Multimed Tools Appl
https://doi.org/10.1007/s11042-018-6319-4

RGB-D image saliency detection from 3D perspective

Zhengyi Liu [1,2] · **Tengfei Song** [1,2] · **Feng Xie** [1,2]

Received: 11 April 2017 / Revised: 15 June 2018 / Accepted: 25 June 2018

Abstract With the advent of stereo camera saliency object detection for RGB-D image is attracting more and more interest. Most existing algorithms treat RGB-D image as one RGB image and one depth map, then measure saliency map independently, and last fuse them. They disregard the fact that human visual system operates in real 3D environments. The paper proposed saliency object detection for RGB-D image from 3D perspective. It regards object as three dimensional structures, and redefines boundary conception in RGB-D image, and regards space boundary including top, down, left, right, front, back plane in real 3D environment as background. It incorporates 3D compactness feature, in which salient objects typically have 3D compact spatial distributions, into color and depth feature to express similarity among supervoxels and applies manifold ranking by six boundary planes to generate six saliency maps, and then integrates them to get the RGB-D saliency map from background view. In the end it refines saliency map by high confident salient seeds from foreground view. Experiment results show that six planes of RGB-D image are superior to four sides of RGB image as background seeds and 3D compactness plays an important role in saliency measurement. Our approach outperforms other state-of-the-art algorithms on NLPR RGBD 1000 benchmark.

Keywords RGB-D image saliency · 3D boundary prior · 3D compactness · Manifold ranking

✉ Zhengyi Liu
22927463@qq.com

Tengfei Song
sdtastf@163.com

Feng Xie
811173098@qq.com

[1] Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education, Anhui University, Hefei, China

[2] Co-Innovation Center for Information Supply & Assurance Technology, Anhui University, Hefei, China

Published online: 31 July 2018

1 Introduction

Saliency detection aims to detect the most attractive objects to human viewers in an image, which has been widely used as an important fundamental of various visual applications, including object recognition [15], image segmentation [17, 20], image compression [14, 28, 37], image retrieval [2, 9, 10, 18, 22, 29, 35], visual tracking [8, 40, 41], scene classification [3, 24, 36], human-robot interaction [30] etc. It can significantly reduce the complexity of high-level vision tasks which only need to be concentrated on a few regions of interests instead of entire images.

With the advent of stereo camera such as Microsoft Kinect, saliency object detection for RGB-D image is attracting more and more interest. Each RGB-D image can be decomposed into a color image with R, G and B color channels and an aligned gray-level depth map. Many researches focus on how to use depth information for saliency detection. One assumption is that the nearer objects attract more human attention [16]. But saliency detection result based on such assumption from depth map is relatively coarse [4, 6, 7, 12, 13, 23, 26, 31] and fusion with color feature and other prior is imperative. One research line [4, 5, 7, 23, 25, 31–33] treats the depth map independently to generate a depth-induced saliency map and then combine the depth-induced saliency map with a color-produced saliency map via some heuristic or machine learning approaches. The other line [11, 23, 25, 26] combines depth feature with color feature to guide saliency propagation and iteration. Most existing algorithms disregard the fact that human visual system operates in real 3D environments when utilizing depth cues.

In fact RGB-D image composed of RGB image and depth image demonstrates 3D world as shown in Fig. 1. Each superpixel not only has x and y coordinates in RGB image, but also has z coordinate expressed by depth value. Different scenes have different depth-of-fields (DOFs) [16] and absolute depth values from depth image are normalized to the range [0, 1] in order to get rid of the influence of DOFs.

From Fig. 1 we can see that depth values of salient bottle are about 0.3, while depth values of surrounding are about 0.8, and depth values of image's bottom side are about 0.06. Through observation we can find that 3D world demonstrated by RGB-D image has the following characteristics:

(1) Saliency objects are much less connected to 3D scene boundaries including six planes such as top, down, left, right, front and back plane. From Fig. 2 we can see that each superpixel is no longer superpixel in the plane but supervoxel in three dimensional spaces for it has the depth value in RGB-D image. Salient bottle is obviously near the center of 3D space and apart from top, down, left, right, front and back plane. We call it 3D Boundary Prior.

(2) Salient objects typically have 3D compact spatial distributions, whereas background regions have a wider distribution over 3D world. From Fig. 2 we can see that the depth value of supervoxel on the bottle is fairly close and different from its surroundings. That is to say, salient objects not only have compact spatial distributions in RGB image illustrated in [39], but also group together in depth direction. We call it 3D Compactness.

Based on above characteristics, the paper proposes saliency object detection for RGB-D image from 3D perspective. It regards object as three dimensional structures, and has 3D

Fig. 1 Superpixels with normalized depth value built from RGB-D image

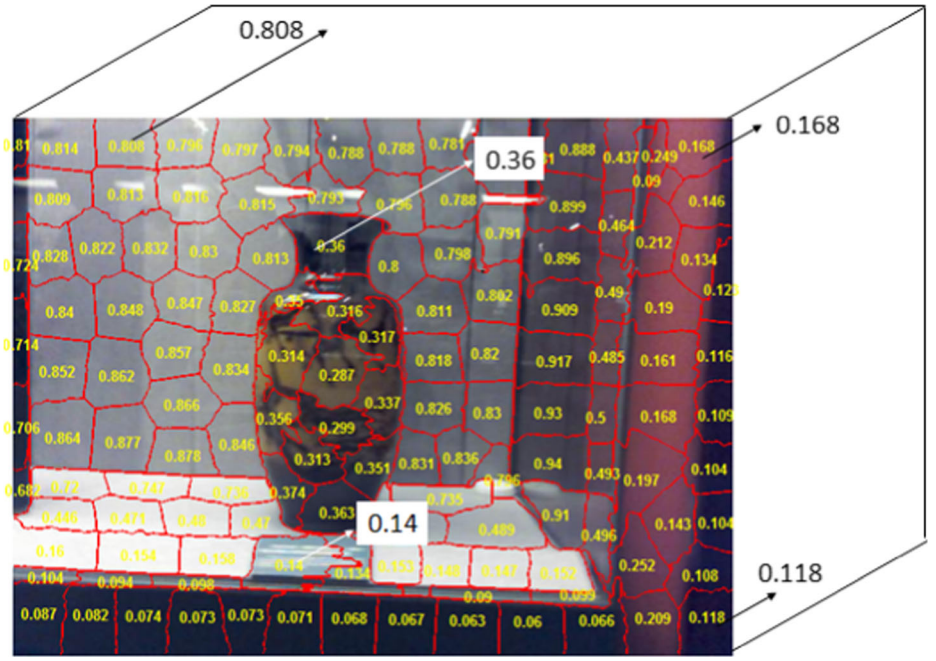


Fig. 2 Supervoxels deviated from RGB plane by depth value in 3D scene

compactness feature, and combines color and depth feature together to express similarity among supervoxels. It redefines boundary conception in RGB-D image and regards top, down, left, right, front, back plane in real 3D environment as space boundary. It constructs six saliency maps using supervoxels in six planes as background seeds [34], i.e. labeled data and applies manifold ranking [38] to rank the relevance of all the supervoxels and then integrates them to get the RGB-D saliency map by background prior, and gets high confident salient seeds. In the end, it refines saliency map by salient seeds base on foreground prior. The framework of our RGB-D salient object detection is shown in Fig. 3.

The main contributions of this work include:

(1) It proposes boundary prior in 3D perception, which saliency objects are much less connected to 3D scene boundaries such as top, down, left, right, front and back plane.
(2) It proposes 3D compactness feature, which saliency objects have compact spatial distributions in 3D scene and group together not only in RGB image plane but also in depth direction.
(3) It proposes RGB-D image saliency detection method from 3D perception, which uses color, depth and 3D compactness feature, and applies 3D boundaries as background seeds, and then ranks the saliency from background perspective, and last refines saliency map from foreground perspective.

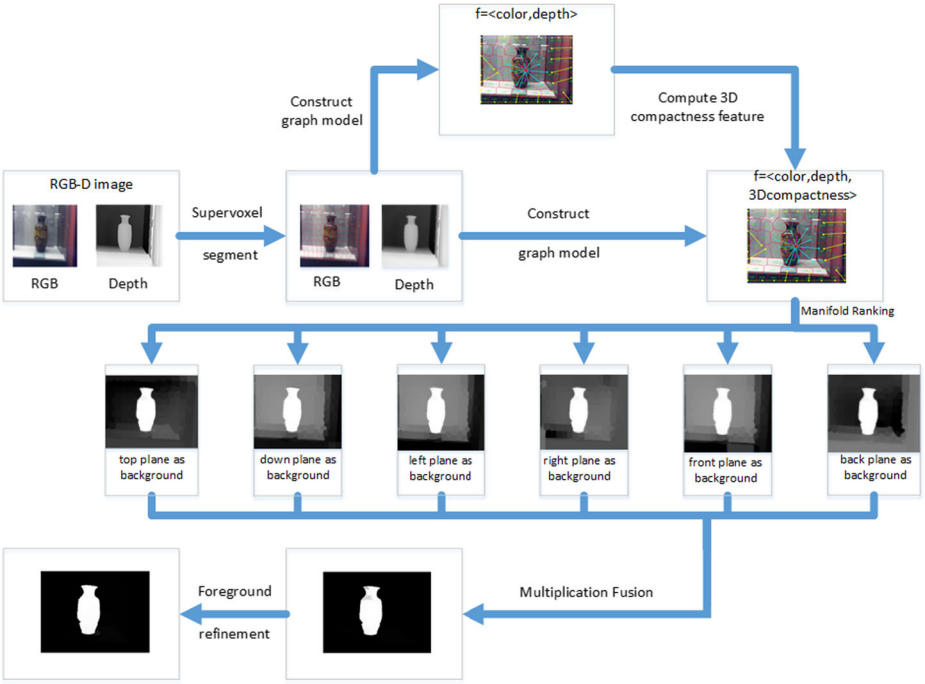


Fig. 3 Framework of our method. The RGB-D image is segmented into supervoxels by color and depth cue, and then constructs graph model by feature = <color,depth > to compute 3D compactness feature, and then applies manifold ranking in graph model with feature = <color,depth,3D compactness> by 3D boundary prior that supervoxels in top, down, left, right, front and back plane are background seeds, and then fuses six coarse saliency maps by multiplication operation and last refines the saliency result from foreground perspective

2 Related works

Saliency object detection for RGB-D image is attracting more and more interest recently. Most researches focus on getting better saliency map from depth image. Cheng et al. [4], Ren et al. [26], Guo et al. [7], Tang et al. [31] measure depth contrast using absolute depth difference between regions. Peng et al. [23] propose a multi-contextual contrast model including local contrast, global contrast and background contrast to detect salient object from depth map. Ju et al. [12, 13] define the depth saliency of a point as how much it outstands from surroundings which are measured using an anisotropic center-surround operator. Fent et al. [6] propose Local Background Enclosure feature which employs an angular density component and an angular gap component to measure the proportion of the object boundary in front of the background. But saliency detection result from depth map is relatively coarse and fusion with color feature and other prior is imperative.

Cheng et al. [4] measure each region's salient value using color contrast, depth contrast and spatial bias. Each region is represented by its center in a 6-dimensional vector: RGB color (3D), 3D spatial information (image coordinate and depth). Contrast cue is based on feature distance and spatial distance weighting. Depth contrast is represented by absolute value between regions.

Xue et al. [33] propose a saliency object detection model integrating RGB and depth cues via mutual manifold ranking of RGB image and depth map, and take color and depth feature as the measurement of similarity, and define four corners as background queries. It ranks the saliency of the RGB image with composition feature including color and depth cues, and it then makes use of the RGB saliency as the prior to rank the saliency of the depth map, and last fuses them.

Ren et al. [26] propose fusion method of region contrast and global priors such as background, depth and surface orientation priors for RGB-D saliency detection. In region contrast measurement color vector is augmented with depth to indicate its color and depth contrast to all other superpixels.

Tang et al. [31] present a two-stage framework for detecting salient objects in challenging images and each stage combines color and depth features. In the object location stage, region contrast and depth prior measurement produces a noise-filtered salient patch, which indicates the location of the object. In the object boundary inference stage, boundary information is encoded in a graph using both depth and color information, and then heat diffusion is employed to infer more reliable boundaries and obtain the final saliency map.

Guo et al. [7] propose a saliency fusion and propagation strategy based salient object detection method for RGB-D images, in which color cue, location cue and depth cue are fused to provide high precision detection result and saliency propagation is utilized to improve the completeness of salient objects.

Guo et al. [25] propose a novel salient object detection method for RGB-D images using saliency evolution strategy. It firstly over-segments a RGB-D image into super-pixels with the extended SLIC algorithm based on both color cue and depth cue. Then it estimates two saliency maps based on color cue and depth cue independently, and fuse them with refinement to obtain the initial saliency map with high precision. Finally, it iteratively propagates saliency over the whole image on a graph-based model and generates the final saliency map.

Ju et al. [12, 13] propose a saliency method that works on depth images based on anisotropic center-surround difference. Instead of depending on absolute depth, it measures the saliency of a point by how much it outstands from surroundings, which takes the global depth structure into consideration. Depth prior and location prior are also used for refinement.

Peng et al. [23] separate the input RGB-D image into two independent components: a RGB image and a depth map, and then calculate their own saliency maps and last fuse them.

Song el al. [27] propose the depth-based object probability to enhance color contrast and depth contrast, and effectively utilize the region merging method for saliency refinement.

Most existing algorithms disregard the fact that human visual system operates in real 3D environments. In order to simulate human visual system more really, RGB-D image composed of RGB image and depth image should not be treat as two plane images, and it can exhibit 3D characteristics and be employed in saliency detection.

3 3D boundary prior

In [23, 33, 34, 42], four image boundaries or four corners of the image are taken as the boundary, whereas this assumption is unsuitable for RGB-D image saliency detection. Salient objects correspond to real objects and generally be more likely to be placed near the center of 3D scene and apart from 3D scene boundaries composed of top, down, left, right, front, back plane. Thus four boundary sides or four corners are no longer the best selection of boundary of RGB-D image. We establish coordinate system illustrated in Fig. 4, and define 3D boundary of RGB-D image as follows:

$$P_{top} = \{x, y, z | 0 < x < width, y = 0, 0 < z < 1\} \tag{1}$$

$$P_{down} = \{x, y, z | 0 < x < width, y = height, 0 < z < 1\} \tag{2}$$

$$P_{left} = \{x, y, z | x = 0, 0 < y < height, 0 < z < 1\} \tag{3}$$

$$P_{right} = \{x, y, z | x = width, 0 < y < height, 0 < z < 1\} \tag{4}$$

$$P_{front} = \{x, y, z | 0 < x < width, 0 < y < height, z = 0\} \tag{5}$$

$$P_{back} = \{x, y, z | 0 < x < width, 0 < y < height, z = 1\} \tag{6}$$

The definition of 3D boundary is intuitive but difficult to compute the proportion between plane coordinate and depth value. Fortunately we get geometry theory of 3D boundary and propose a “soft” approach to define supervoxels in 3D boundary.

Each superpixel in RGB-D image not only has x and y coordinates, but has z coordinate, and is called supervoxel. Take supervoxels in the left plane as an example, each supervoxel deviates from front plane by depth value along Z direction, thus supervoxels in the left plane are just superpixels in the left side of RGB image. And supervoxels in right/top/down plane in 3D boundary are just superpixels in the right/top/down side in RGB image similarly. Thus how to express supervoxels in the front plane and back plane is rather different.

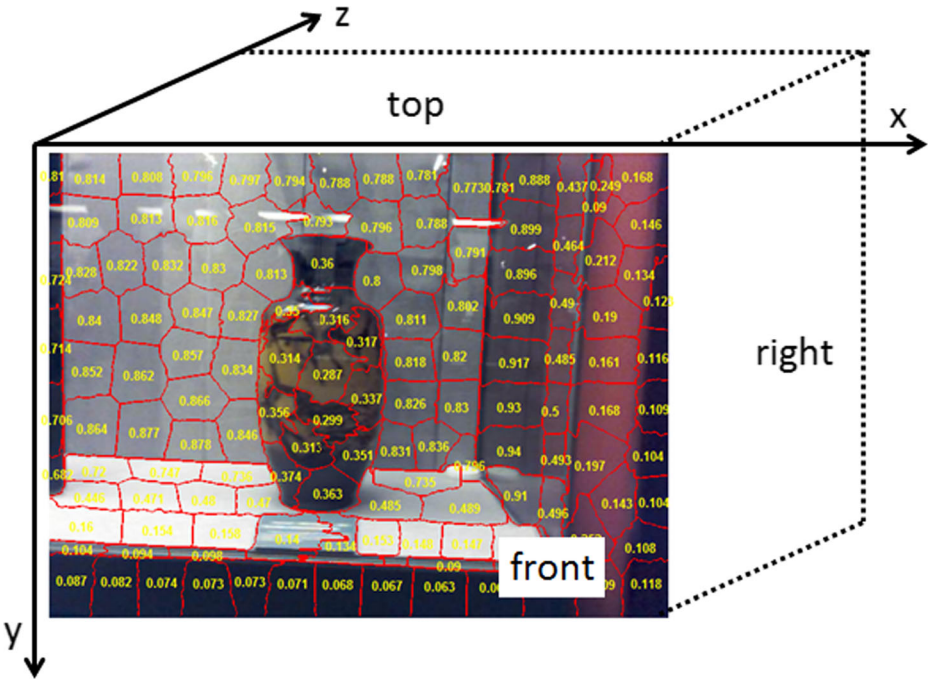


Fig. 4 Coordinate system of 3D scene and definition of top, front, right plane and opposite planes including down, back, left plane

To get the supervoxels in the front and back plane, we sort all the supervoxels by normalized depth value, and choose supervoxels with the highest depth value as the supervoxels in the back plane and choose supervoxels with the lowest depth value as the supervoxels in the front plane. Such definition is proposed considering that salient object are more likely away from back plane and closer to front plane, but they are not supervoxels in the front plane. Therefore we can get supervoxels in six planes as boundary seeds and apply manifold ranking to calculate RGB-D image saliency.

4 3D compactness feature

Salient objects generally correspond to real objects in 3D scene. They are grouped together into connected region in 3D scene and surrounded by background regions. Thus in the three dimensional spatial domain, salient objects typically have 3D compact spatial distributions, whereas background regions have a wider distribution over the 3D world.

We first define the feature vector as $f=[c,d]^T$, where c_i and d_i are normalized average CIELab color value and depth value. And then we define the similarity a_{ij} between a pair of supervoxels v_i and v_j using:

$$a_{ij} = \exp\left[-\left(\frac{\|c_i - c_j\|}{\sigma_c^2} + \frac{\|d_i - d_j\|}{\sigma_d^2}\right)\right] \tag{7}$$

where σ_c and σ_d are constants that control the strength of the weights respectively.

We construct extended graph model $G_1 = <V, E>$ enlightened by [34, 39], where V is a set of nodes, i.e., supervoxels generated through SLIC algorithm [1] and E is a set of undirected edges which is weighted by an affinity matrix $P = [p_{ij}]_{n \times n}$:

$$p_{ij} = \begin{cases} a_{ij} & if \quad j \in N_i \\ 0 & otherwise \end{cases} \tag{8}$$

where n denotes the number of supervoxels, and N_i denotes the set of neighbors of v_i. Note that each node is not only connected to those nodes neighboring it, but also connected to the nodes sharing common boundaries with its neighboring node, and all nodes around 3D image borders including supervoxels in four image sides and those supervoxels with the highest and lowest depth values are considered neighbors of each other as illustrated in Fig. 5.

And then we propagate the similarity using the manifold ranking through the constructed graph. That is,

$$H^T = (D - \alpha P)^{-1} P \tag{9}$$

where $P = [p_{ij}]_{n \times n}$ is the affinity matrix, $D = diag\{d_{11}, \cdots, d_{nn}\}$ is degree matrix where $d_{ii} = \sum_j p_{ij}$, $H = [h_{ij}]_{n \times n}$ is the similarity matrix after the diffusion process.

We define the spatial variance of supervoxel v_i as:

$$sv(i) = \frac{\sum_{j=1}^{n} h_{ij} \cdot n_j \cdot \left\| b_j - \mu_i \right\|}{\sum_{j=1}^{n} h_{ij} \cdot n_j} \tag{10}$$

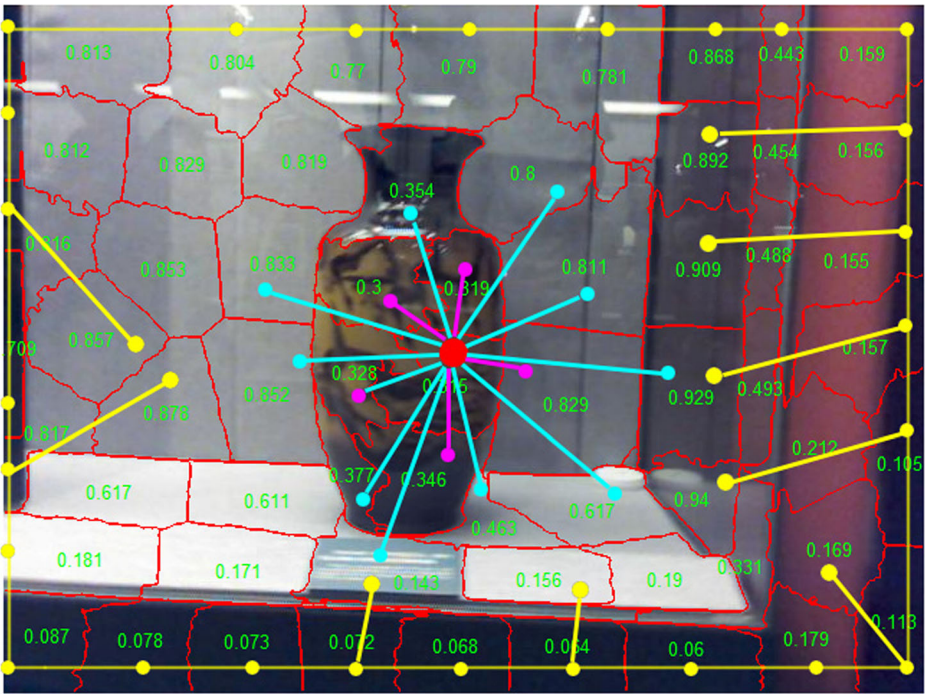


Fig. 5 Graph model. The yellow line connects supervoxels in four sides and supervoxels with the highest and lowest depth value, and indicates that all the 3D boundary nodes are connected with each other

where n_j is the number of pixels that belong to supervoxel v_j, $b_j = \left[b_j^x, b_j^y, b_j^z\right]$ is the 3D center of the supervoxels v_j, and the spatial mean $\mu_i = \left[\mu_i^x, \mu_i^y, \mu_i^z\right]$ is defined as:

$$\mu_i^x = \frac{\sum_{j=1}^{n} h_{ij} \cdot n_j \cdot b_j^x}{\sum_{j=1}^{n} h_{ij} \cdot n_j} \tag{11}$$

$$\mu_i^y = \frac{\sum_{j=1}^{n} h_{ij} \cdot n_j \cdot b_j^y}{\sum_{j=1}^{n} h_{ij} \cdot n_j} \tag{12}$$

$$\mu_i^z = \frac{\sum_{j=1}^{n} h_{ij} \cdot n_j \cdot b_j^z}{\sum_{j=1}^{n} h_{ij} \cdot n_j} \tag{13}$$

where b_j^x and b_j^y are expressed by absolute x and y value relative to width and height of RGB image, while b_j^z is expressed by normalized depth value from depth image. The proportion between them is difficult to compute to express the real 3D objects. But we find that 3D compactness characteristic still exists even if the proportion is not exactly the right proportion. From Fig. 2 we can see if the depth value is larger than plane value x and y, salient bottle still has 3D compactness characteristic but salient bottle is stretched along depth direction, and if the depth value is smaller than plane value x and y, 3D compactness characteristic of salient bottle is more distinct because salient bottle is compressed along depth direction. Therefore there is no need to find the exact proportion between depth and plane direction value. The method we use is to get an appropriate proportion by experiment. Of course if we can get the exact proportion, then the result is more excellent.

Salient objects are more likely to be placed near the center of 3D images and the background generally spreads over the 3D scene. Salient objects locate in the relatively closer depth levels with respect to the camera than background regions, but they are not the nearest ones and in fact there are usually some background regions closer to the camera than object regions.

Motivated by this, we calculate the spatial distance of supervoxels from the center of RGB-D images using:

$$sd(i) = \frac{\sum_{j=1}^{n} h_{ij} \cdot n_j \cdot \left\| b_j - p \right\|}{\sum_{j=1}^{n} h_{ij} \cdot n_j} \tag{14}$$

where $p = [p_x, p_y, p_z]$ is the spatial coordinate of RGB-D image center. p_x, p_y are the spatial coordinate of RGB image center, p_z is set to the least 2% of depth value, and the proportion between depth and plane direction value is the same as discussed before.

Supervoxels that exhibit large spatial variance across the RGB-D image are less likely to be salient. Supervoxels that exhibit large spatial distance from the center of RGB-D image are less likely to be salient. 3D Compactness feature of supervoxel v_i is consequently defined as:

$$s(i) = 1 - Norm(sv(i) + sd(i)) \tag{15}$$

where $Norm(x)$ is a function that normalized x to the range [0,1].

5 Saliency detection in RGB-D image

In RGB-D image salient objects typically have 3D compactness and 3D boundary prior characteristics. Therefore we construct graph model based on color, depth and 3D compactness feature from 3D boundary prior perspective to guide the saliency ranking.

5.1 Graph construction

Incorporating the depth cues and 3D compactness feature, we define the feature vector as $f=[c,d,s]^T$, where c, d, s are the normalized average CIELab color value, depth value and 3D compactness value in each supervoxel. We construct the undirected graph model $G_2=(V,E)$ as illustrated in Fig. 5, in which affinity matrix $W=[w_{ij}]_{n*n}$ is defined as:

$$w_{ij}=\begin{cases}\exp\left[-\left(\frac{\|c_i-c_j\|}{\sigma_c^2}+\frac{\|d_i-d_j\|}{\sigma_d^2}+\frac{\|s_i-s_j\|}{\sigma_s^2}\right)\right] & if \quad j\in N_i\\ 0 & otherwise\end{cases} \tag{16}$$

where σ_c, σ_d, σ_s are constants that control the strength of the weights respectively.

5.2 3D boundary definition

We choose supervoxels in six planes as background seeds, and as analyzed before they are supervoxels in four sides of RGB image, supervoxels of the highest depth value as supervoxels in back plane and supervoxels of the lowest depth value as supervoxels in front plane. The highest proportion and lowest proportion are set as δ_h, δ_l respectively.

5.3 Ranking with 3D boundary queries

We use the supervoxels on 3D Boundary as background seeds, i.e., the labeled data to rank the relevance of all the other regions. Note that we construct six saliency maps using six planes as 3D boundary prior and then integrate them for the final map.

For each 3D image boundary, we rank the relevance of all the supervoxels to this 3D boundary using:

$$\underset{p\in\{\text{top,down,left,right,front,back}\}}{f_p^{\,*}}=(\mathrm{D}-\alpha\mathrm{W})^{-1}y_{3D} \tag{17}$$

where indicator vector y_{3D} indicates the relevance of a superpixel to 3D background seed, $W=[w_{ij}]_{n\times n}$ is the affinity matrix, $D=diag\{d_{11},\cdots,d_{nn}\}$ is degree matrix where $d_{ii}=\sum_j w_{ij}$.

We normalize f^* to range [0, 1], and the saliency map S_p using boundary prior can be written as:

$$\underset{p\in\{\text{top,down,left,right,front,back}\}}{S_p}=1-\overline{f}^{\,*} \tag{18}$$

where $\overline{f}^{\,*}$ denotes the normalized vector.

Saliency maps are integrated by the following process:

$$S_{bg} = \prod_{p\in\{\text{top,down,left,right,front,back}\}} S_p \tag{19}$$

By multiplication operation and OTSU threshold method [21] we get high confident supervoxel seeds in saliency measurement.

5.4 Refinement

We sample the high confident supervoxels as seeds, while leaving the others as unknown. The saliency of unknown supervoxels is generated by the optimization of manifold ranking using:

$$f^* = (\mathrm{D}-\alpha\mathrm{W})^{-1}y \tag{20}$$

The indicator vector y indicates the relevance of a supervoxel to high confident salient seeds. Final saliency map is obtained using:

$$S = \overline{f}^* \tag{21}$$

6 Experiments and comparison

The performance of our proposed method is evaluated on NLPR RGBD1000 benchmark [23]. It contains 1000 RGB-D images captured by Kinect, along with manually labeled ground truths for salient objects.

Evaluation metrics We adopt the precision-recall curve and average precision, recall and F-measure score proposed in [19] to evaluate algorithm performance as follows:

$$precsioin = \frac{TP}{TP + FP} \tag{22}$$

$$recal = \frac{TP}{TP + FN} \tag{23}$$

$$F_\beta = \frac{\left(1+\beta^2\right) \times precision \times recall}{\beta^2 \times precision + recall} \tag{24}$$

where TP, FP and FN are the weighted true-positive, false-positive and false-negative, F_β denotes the weighted F-measure and $\beta^2 = 0.3$ to emphasize the precision.

Experiment setup Throughout all the experiments, the parameters involved in our algorithm are empirically assigned and proved the best by experiments as follows:

The number of the supervoxel nodes is set as $n = 200$. Constants that control the strength of the weights in graph model G_1 are set as $\sigma_c = 1.3$, $\sigma_d = 1.6$. Constants that control the strength of the weights in graph model G_2 are set as $\sigma_c = 1.4$, $\sigma_d = 1$, $\sigma_s = 1.4$. Proportions of the highest and lowest depth value are set as $\delta_h = 30\,\%$, $\delta_l = 3\%$.

6.1 3D boundary prior evaluation

In order to verify the contribution of using 3D boundary, we use four sides of RGB image and six planes of RGB-D image to measure saliency performance respectively, and results are illustrated in Fig. 6a, b. Note that color and depth feature are selected in the experiment and 3D compactness feature is not included.

Evaluation results show that method with six planes is superior to that with four sides as background seeds.

6.2 3D compactness feature evaluation

In order to evaluate the effect of 3D compactness feature, we use color, color+depth, and color+depth + 3D compactness feature to evaluate saliency performance respectively, and results are illustrated in Fig. 7a, b. Note that six planes are selected as background seeds in the experiment.

Evaluation results show that F-measure score is increased obviously and incorporation 3D compactness feature into color and depth feature plays an important role on RGB-D saliency detection.

6.3 Comparison with other method

The proposed RGB-D saliency method is compared against the following state-of-the-art RGB-D saliency detection model: LMH [23], ACSD [12], GP [26], FP [7], MGMR [33], SE [25]. Experiment results illustrated in Fig. 8 show that the proposed method consistently outperforms all the other RGB-D saliency methods. Especially our method is the improvement for MR [34], and is the same as but superior to MGMR [33]. In addition GP [26] has the excellent performance among these RGB-D saliency methods, but its F-measure score is only 0.7220 and is slightly lower than ours 0.7253 even if it has higher precision. Visual performance comparisons of the proposed method and six existing RGB-D methods are illustrated in Fig. 9.

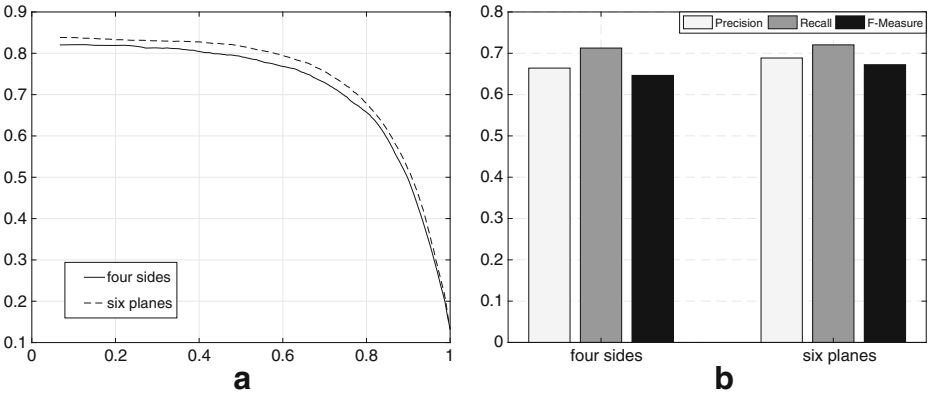


Fig. 6 Evaluation between four sides and six planes as background seeds. **a** Precision–recall curve **b** F-measure

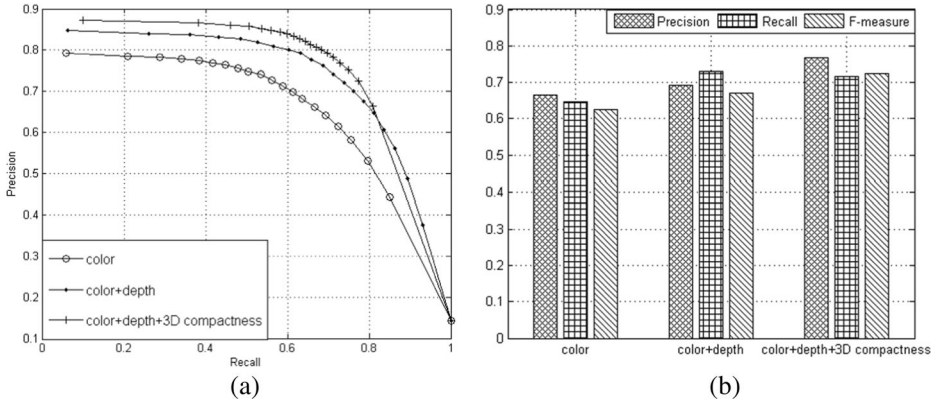


Fig. 7 Evaluation between color, color+depth and color+depth + 3D compactness feature. **a** Precision–recall curve **b** F-measure

6.4 Computational cost and runtime

Time complexity of the proposed algorithm is $O(n^2)$, where n indicates the number of superpixels in each image. Algorithm consists of four parts: (1) superpixel segmentation (2) 3D compactness feature measurement (3) graph model construction (4) manifold ranking. Time complexity of the first part is $O(n^*t)$, where t indicates the number of iterations. Time complexity of the second part is $O(n^2)$, in which graph model construction with color and depth feature costs $O(n^2)$, sv value and sd value cost $O(n^2)$. Time complexity of the third part is $O(n^2)$, in which graph model construction using color, depth and 3D compactness feature costs $O(n^2)$. Time complexity of the fourth part is $O(n^2)$ too, in which manifold ranking is executed in background and foreground perspective respectively.

Average runtime comparison for the proposed method and six existing RGB-D methods is illustrated in Table 1. The measurement environment was Inter(R) Core(TM)

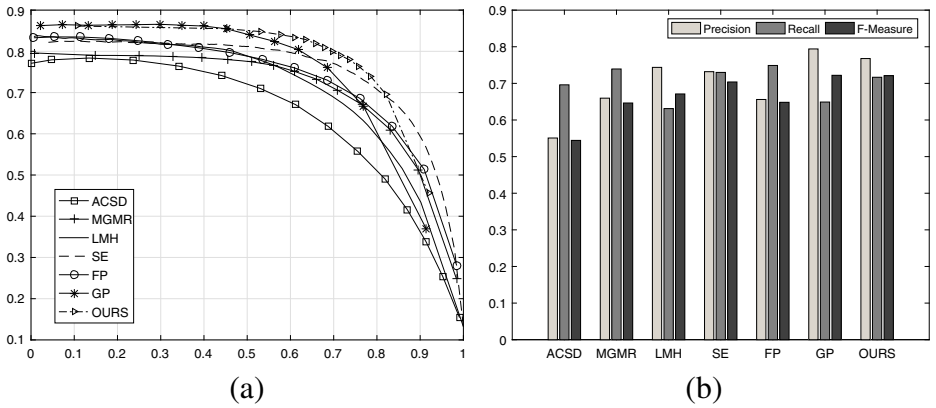


Fig. 8 Quantitative comparisons for the proposed method and six existing RGBD method: **a** Precision–recall curve **b** F-measure

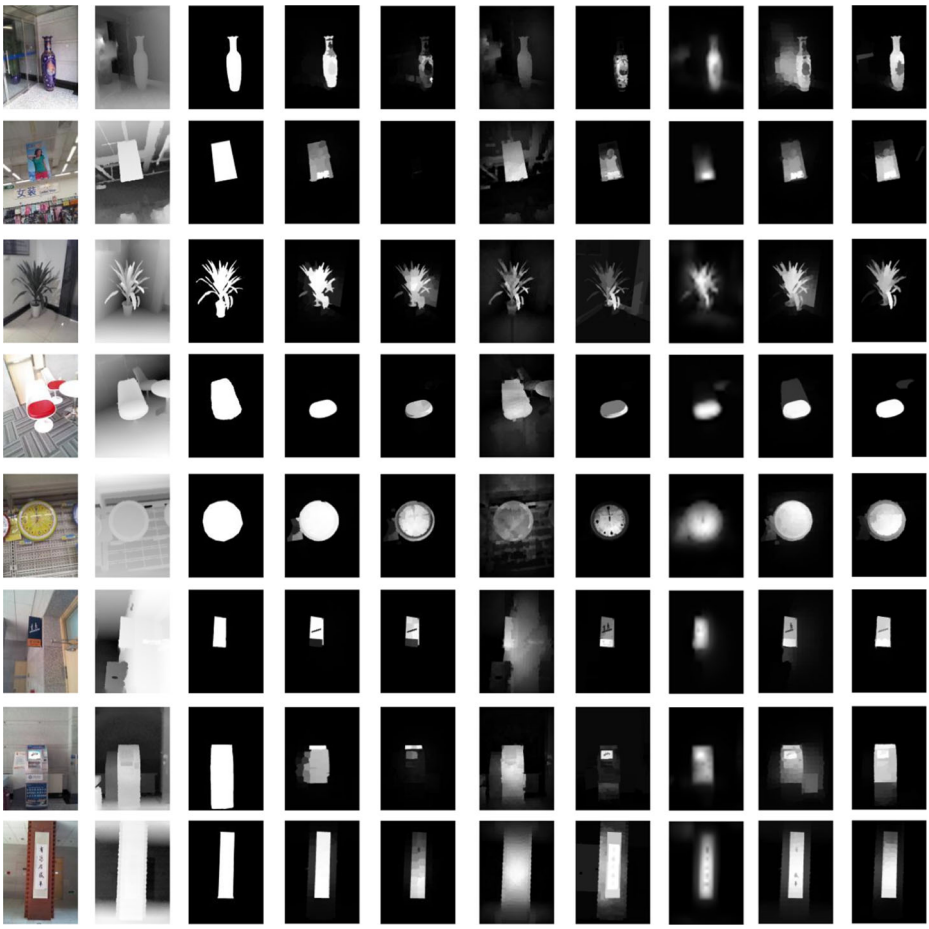


Fig. 9 Visual performance comparisons of the proposed method and six existing RGBD methods. From left to right: RGB image, Depth image, Ground Truth, Ours, LMH [23], ACSD [12], GP [26], FP [7], MGMR [33] and SE [25]

i5–4700 3.20GHz, RAM 8GB. All the algorithms directly operate the source codes published by the author under Matlab 2015b platform except for ACSD [12] under visual studio 2010. From the results we can see that GP [26] costs the longest time to get higher precision, while our method has highest F-measure score and less time cost against the others.

Table 1 Comparison of average runtime per image for the proposed method and six existing RGB-D method

Method	ACSD [12]	MGMR [33]	LMH [23]	SE [25]	FP [7]	GP [26]	OURS
Time(s)	0.062	2.889726	1.808605	2.54974	0.298109	7.7697772	1.9510697
Code	C++	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab

7 Conclusion

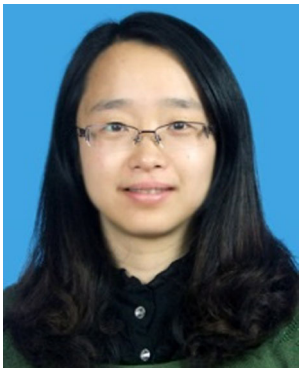
In this paper, we have proposed 3D boundary prior in which salient object are more likely apart from top, down, left, right, front, back plane in 3D scene rather than image borders consists of four sides of RGB image. Then we combine 3D compactness, in which salient objects typically have 3D compact spatial distributions, with color and depth feature to express similarity between adjacent supervoxels. Last we apply manifold ranking to get background prior saliency map and then refine it from foreground view. Both evaluation and comparison results on publicly available RGB-D datasets show the effectiveness, superiority and effectiveness of the proposed method.

Acknowledgements We thank Prof. Jianguo Wu from Anhui University for helping with acquisition of funding. We also thank all anonymous reviewers for their valuable comments. This research is supported by National Key Technology Research and Development Program of the Ministry of Science and Technology of China (2015BAK24B01), Key Program of Natural Science Project of Educational Commission of Anhui Province, China (KJ2015A009), Open issues on Co-Innovation Center for Information Supply & Assurance Technology, Anhui University (ADXXBZ201610).

References

1. Achanta R, Shaji A, Smith K et al (2012) SLIC superpixels compared to state-of-the-art superpixel methods [J]. IEEE Trans Pattern Anal Mach Intell 34(11):2274–2282
2. Ahmad J, Sajjad M, Mehmood I et al (2015) Saliency-weighted graphs for efficient visual content description and their applications in real-time image retrieval systems[J]. J Real-Time Image Proc 13(3): 431–447
3. Chaib S, Gu Y, Yao H et al (2016) A VHR scene classification method integrating sparse PCA and saliency computing[C]. Geoscience and Remote Sensing Symposium (IGARSS). 2016 IEEE International, pp 2742–2745
4. Cheng Y, Fu H, Wei X et al (2014) Depth enhanced saliency detection method[C]. Proceedings of international conference on internet multimedia computing and service. ACM, pp 23
5. Desingh K, Madhava K K, Rajan D et al (2013) Depth really matters: improving visual salient region detection with depth[C]. British machine vision conference (BMVC), pp 98.1–98.11
6. Feng D, Barnes N, You S et al (2016) Local background enclosure for RGB-D salient object detection[C]. Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on. IEEE, pp 2343–2350
7. Guo J, Ren T, Bei J et al (2015) Salient object detection in RGB-D image based on saliency fusion and propagation[C]. Proceedings of the 7th international conference on internet multimedia computing and service. ACM, pp 59
8. Hong S, You T, Kwak S et al (2015) Online tracking by learning discriminative saliency map with convolutional neural network[C]. International Conference on Machine Learning (ICML), pp 597–606
9. Hussain CA, Rao DV, Masthani SA (2016) Robust pre-processing technique based on saliency detection for content based image retrieval systems[J]. Procedia Computer Science 85:571–580
10. Hussain CA, Rao DV, Masthani SA (2016) Image retrieval using saliency content[C]. Electrical, electronics, and optimization techniques (ICEEOT), international conference on. IEEE, pp 840–844
11. Jiang L, Koch A, Zell A (2015) Salient regions detection for indoor robots using RGB-D data[C]. 2015 IEEE International conference on robotics and automation (ICRA). IEEE, pp 1323–1328
12. Ju R, Ge L, Geng W et al (2014) Depth saliency based on anisotropic center-surround difference[C]. 2014 IEEE international conference on image processing (ICIP). IEEE, pp 1115–1119
13. Ju R, Liu Y, Ren T et al (2015) Depth-aware salient object detection using anisotropic center-surround difference[J]. Signal Process Image Commun 38:115–126
14. Khanna MT, Rai K, Chaudhury S et al (2015) Perceptual depth preserving saliency based image compression[C]. Proceedings of the 2nd international conference on perception and machine intelligence. ACM, pp 218–223

15. Krause EA, Zillich M, Williams TE et al (2014) Learning to recognize novel objects in one shot through human-robot interactions in natural language dialogues[C]. AAAI, pp 2796–2802
16. Lang C, Nguyen T V, Katti H et al (2012) Depth matters: influence of depth cues on visual saliency[M]. Computer Vision–ECCV 2012. Springer Berlin Heidelberg, pp 101–115
17. Li Z, Liu G, Zhang D et al (2016) Robust single-object image segmentation based on salient transition region[J]. Pattern Recogn 52:317–331
18. Liu GH, Yang JY, Li ZY (2015) Content-based image retrieval using computational visual attention model[J]. Pattern Recogn 48(8):2554–2566
19. Margolin R, Zelnik-Manor L, Tal A (2014) How to evaluate foreground maps?[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, pp 248–255
20. Oh SJ, Benenson R, Khoreva A et al (2017) Exploiting saliency for object segmentation from image level labels[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4410–4419
21. Otsu N (1979) A threshold selection method from gray-level histograms[J]. IEEE Trans Syst Man Cybern 9(1):62–66
22. Papushoy A, Bors AG (2015) Image retrieval based on query by saliency content[J]. Digital Signal Processing 36:156–173
23. Peng H, Li B, Xiong W et al (2014) Rgbd salient object detection: a benchmark and algorithms[C]. European conference on computer vision. Springer, Cham, pp 92–109
24. Qi M, Wang Y (2016) DEEP-CSSR: scene classification using category-specific salient region with deep features[C]. 2016 IEEE international conference on image processing (ICIP). IEEE, pp 1047-1051
25. Quo J, Ren T, Bei J (2016) Salient object detection for RGB-D image via saliency evolution[C]. 2016 IEEE international conference on multimedia and expo (ICME). IEEE, pp 1–6
26. Ren J, Gong X, Yu L et al (2015) Exploiting global priors for RGB-D saliency detection[C]. 2015 IEEE conference on computer vision and pattern recognition workshops (CVPRW). IEEE, pp 25–32
27. Song H, Liu Z, Du H et al (2015) Saliency detection for RGBD images[C]. Proceedings of the 7th international conference on internet multimedia computing and service. ACM, pp 72
28. Srivastava S, Mukherjee P, Lall B (2016) Adaptive image compression using saliency and KAZE features[C]. International conference on signal processing and communications, pp:1–5
29. Sun J, Wu J, Yu H et al (2016) Boosting image retrieval framework with salient objects[C]. Audio, language and image processing (ICALIP), 2016 international conference on. IEEE, pp 241–245
30. Tamura Y, Akashi T, Yano S et al (2016) Human visual attention model based on analysis of magic for smooth human–robot interaction[J]. Int J Soc Robot 8(5):685–694
31. Tang Y, Tong R, Tang M et al (2016) Depth incorporating with color improves salient object detection[J]. Vis Comput 32(1):111–121
32. Wu P, Duan L, Kong L (2015) RGB-D salient object detection via feature fusion and multi-scale enhancement[C]. CCF Chinese Conference on Computer vision. Springer Berlin Heidelberg, pp 359–368
33. Xue H, Gu Y, Li Y et al (2015) RGB-D saliency detection via mutual guided manifold ranking[C]. Image Processing (ICIP), 2015 IEEE international conference on. IEEE, pp 666–670
34. Yang C, Zhang L, Lu H et al (2013) Saliency detection via graph-based manifold ranking[C]. Computer vision and pattern recognition (CVPR), 2013 IEEE Conference on. IEEE, pp 3166–3173
35. Zhang X, Chen X (2016) Robust sketch-based image retrieval by saliency detection[C]. International conference on multimedia modeling. Springer, Cham, pp 515–526
36. Zhang F, Du B, Zhang L (2015) Saliency-guided unsupervised feature learning for scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing 53(4):2175–2184
37. Zhang L, Chen J, Qiu B (2016) Region-of-interest coding based on saliency detection and directional wavelet for remote sensing images[J]. IEEE Geosci Remote Sens Lett 14(1):23–27
38. Zhou D, Weston J, Gretton A et al (2004) Ranking on data manifolds[C]. Adv Neural Inf Proces Syst 16: 169–176
39. Zhou L, Yang Z, Yuan Q et al (2015) Salient region detection via integrating diffusion-based compactness and local contrast[J]. IEEE Trans Image Process 24(11):3308–3320
40. Zhu S, Bo Y, He L (2016) Robust multi-feature visual tracking with a saliency-based target descriptor[C]. Control conference (CCC), 2016 35th Chinese. IEEE, pp 5008–5013
41. Zhu G, Wang J, Wu Y et al (2016) MC-HOG correlation tracking with saliency proposal[C]. Thirtieth AAAI Conference on Artificial Intelligence. AAAI Press, pp 3690–3696
42. Zhu W, Liang S, Wei Y et al Saliency optimization from robust background detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2814–2821



Zhengyi Liu is an associate professor in School of Computer Science and Technology, Anhui University, China. She received her B.S., M.S., and Ph.D. from Anhui University, China in 2001, 2004 and 2007 respectively. Her research interests include image and video processing, computer vision and machine learning.

Tengfei Song is a M.S. Candidate of Anhui University. He received his B.S. from Taishan College of Science and Technology, China in 2015. His research interests include image and video processing and computer vision.

Feng Xie is a M.S. Candidate of Anhui University. He received his B.S. from Anhui University, China in 2016. His research interests include image and video processing and computer vision.