RGB-D Salient Object Detection via Feature Fusion and Multi-scale Enhancement

Peiliang Wu(✉), Liangliang Duan, and Lingfu Kong

School of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei Province, China
peiliangwu@ysu.edu.cn

Abstract. Salient object detection, especially for multi-object detection in complex scene, is a very challenging issue in computer vision. With the emergence and promotion of somatosensory sensors such as Kinect, RGB-D data jointing color and depth information can be obtained easily and inexpensively. This paper focuses on the RGB-D salient object detection. Firstly, the RGB image is converted into Lab color space and superpixels are segmented according to color and merged according to depth. Then, color contrast features and depth contrast features are calculated to construct an effective multi-feature fusion to generate saliency map. Finally, multi-scale enhancement is operated on the saliency map to further improve the detection precision. Experiments on the public data set NYU depth V2 show that the proposed method can effectively detect each salient object in multi-object scenes, and can also highlight the each object entirely.

Keywords: RGB-D · Superpixel segmentation · Salient object detection · Multi-feature fusion

1 Introduction

The human visual system has the capability to locate the most interest region in a cluttered visual scene, this selective attention mechanism allows us to effectively capture prey and escape predators, which is a very important survival skill for human. Due to its biological importance, many research efforts have been made to find the essence of attention mechanism. A variety of calculation models are proposed to simulate such biological mechanisms in the recent period of time. Visual saliency detection is a very important research in the field of visual attention mechanism. Currently, visual saliency detection is roughly divided into two aspects. One is using High-level visual prior knowledge to mimic human's top-down saliency computational model, whose basic idea is firstly to cluster image pixels into block feature, and through some prior knowledge to simulate the human eye's ability which can identify different objects. The other is the bottom-up detection model based on low-level visual features, whose basic idea is based on visual feature of gray, color or direction for forming the feature map of each feature dimension and merging into final saliency map. Computing visual saliency has very important applications such as image segmentation [1], image classification [2], object

H. Zha et al. (Eds.): CCCV 2015, Part II, CCIS 547, pp. 359–368, 2015.
DOI: 10.1007/978-3-662-48570-5_35

recognition [3] and other fields, and can optimize the allocation of various computing resources.

At present, many RGB-D sensors, such as Bumbbee camera, PMD camera, especially Kinect have been developed. As the perfect combination of visual camera and ultrasonic sensor, this kind of sensor can obtain scene and object RGB image and depth image simultaneously and is becoming a simple, cheap, convenient environment data acquisition equipment. Studies have been launched, but mainly focus on color RGB-D point cloud data registration [4], 3D reconstruction [5], etc.

Considering the color and depth information are important external data obtained by human vision, RGB-D data will be an important role in promoting research on human visual attention mechanism. In most recent years, salient object detection for RGB-D data gains much attention. Referring to the working mechanism of the human visual system, we propose a saliency calculation framework for RGB-D salient object detection. First of all, superpixels are segmented with Lab color and merged with depth. Then color contrast features and depth contrast features are calculated and fused based on background contrast. Finally, RGB-D image saliency calculation framework is proposed and improved with multi-scale saliency enhancement.

2 Related Work

Recently, RGB image saliency detection has been studied widely and deeply, in which low-level image contrast features play a very important role. The most influential model was proposed by Itti[6] in 1998, by combining low-level image feature such as color, edge and direction etc. and center-surround difference to calculate the image salient region. Harel et al. [7] developed Itti method to generate saliency map and perform the normalized operation based on graph method. Hou et al. [8] proposed a method based on the calculation of the residual spectrum, using the amplitude spectrum information generated by the Fourier transform of the image. Achanta[9] proposed a frequency-tuned approach, in which the distance between the image pixel and the average values of image are calculated as the pixel saliency. Cheng et al. [10] extended color histogram to 3D color space and proposed saliency analysis method based on the color region histogram. Perazzi et al. [11] combine color contrast and color distribution information for image saliency analysis. Margolin et al. [12] proposed a method that combined pattern and color into a model. The above methods can get good results when processing simple images. But when dealing with images containing complex background and several objects, the detection results are bad. Therefore, more saliency factors need to be integrated to solve this problem. RGB-D sensors collecting the color and depth information of the scenes at the same time, is expected to provide depth saliency factor in addition to color. But in terms of salient object detection based on RGB-D data, although several prior works[13-16] aim to explore the saliency analysis of RGB-D, they are still at the initial stage.

3 RGB-D Salient Object Detection

3.1 RGB-D Salient Object Detection Framework

The framework of our RGB-D salient object detection is shown in Fig. 1. For the input RGB-D image, firstly converte RGB into CIELab space and normalize depth into [0-255], secondly, segment superpixels according to Lab color. Thirdly, considering each superpixel as a processing unit, calculate the average depth of each processing unit and merge Lab-based superpixels according to their difference value of average depth(In this paper two superpixel will be merged when their difference value of average depth<10). Then, fuse Lab contrast features and depth contrast features of each merged superpixel to get the global saliency map, and finally, the multi-scale enhancement is designed to improve the detection precision.

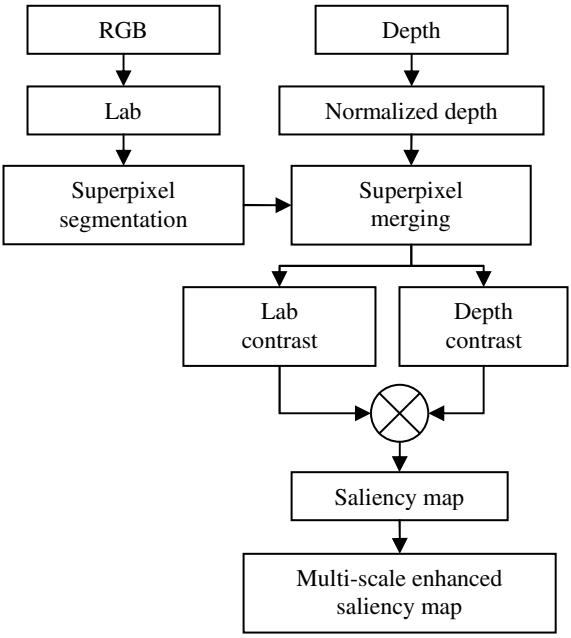

Fig. 1. The framework of RGB-D salient object detection

3.2 RGB-D Superpixel Segmentation

Early salient object detection methods are mainly based on pixel or regularized image unit, and the detection results is unsatisfied. Current methods based on irregular image unit (superpixel) become very popular, including graph cutting [17], Mean shift [18] and SLIC [19], these methods can significantly improve the saliency detection results and generate saliency map with high quality. This paper develops graph cutting segmentation [17] to handle RGB-D images, The details are as follows.

(1) Convert RGB to Lab color, segment the Lab into several disjoint regions $\{O_i\}_{i=1,2,...,N}$ according to [17], Where N is the number of segmented region. $f_c^{(i)}$ denotes the color feature of i^{th} region, where $f_c^{(i)} = \sum_{p \in O_i} V_p / U_i$ denotes the average Lab color of all the pixels in this region, p is the pixel within the region of O_i, V_p denotes Lab feature vector of pixel p, U_i is the number of pixel within region O_i.

(2) Normalize raw depth data. and mapping the normalization results to the range of 0~255. In the segmented regions obtained from (1), calculate the average depth of each region $f_d^{(i)} = \sum_{p \in O_i} D_p / U_i$, where D_p denotes normalized depth of p with value in [0-255].

(3) Merge adjacent segmentation regions when their depth difference<10. And then, recalculate the number of regions N, as well as color feature $f_c^{(i)}$ and depth feature $f_d^{(i)}$.

3.3 RGB-D Contrast Features

Contrast is the most important factor in the low visual saliency calculation. Because the size of each superpixels segmented above is obvious different, we need to consider the size factor to calculate RGB-D contrast of each segmented region $\{O_i\}_{i=1,2,...,N}$.

Color Contrast

Considering the color difference between the segmented image regions(In the Lab color space), the distance of between two regions (in depth channel) and the size of segmented region, the global color contrast feature of a segmented region is defined as follows.

$$S_c^{(i)} = \sum_{k=1,k\neq i}^{N} \varphi(O_i, O_k) \cdot \tau(O_i, O_k) \cdot U_k \tag{1}$$

where $\varphi(O_i, O_k)$ is the color difference between segmented region O_i and O_k (measured in the CIELab color space), the definition as shown in (2).

$$\varphi(O_i, O_k) = \left\| f_c^{(i)} - f_c^{(k)} \right\| \tag{2}$$

$\tau(O_i, O_k)$ denotes smooth term measuring the distance between the different segmented regions of image, which is used to balance the impact of saliency between different positions within the image space.

$$\tau(O_i, O_k) = 1 - D(O_i, O_k) \tag{3}$$

where $D(O_i, O_k) = \left\| c^{(i)} - c^{(k)} \right\|$ denotes spatial distance between region O_i and O_k. When calculating color contrast of region, it has a great impact on adjacent neighbor regions, on the contrary, it has a little impact on long distance regions.

Depth Contrast

Considering the depth difference between two segmented regions (in the depth channel) and the size of the segmented regions, the depth contrast feature of the segmented region O_i is defined as follows.

$$S_d{}^{(i)} = \sum_{k=1,k\neq i}^{N} \phi(O_i, O_k) \cdot \tau(O_i, O_k) \cdot U_k \tag{4}$$

where $\phi(O_i, O_k)$ is the depth difference between two segmented region O_i and O_k of image.

$$\phi(O_i, O_k) = \left\| f_d^{(i)} - f_d^{(k)} \right\| \tag{5}$$

$\tau(O_i, O_k)$ denotes smooth term measuring the distance between the different segmented regions of image which is defined as equation (3).

3.4 Saliency Features Fusion

When the scene image contains complex background and a variety of objects, it is difficult to detect salient objects accurately only use one single cue. Saliency cues of both color contrast and depth contrast reflect image saliency from different perspectives. Simple linear fusion may make saliency detection bad[20], so it is necessary to design an effective strategy to integrate these saliency cues. In order to highlight each salient object uniformly in a complex and multi-object scene, we use the following feature fusion approach.

$$S_{original}{}^{(i)} = \exp(S_c{}^{(i)}) \times S_d{}^{(i)} \tag{6}$$

So far, the saliency map is obtained by multi-feature fusion with both color and depth channels.

3.5 Multi-scale Saliency Enhancement

Under a single scale, saliency image analysis are often not comprehensive [6,21]. When changing the resolution of the image, the image structure will show different features, so it is very necessary for saliency analysis under multiple scales.

In this paper we use the multi-scale representation of the image to further enhance the saliency detection results, and achieve the goal that highlight each salient object uniformly. In this paper, we down sample RGB-D images into four different scales. Finally, the definition of fusion type of saliency image at multi-scales is defined in (7).

$$S_{final}(I) = \bigoplus_{h=1}^{H} S_{original}(I^h) \tag{7}$$

where I^h are images at different scales, the image of the original scale is I whose h=1. $S_{original}(I^h)$ is the saliency detection result in the single h-scale based on above section. We normalize $S_{final}(I)$, and get the final multi-scale salient object detection results.

4 Experiments and Analysis

We chose NYU Depth V2 as data set, which is comprised of video sequences from a variety of indoor scenes as recorded by both the RGB and Depth cameras from the Microsoft Kinect.

4.1 Comparison of Superpixel Segmentation Results

In order to evaluate the advantages of superpixel segmentation jointing color and depth, we compared graph-cutting based superpixel segmentation on (1) depth, (2) Lab color, and (3) depth + Lab. The three segmentation results are shown in Fig. 2.

It can be seen from the figure that for superpixel segmentation only based on depth, the segmented regions are too large, and adhesion appears easily at the bottom of objects placed on table. For superpixel segmentation only based on color, the segmented regions almost too small, and the whole object is segmented into many small parts, which is called as over-segmentation. Over-segmentation always happens on the texture objects such as boxes and right-side cap in Fig. 2. For our RGB-D superpixel segmentation, because the combination of both color cue and depth cue, superpixels are segmented neither too large nor too small, and the boundary of each object is distinct.

4.2 Salient Object Detection Results

In order to evaluate the advantages of jointing depth and color data to detect salient object, we compared salient object detection at a single scale and at multi-scales respectively based on (1) depth, (2) Lab image, and (3) depth + Lab. The obtained saliency maps are shown in Fig. 3.

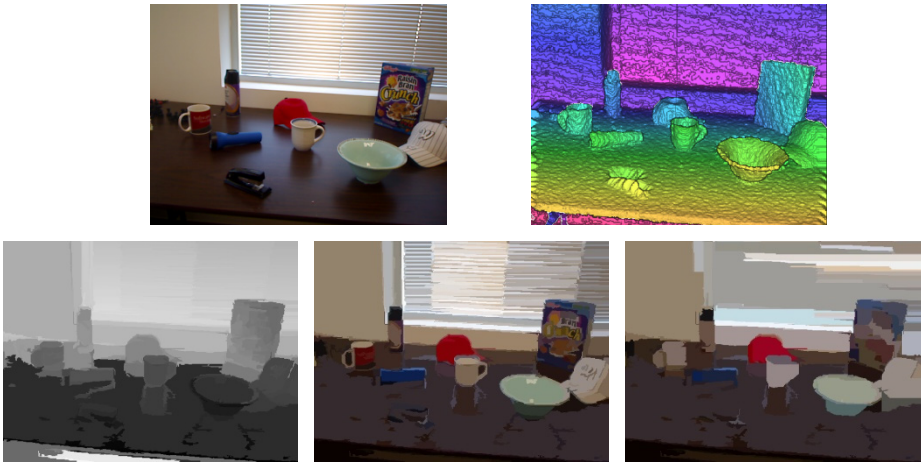

Fig. 2. Graph-cut based Superpixel segmentation on three type of data (top-left: raw RGB, top-right: raw depth; bottom-left: segmentation on depth, bottom-middle: segmentation on Lab, bottom-right: segmentation on Depth + Lab)

As can be seen from Fig. 3, (1) the result of salient object detection is unsatisfied under a single scale, while after multi-scale enhancing, the object regions are highlighted. (2) Over-segmentation in saliency map is distinct when only using color cue, suppression of background is not good when only using depth cue. While the salient object detection jointing color and depth can detect almost every salient object and meanwhile highlight the outline of each object.

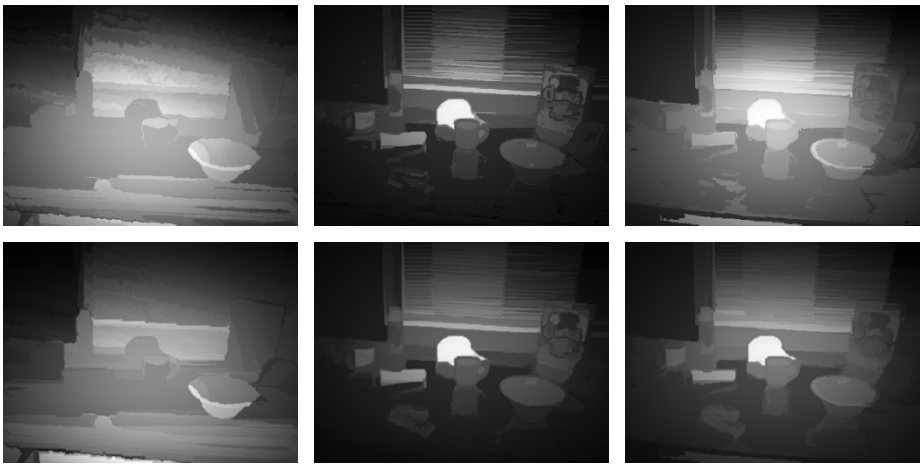

Fig. 3. Salient object detection on three type of data (top: salient object map under single scale, top-left: only depth cue, top-middle: only Lab color cue, top-right: Lab+depth cues. bottom: salient object map with multi-scales enhancement, bottom-left: only depth cue, bottom -middle: only Lab color cue, bottom-right: Lab+Depth cues)

4.3 Contrasts of Salient Object Detection Results

In order to evaluate the proposed salient object detection method jointing depth and color cues, we compared our salient object detection results with early work [16]. The experimental RGB-D images are chosen form four scenes (including desk, kitchen_small and meeting_small, table) from NYU Depth V2. The obtained saliency maps are shown in Fig. 4.

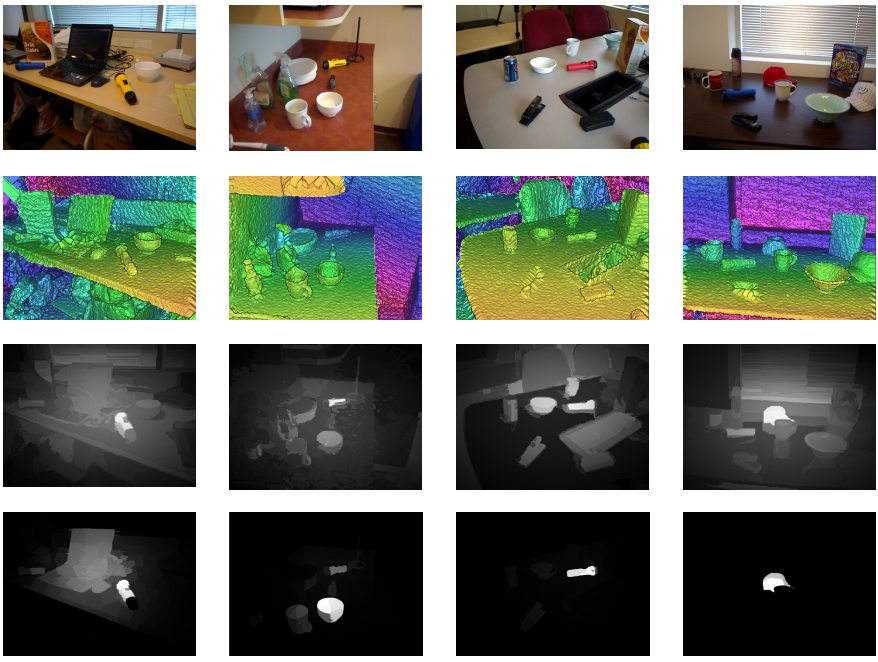

Fig. 4. Contrasts of RGB-D salient object detection on four type of scenes (top: original RGB image; second row: the corresponding depth image; Third row: salient map of our approach; bottom: salient map of [16].)

As we can see from Fig. 4, on contrary, our approach perform better on multi-object scenes, and detect almost each of the salient objects. While [16] only detects one or two objects in the middle part of image.

5 Conclusion

In this paper, we propose a multi-feature fusion framework for RGB-D salient object detection. As a preprocessing stage, RGB-D superpixels are segmented based on graph-cut algorithm. Then color contrast feature and depth contrast feature are extracted and integrated from each superpixel. Under different scales, multi-scale enhancement is designed. The proposed method can produce high quality salient object

map which not only highlights each salient object in the multi-object scene, but also can effectively alleviate the over-segmentation.

Because of the complex of the scenes, although our approach can detect almost each of the salient objects, but some background is also highlighted. The next step of our work is to reduce the impact of complex background and improve detection accuracy.

Acknowledgments. This work is supported by National Natural Science Foundation of China No. 61305113.

References

1. Chang, K.Y., Liu, T.L., Lai, S.H.: From co-saliency to co-segmentation: an efficient and fully unsupervised energy minimization model. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2129–2136 (2011)
2. Sharma, G., Jurie, F., Schmid, C.: Discriminative spatial saliency for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3506–3513 (2012)
3. Ren, Z., Gao, S., Chia, L.T., et al.: Region-based Saliency Detection and Its Application in Object Recognition. IEEE Transactions on Circuits and Systems for Video Technology **24**(5), 769–779 (2014)
4. Hao, M., Biruk, G., Kishore, P.: Color point cloud registration with 4D ICP algorithm. In: IEEE International Conference on Robotics and Automation. Shanghai, China, pp. 1511–1516 (2011)
5. Zollhofer, M., Niebner, M., et al.: Real-time Non-rigid Reconstruction using an RGB-D Camera. ACM Transactions on Graphics **33**(4), 1–12 (2014)
6. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(11), 1254–1259 (1998)
7. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Annual Conference on Neural Information Processing Systems, pp. 545–552 (2006)
8. Hou, X.D., Zhang, L.Q.: Saliency detection: a spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE, Minneapolis (2007)
9. Achanta, R., Hemami, S., Estrada, F., et al.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1597–1604. IEEE, Miami Beach (2009)
10. Cheng, M.M., Zhang, G.X., Mitra, N.J., et al.: Global contrast based salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 409–416. IEEE, Colorado Springs (2011)
11. Perazzi, F., Krahenbuhl, P., Pritch, Y., et al.: Saliency filters: contrast based filtering for salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 733–740 (2012)
12. Margolin, R., Tal, A., Zelnik-Manor, L.: What makes a patch distinct? In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1139–1146 (2013)
13. Ciptadi, A., Hermans, T., Rehg, J.M.: An in Depth View of Saliency. In: BMVC, pp. 1–11 (2013)
14. Desingh, K., Krishna, K.M., Jawahar, C.V., Rajan, D.: Depth really matters: improving visual salient region detection with depth. In: BMVC, pp. 1–11 (2013)

15. Lang, C., Nguyen, T.V., Katti, H., Yadati, K., Kankanhalli, M., Yan, S.: Depth Matters: Influence of Depth Cues on Visual Saliency. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 101–115. Springer, Heidelberg (2012)
16. Peng, H., Li, B., Xiong, W., Hu, W., Ji, R.: RGBD Salient Object Detection: A Benchmark and Algorithms. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part III. LNCS, vol. 8691, pp. 92–109. Springer, Heidelberg (2014)
17. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International Journal of Computer Vision **59**, 167–181 (2004)
18. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**, 603–619 (2002)
19. Achanta, R., Shaji, A., Smith, K., et al.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**, 2274–2282 (2012)
20. Gopalakrishnan, V., Hu, Y., Rajan, D.: Salient region detection by modeling distributions of color and orientation. IEEE Transactions on Multimedia. **11**(5), 892–905 (2009)
21. Goferman, S., Zelnik-Manor, L., Tal, A.: Context- aware saliency detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(10), 1915–1926 (2012)