Object Proposal via Depth Connectivity Constrained Grouping

Yuantian Wang[1], Lei Huang[1], Tongwei Ren[1,*], Sheng-Hua Zhong[2], Yan Liu[3], and Gangshan Wu[1]

[1] State Key Laboratory for Novel Software Technology, Nanjing University, China
[2] College of Computer Science and Software Engineering, Shenzhen University, China
[3] Computing Department, The Hong Kong Polytechnic University, Hong Kong, China
wangyt@smail.nju.edu.cn,{leihuang,rentw}@nju.edu.cn,csshzhong@szu.edu.cn, csyliu@comp.polyu.edu.hk,gswu@nju.edu.cn

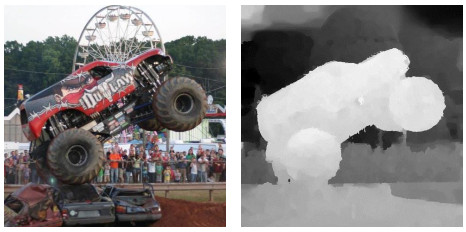
Abstract. Object proposal aims to detect category-independent object candidates with a limited number of bounding boxes. In this paper, we propose a novel object proposal method on RGB-D images with the constraint of depth connectivity, which can improve the key techniques in grouping based object proposal effectively, including segment generation, hypothesis expansion and candidate ranking. Given an RGB-D image, we first generate segments using depth aware hierarchical segmentation. Next, we combine the segments into hypotheses hierarchically on each level, and further expand these hypotheses to object candidates using depth connectivity constrained region growing. Finally, we score the object candidates based on their color and depth features, and select the ones with the highest scores as the object proposal result. We validated the proposed method on the largest RGB-D image data set for object proposal, and our method is superior to the state-of-the-art methods.

Keywords: Object proposal, RGB-D image, depth connectivity, constrained grouping

1 Introduction

Object proposal aims to indicate the positions of category-independent object candidates in a given image with bounding boxes [1]. It can be used as a fundamental of numerous multimedia applications, such as object recognition [11], segmentation [14], tracking [18], image annotation [19], saliency analysis [16] and information retrieval [24]. Two paradigms are mainly used in current object proposal methods, named window scoring and grouping [26]. The former samples bounding boxes in a given image, measures the probability of each candidate box in containing an object, *i.e.*, "*objectness*", and selects the boxes with the highest objectness scores as object candidates; the latter over-segments a given image into amounts of segments, and groups these segments into object candidates, which probably enclose objects.

A key challenge in object proposal is the diversity and complexity of object appearance in color representation. Figure 1 shows an example. The



(a) (b)

Fig. 1. An example of the difference between color appearance and depth appearance in objectness representation. The monster truck is easier to distinguish in depth appearance (b) than in color appearance (a).

monster truck in Fig. 1(a) has complicate color appearance, which is difficult to distinguish from the complex scene. In contrast, it can be easily identified from depth appearance in Fig. 1(b), because its surface is connected in depth while its boundary is disconnected from background. We can see that depth provides a powerful cue in detecting object candidates in RGB-D images [23]. However, RGB-D images usually suffer from low quality problem on depth appearance, which is caused by the limitation of capture devices and estimation algorithms, including inaccurate boundary and serious noise. It hampers the performance of object proposal methods, especially for the ones using pixel-level features, such as edges [12]. Compared to window scoring used in most existing object proposal methods on RGB-D images, grouping strategy has its natural advantage in combining depth cue into object proposal, because they work on region level necessarily. It helps to improve the robustness in handling low quality depth.

In this paper, we propose a novel object proposal method on RGB-D images with the constraint of *depth connectivity*. It can improve the key techniques in grouping based object proposal effectively, namely segment generation, hypothesis expansion and candidate ranking. Figure 2 shows an overview of the proposed method. We first generate segments using depth aware hierarchical segmentation on ultra-metric contour map. Next, we combine the segments into hypotheses hierarchically, and further expand these hypotheses to object candidates with depth connectivity constrained region growing. Finally, we score the object candidates based on their color and depth features, and select the ones with the highest scores as the object proposal result. We validate the proposed method on the largest public RGB-D image data set for object proposal, named *NJU1800*. Our method is superior the state-of-the-art methods.

Our contributions mainly include:

- We define the depth connectivity between two segments, and utilize it to measure inner depth connectivity and boundary depth connectivity of an object candidate.
- We propose an object proposal method on RGB-D images with the constraint of depth connectivity, which improves hierarchical segmentation, hypothesis expansion and candidate ranking in grouping.
- We validate our method on the largest RGB-D image data set for object proposal, and our method outperforms the state-of-the-art methods.

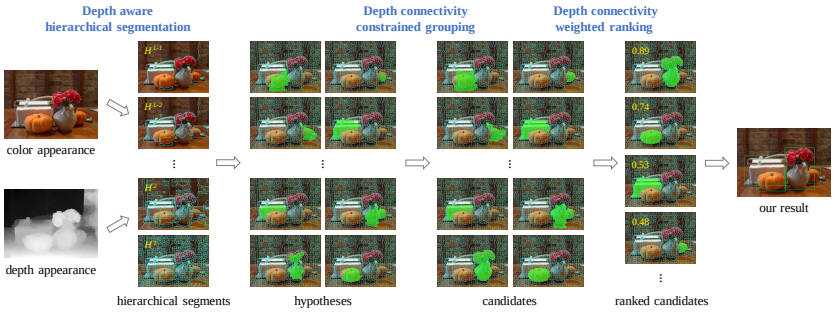


Fig. 2. An overview of our proposed method. To a given RGB-D image, we first over-segment it into segments. Next, we combine the segments into hypotheses and further expand the hypotheses to object candidates. Finally, we score the object candidates and select the ones with the highest scores as our result.

2 Related Work

2.1 Object proposal on RGB images

Two paradigms are mainly used in the existing object proposal methods, named window scoring and grouping.

Window scoring based methods focus on the objectness measurement of the sampled bounding boxes. Some hand-crafted features were proposed to measure objectness, including object location and geometry properties [1], structured edge [26] and binarized normed gradient [7]. The boxes were ranked based on the extracted features, and the ones with the highest scores are selected as object candidates. Window scoring based methods are usually efficient, but they are hard to generate accurate candidates due to the limitation of discrete sampling.

In comparison, grouping based methods focus on segment generation and grouping. Carreira *et al.* [4] utilized constrained parametric mincuts and merged them based on object features, which was improved by applying edge detectors and multiple gragh cut segmentations [9] . Uijlings *et al.* [20] proposed selective search algorithm to merge similar super-pixels greedily, which could benefit from the combination with multiple features [17], multi-branch hierarchical segmentation [21], and region merging in high-complexity scenarios [22]. Manen *et al.* [15] merged randomized super-pixel connectivity graph with learned features. Arbeláez *et al.* [3] utilized multiscale hierarchical segmentation and combinatorial grouping with Pareto front model. Krähenbühl *et al.* [10] set object-like seeds and used classifiers in geodesic transform as object proposal results. Grouping based methods can provide more accurate candidates, but they usually suffer from low efficiency problem caused by iterative grouping.

A proposal refinement strategy was proposed in [6], which refined object candidates generated by different object proposal methods. The integration of window scoring based methods and the refinement strategy can obtain a good trade-off between proposal accuracy and efficiency [13].

2.2 Object proposal on RGB-D images

Object proposal methods on RGB-D images mainly focus on exploiting the effect of depth cue and integrating it into object proposal methods on RGB images. Xu *et al.* [23] first brought depth into objectness measurement by adaptively integrating color gradient and depth gradient. Liu *et al.* [12] detected multi-layered structured edges by decomposing the sparse edge map according to the corrected depth map, and ranked the bounding boxes with its maximum scores on all the depth layers. Liu *et al.* [13] generated bounding boxes by edge boxes method [26] and refined them through repartioning the super-pixels on their boundaries. Zhang *et al.* [25] provided a proposal refinement strategy with multiple trained high-level features, including CNN feature, depth geometric feature and semantic context feature.

The exiting object proposal methods on RGB-D images concentrate on extending windows grouping based methods and refinement strategies, but ignore the improvement of grouping based methods, which may impede them from generating object candidates with high accuracy.

3 Our Method

3.1 Depth connectivity measurement

Depth connectivity is the basic concept in our method, which is utilized to improve the performance of the key procedures. In this subsection, we introduce the measurement of depth connectivity.

In grouping based object proposal methods, a given image is first over-segmented into many segments. Assume s_i and s_j are two segments in the given image, and the average depths of all the pixels within them are d_i and d_j, respectively. Here, depth is normalized to the value range of $[0, 1]$, and larger depth value means image content is nearer. If s_i and s_j are adjacent, their depth connectivity $\phi_{i,j}$ is defined as:

$$\phi_{i,j} = 1 - |d_i - d_j|. \tag{1}$$

If s_i and s_j are not adjacent ,but they belong to a segment combination, their depth connectivity is defined as:

$$\phi_{i,j} = \max_{p_k \in \mathrm{P_{i,j}}} \varphi_k, \tag{2}$$

where $\mathrm{P_{i,j}}$ is the set of all the connected paths between s_i and s_j within the segment combination; φ_k is the depth connectivity of p_k. Let $p_k : s_i \to s_{k_1} \to \cdots \to s_{k_{N_k}} \to s_j$, where N_k is the number of segments in p_k except s_i and s_j, φ_k is calculated as:

$$\varphi_k = \min\{\phi_{i,k_1}, \cdots, \phi_{k_{N_k},j}\}. \tag{3}$$

From Eq. (1)-(3), we can see that depth connectivity between two segments is in the value range of $[0, 1]$. Larger depth connectivity value means two segments are more connected in depth.

Based on depth connectivity between two segments, we further define inner depth connectivity and boundary depth connectivity of an object candidate. To an object candidate c, its inner depth connectivity is measured as follows:

$$\psi^{in} = \min_{s_i, s_j \in \mathrm{S_c}} \phi_{i,j}, \tag{4}$$

where $\mathrm{S_c}$ is the set of all the segments within c. The boundary depth connectivity of c is measured as follows:

$$\psi^{bd} = \frac{1}{|\mathrm{B_c}|} \sum_{s_i \in \mathrm{B_c}} \min_{s_j \in \Omega_i \setminus \mathrm{S_c}} \phi_{i,j}, \tag{5}$$

where $\mathrm{B_c}$ is the set of all the segments in the boundary of c; Ω_i is the set of all the segments surrounding s_i; $|.|$ denotes the cardinality of a set.

3.2 Depth aware hierarchical segmentation

We first generate the ultra-metric contour map using [2], which contains the contours weighted by brightness, color and texture gradients. The regions surrounded by the contours are treated as the segments.

Since all the segments are separated by the contours, there is one and only contour between every two segments. To two segments s_i and s_j, we denote the contour part between them as $e_{i,j}$, and measure its strength as follows:

$$\omega_{i,j} = \lambda \omega^U_{i,j} + (1 - \lambda)(1 - \phi_{i,j}), \tag{6}$$

where $\omega^U_{i,j}$ is the weight of $e_{i,j}$ in the ultra-metric contour map referring to [3]; $\phi_{i,j}$ is the depth connectivity between s_i and s_j; λ is a parameter for linear combination, which equals 0.7 in our experiments. For the value ranges of both $\omega^U_{i,j}$ and $\phi_{i,j}$ are $[0, 1]$, the value range of $\omega_{i,j}$ is $[0, 1]$.

Based on edge strength, we further merge the segments into different hierarchies $\{\mathcal{H}^*, \mathcal{H}^1, \mathcal{H}^2, ..., \mathcal{H}^L\}$ with Platt's method [5]. Here, $\mathcal{H}^*$ is the original segments before merging and $\mathcal{H}^L$ is the whole image. Based on the depth connectivity between two segments, *i.e.*, the item $1-\phi_{i,j}$ in Eq. (6), we can merge the adjacent segments with similar depth values, and prevent the merging of two segments which are not connected in depth but similar in color appearance.

3.3 Depth connectivity constrained grouping

Based on $\{\mathcal{H}^*, \mathcal{H}^1, \mathcal{H}^2, ..., \mathcal{H}^L\}$ generated in hierarchical segmentation, we further generate hypotheses by combining the segments into singletons, pairs, triplets, and four-tuples on $\mathcal{H}^1, \mathcal{H}^2, ..., \mathcal{H}^{L-1}$, respectively. Inspired by [3], the adjacent segments without intersection on different hierarchies are preferred in hypothesis generation, and only the top fixed-number of hypotheses are retained.

Because the hypotheses are usually incomplete as compared to objects, we expand the hypotheses to generate object candidates. Considering the surface of

an object is usually connected in depth, we use a greedy region growing strategy constrained by depth connectivity in hypothesis expansion. Assume Δ_h is the set of all the segments adjacent to a hypothesis h, we expand h iteratively till no segment can be grouped:

$$h^* \leftarrow h \cup \{s_i | s_i \in \Delta_h, \phi_{i,j} \geq \tau\}, \tag{7}$$

where s_j is a segment within h and it is adjacent to s_i; τ is a threshold, which equals 0.95 to avoid over-expansion in our experiments. We expand all the hypotheses and remove the repeated ones. The retained hypotheses after expansion are treated as object candidates.

3.4 Depth connectivity weighted ranking

We score object candidates according to their color and depth features, and select the ones with the highest scores for object proposal.

In object candidate scoring based on color feature, we use a trained maximum marginal relevance model provided by [3], which uses the low-level features including size, location, shape and boundary contour strength.

In object candidate scoring based on depth feature, we use both inner depth connectivity in Eq. (4) and boundary depth connectivity in Eq. (5). A candidate probably containing an object usually has high inner depth connectivity, because the surface of an object is connected in depth, and low boundary depth connectivity, because an object is usually disconnected from background in depth. However, the overemphasis of inner depth connectivity or boundary depth connectivity may degrade the performance of object proposal. Specifically, the overemphasis of inner depth connectivity may cause the preference of partial objects with similar depth, while the overemphasis of boundary depth connectivity may increase the rankings of the combinations of multiple objects with obvious boundaries. Hence, we balance the influences of inner depth connectivity and boundary depth connectivity in scoring object candidates based on depth features:

$$S^d = (\psi^{in})^\gamma - (\kappa(\psi^{bd}, \delta))^\gamma, \tag{8}$$

where γ is a parameter to nonlinearly emphasize high depth connectivity, which equals 4 in our experiments; κ is a function to punish high boundary depth connectivity with a parameter δ, which returns ψ^{bd} when it is smaller than δ, and 1 otherwise; δ equals 0.5 in our experiments. S^d is normalized to the value range of $[0, 1]$.

We combine the scores based on color and depth features linearly to obtain the final score of each object candidate:

$$S = \alpha S^c + (1 - \alpha) S^d, \tag{9}$$

where S^c is the score based on color feature; α is a parameter for combination, which equals 0.5 in our experiments.

Finally, we select the object candidates with the highest scores and generate their bounding boxes as the object proposal result.

4 Experiments

4.1 Data set and experiment settings

We validated our method on the largest public RGB-D data set for object proposal *NJU1800*, which contains 1,800 RGB-D images with manually labelled ground truth [13].

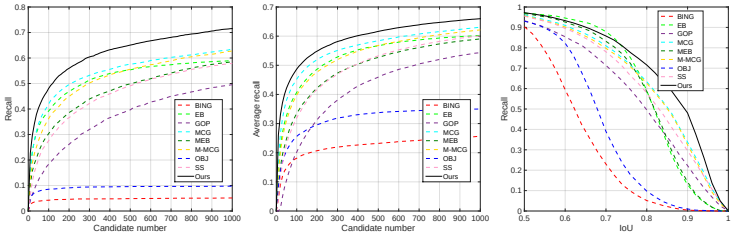
All the experiments were conducted on a computer with Intel i5 2.8GHz CPU and 8GB memory. For all the other methods in comparison, we used their default settings suggested by the authors.

4.2 Experimental results

We first compare our method with eight object proposal methods on RGB images, namely binarized normed gradients (BING) [7], edge boxes (EB) [26], objectness (OBJ) [1], geodesic object proposal (GOP) [10], multiscale combinatorial grouping (MCG) [3], selective search (SS) [20], multi-thresholding straddling expansion of edge boxes (M-EB) and multiscale combinatorial grouping (M-MCG) [6]. Figure 3 shows the comparison results on recall *vs.* candidate number under IoU=0.8, average recall *vs.* candidate number [8] and recall *vs.* IoU on the top 1,000 candidates. It shows that our method outperforms all the methods on RGB images. It illustrates that our exploitation of depth in object proposal is effective, because inappropriate usage of depth will not improve object proposal performance [12].

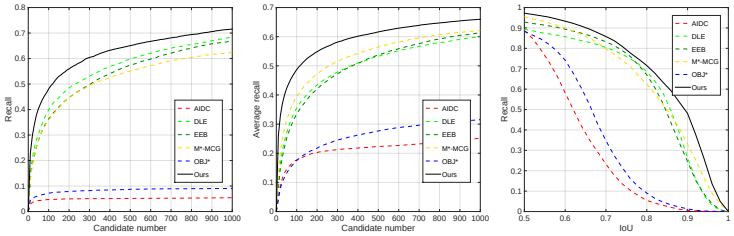
We further compare our method with three object proposal methods on RGB-D images, namely adaptive integration of depth and color (AIDC) [23], depth-aware layered edge (DLE) [12] and elastic edge boxes (EEB) [13]. They can be treated as the extensions of BING, EB and M-EB by integrating depth, respectively. To provide more comprehensive evaluation, we adopt two baselines extended from other two open-source object proposal methods on RGB images, namely OBJ and M-MCG, by referring to [13], and denote them with OBJ* and M*-MCG. Figure 4 shows the comparison results under the same criteria to those in Fig. 3. It shows that our method is superior to other methods on RGB-D images. Figure 5 shows some examples of object proposal results generated by different methods on RGB-D images. The best bounding boxes as compared to the ones in ground truth within the top 1,000 candidates under IoU=0.8 of each image are denoted with green boxes, and the omitted ones in ground truth are denoted with red boxes. we can see that our method can propose all the objects on various images, but other methods fail in some cases.

We also validate the efficiency of our method. Table 1 shows the running time of our method and other methods on RGB-D images. We can see that the running time of our method is similar to other grouping based methods with comparable performance, such as M*-MCG.



(a) (b) (c)

Fig. 3. Comparison with the state-of-the-art methods on RGB images. (a) Curve of recall *vs.* candidate number (IoU = 0.8). (b) Curve of average recall *vs.* candidate number. (c) Curve of recall *vs.* IoU on the top 1,000 candidates.



(a) (b) (c)

Fig. 4. Comparison with the state-of-the-art methods on RGB-D images. (a) Curve of recall *vs.* candidate number (IoU = 0.8). (b) Curve of average recall *vs.* candidate number. (c) Curve of recall *vs.* IoU on the top 1,000 candidates.

Table 1. Efficiency evaluation of different methods on RGB-D images.

Method	Type	Language	Time per image (s)
AIDC	window	C++	0.08
DLE	window	C++ & Matlab	4.51
EEB	integration	C++ & Matlab	22.34
OBJ*	window	C++ & Matlab	4.19
M*-MCG	grouping	C++ & Matlab	60.41
Ours	grouping	C++ & Matlab	67.53

4.3 Discussion

In our experiments, we find some limitations of our method. For instance, our method may omit some objects in an image containing multiple objects with complex scene, such as the cups and two children in the top example in Fig. 6. Moreover, as shown in the bottom example in Fig. 6, our method fail in providing the accurate bounding boxes when the depth of two aircrafts are partially inaccurate.

5 Conclusion

In this paper, we proposed an object proposal method on RGB-D images with the constraint of depth connectivity. Specifically, depth connectivity is used to

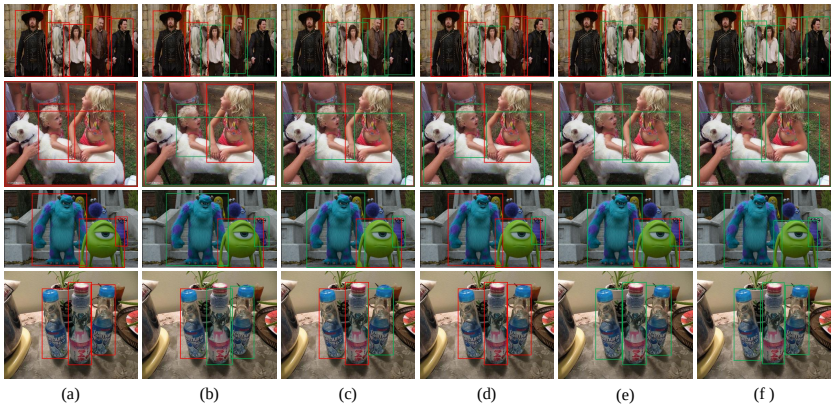


Fig. 5. Examples of object proposal results using different methods on RGB-D images. All the green boxes denote the best bounding boxes to the ones in ground truth within the top 1,000 candidates under IoU=0.8, and the red boxes denote the omitted ones in ground truth. (a) AIDC. (b) DLE. (c) EEB. (d) OBJ*. (e) M*-MCG. (f) Ours.

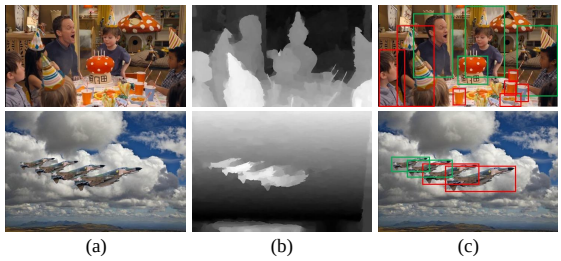


Fig. 6. Examples of our failure results. (a) Color appearance. (b) Depth appearance. (c) Our results (the green boxes and the red boxes have the same denotation to Fig. 5).

improve the key techniques in grouping based object proposal, including segment generation, hypothesis expansion and candidate ranking. The proposed method was validated on the largest RGB-D image data set for object proposal *NJU1800*, and the experimental results showed that it outperforms the state-of-the-art methods on both RGB images and RGB-D images.

6 Acknowledgements

This work is supported by National Science Foundation of China (61321491, 61202320), and Collaborative Innovation Center of Novel Software Technology and Industrialization.

References

1. Alexe, B., Deselaers, T., Ferrari, V.: Measuring the objectness of image windows. TPAMI 34, 2189–2202 (2012)

2. Arbeláez, P.: Boundary extraction in natural images using ultrametric contour maps. In: CVPR Workshop. pp. 182–182 (2006)
3. Arbeláez, P., Pont-Tuset, J., Barron, J.T., Marques, F., Malik, J.: Multiscale combinatorial grouping. In: CVPR. pp. 328–335 (2014)
4. Carreira, J., Sminchisescu, C.: Constrained parametric min-cuts for automatic object segmentation. In: CVPR. pp. 3241–3248 (2010)
5. Chapelle, O., Vapnik, V., Bousquet, O., Mukherjee, S.: Choosing multiple parameters for support vector machines. Machine Learning 46(1), 131–159 (2002)
6. Chen, X., Ma, H., Wang, X., Zhao, Z.: Improving object proposals with multi-thresholding straddling expansion. In: CVPR. pp. 2587–2595 (2015)
7. Cheng, M.M., Zhang, Z., Lin, W.Y., Torr, P.H.S.: Bing: Binarized normed gradients for objectness estimation at 300fps. In: CVPR. pp. 3286–3293 (2014)
8. Hosang, J., Benenson, R., Dollár, P., Schiele, B.: What makes for effective detection proposals? TPAMI 38(4), 814–830 (2016)
9. Humayun, A., Li, F., Rehg, J.M.: Rigor: Reusing inference in graph cuts for generating object regions. In: CVPR. pp. 336–343 (2014)
10. Krähenbühl, P., Koltun, V.: Geodesic object proposals. In: ECCV. pp. 725–739 (2014)
11. Li, X., Jiang, S., Lv, X., Chen, C.: Learning to recognize hand-held objects from scratch. In: PCM. pp. 527–539. Springer (2016)
12. Liu, J., Ren, T., Bao, B.K., Bei, J.: Depth-aware layered edge for object proposal. In: ICME. pp. 1–6 (2016)
13. Liu, J., Ren, T., Wang, Y., Zhong, S.H., Bei, J., Chen, S.: Object proposal on rgb-d images via elastic edge boxes. NEUCOM 236, 134–146 (2017)
14. Luo, B., Li, H., Meng, F., Wu, Q., Huang, C.: Video object segmentation via global consistency aware query strategy. TMM PP(99), 1–1 (2017)
15. Manen, S., Guillaumin, M., Gool, L.J.V.: Prime object proposals with randomized prim's algorithm. In: ICCV. pp. 2536–2543 (2013)
16. Qi, F., Zhao, D., Liu, S., Fan, X.: 3d visual saliency detection model with generated disparity map. Multimedia Tools and Applications 76(2), 3087–3103 (2017)
17. Rantalankila, P., Kannala, J., Rahtu, E.: Generating object segmentation proposals using global and local search. In: CVPR. pp. 2417–2424 (2014)
18. Ren, T., Qiu, Z., Liu, Y., Yu, T., Bei, J.: Soft-assigned bag of features for object tracking. MMSJ 21(2), 189–205 (2015)
19. Sang, J., Xu, C., Liu, J.: User-aware image tag refinement via ternary semantic analysis. IEEE Transactions on Multimedia 14(3), 883–895 (2012)
20. Uijlings, J.R.R., van de Sande, K.E.A., Gevers, T., Smeulders, A.W.M.: Selective search for object recognition. IJCV 104(2), 154–171 (2013)
21. Wang, C., Zhao, L., Liang, S., Zhang, L.: Object proposal by multi-branch hierarchical segmentation. CVPR pp. 3873–3881 (2015)
22. Xiao, Y., Lu, C., Tsougenis, E., Lu, Y., Tang, C.K.: Complexity-adaptive distance metric for object proposals generation. In: CVPR. pp. 778–786 (2015)
23. Xu, X., Ge, L., Ren, T., Wu, G.: Adaptive integration of depth and color for objectnes estimation. In: ICME. pp. 1–6 (2015)
24. Zhang, H., Zha, Z.J., Yang, Y., Yan, S., Gao, Y., Chua, T.S.: Attribute-augmented semantic hierarchy: Towards a unified framework for content-based image retrieval. TOMM 11(1s), 21 (2014)
25. Zhang, H., He, X., Porikli, F., Kneip, L.: Semantic context and depth-aware object proposal generation. In: ICIP (2016)
26. Zitnick, C.L., Dollár, P.: Edge boxes: Locating object proposals from edges. In: ECCV. pp. 391–405 (2014)