., Lu, Y., Tang, C.-K.: Complexity-adaptive distance metric for object proposals generation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 778–786 (2015)
19. Wang, C., Zhao, L., Liang, S., Zhang, L., Jia, J., Wei, Y.: Object proposal by multi-branch hierarchical segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)
20. Manen, S., Guillaumin, M., Van Gool, L.: Prime object proposals with randomized prims algorithm. In: IEEE International Conference on Computer Vision, pp. 2536–2543 (2013)
21. Long, C., Wang, X., Hua, G., Yang, M., Lin, Y.: Accurate object detection with location relaxation and regionlets re-localization. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9003, pp. 260–275. Springer, Heidelberg (2015)
22. Arbelaez, P., Pont-Tuset, J., Barron, J., Marques, F., Malik, J.: Multiscale combinatorial grouping. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 328–335 (2014)
23. Krähenbühl, P., Koltun, V.: Geodesic object proposals. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part V. LNCS, vol. 8693, pp. 725–739. Springer, Heidelberg (2014)

24. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012)
25. Sun, D., Roth, S., Black, M.J.: Secrets of optical flow estimation and their principles. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2432–2439 (2010)