RESEARCH ARTICLE

A salient region detection model combining background distribution measure for indoor robots

Na Li, Hui Xu, Zhenhua Wang, Lining Sun, Guodong Chen*

Robotics and Microsystem Center, Soochow University, Suzhou, Jiangsu, China

* guodongxyz@163.com

OPEN ACCESS

Citation: Li N, Xu H, Wang Z, Sun L, Chen G (2017) A salient region detection model combining background distribution measure for indoor robots. PLoS ONE 12(7): e0180519. https://doi.org/10.1371/journal.pone.0180519

Editor: Dewen Hu, National University of Defense Technology College of Mechatronic Engineering and Automation, CHINA

Received: February 15, 2017

Accepted: June 17, 2017

Published: July 24, 2017

Data Availability Statement: All relevant data are within the paper.

Funding: This project was funded by the National Natural Science Foundation of China (Grant No. 61305118 and No. U1509202) and the National Natural Science Foundation of Jiangsu (Grant No. BK20130329). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Vision system plays an important role in the field of indoor robot. Saliency detection methods, capturing regions that are perceived as important, are used to improve the performance of visual perception system. Most of state-of-the-art methods for saliency detection, performing outstandingly in natural images, cannot work in complicated indoor environment. Therefore, we propose a new method comprised of graph-based RGB-D segmentation, primary saliency measure, background distribution measure, and combination. Besides, region roundness is proposed to describe the compactness of a region to measure background distribution more robustly. To validate the proposed approach, eleven influential methods are compared on the DSD and ECSSD dataset. Moreover, we build a mobile robot platform for application in an actual environment, and design three different kinds of experimental constructions that are different viewpoints, illumination variations and partial occlusions. Experimental results demonstrate that our model outperforms existing methods and is useful for indoor mobile robots.

Introduction

Due to population ageing [1], the cost of health care is raising in recent years and indoor robots will play significant roles in our daily life. Vision system, the most important perception vehicle for indoor robot, obtains a flood of visual information to perceive and understand the surrounding world, but as is often the case that only a few regions are relevant to a given context. Therefore, indoor robot is expected to possess the information management skill to capture useful visual information regard to a specific task. Inspired by the primate-customized visual attention mechanism, which is simple, yet extremely robust: select the most relevant information among the plethora of visual information [2, 3], saliency detection has been extensively studied to deal with the enormous amount of information and enhance computational efficiency. It has been used in many applications including object recognition [4–6], image segmentation [7, 8], image retrieval [9], visual tracking [10–12] and human-robot interaction (HRI) [1, 13–19].

Plenty of saliency detection methods proposed have tested on the public datasets and performed well. However, to be the best of our knowledge, most of the existing methods hold true for natural images with simple background and single salient object, but they cannot work well for indoor scene with complex backgrounds, several salient objects and illumination variations. We also notice that most of the proposed methods are based on visual features in 2D scenes, and the depth information is ignored even though it is essential cues for human beings to perceive the world. Inspired by [20–22], we propose a new combined salient region detection model that integrates background distribution into primary saliency. The framework of the proposed method can be seen in Fig 1.

The rest of this paper is organized as follows. Related works are provided in Section 2. Our model is descried in detail in Section 3. Experiments on two public datasets and the robot platform are introduced and analyzed in Section 4 and Section 5, respectively. Conclusions and future work are listed in Section 6.

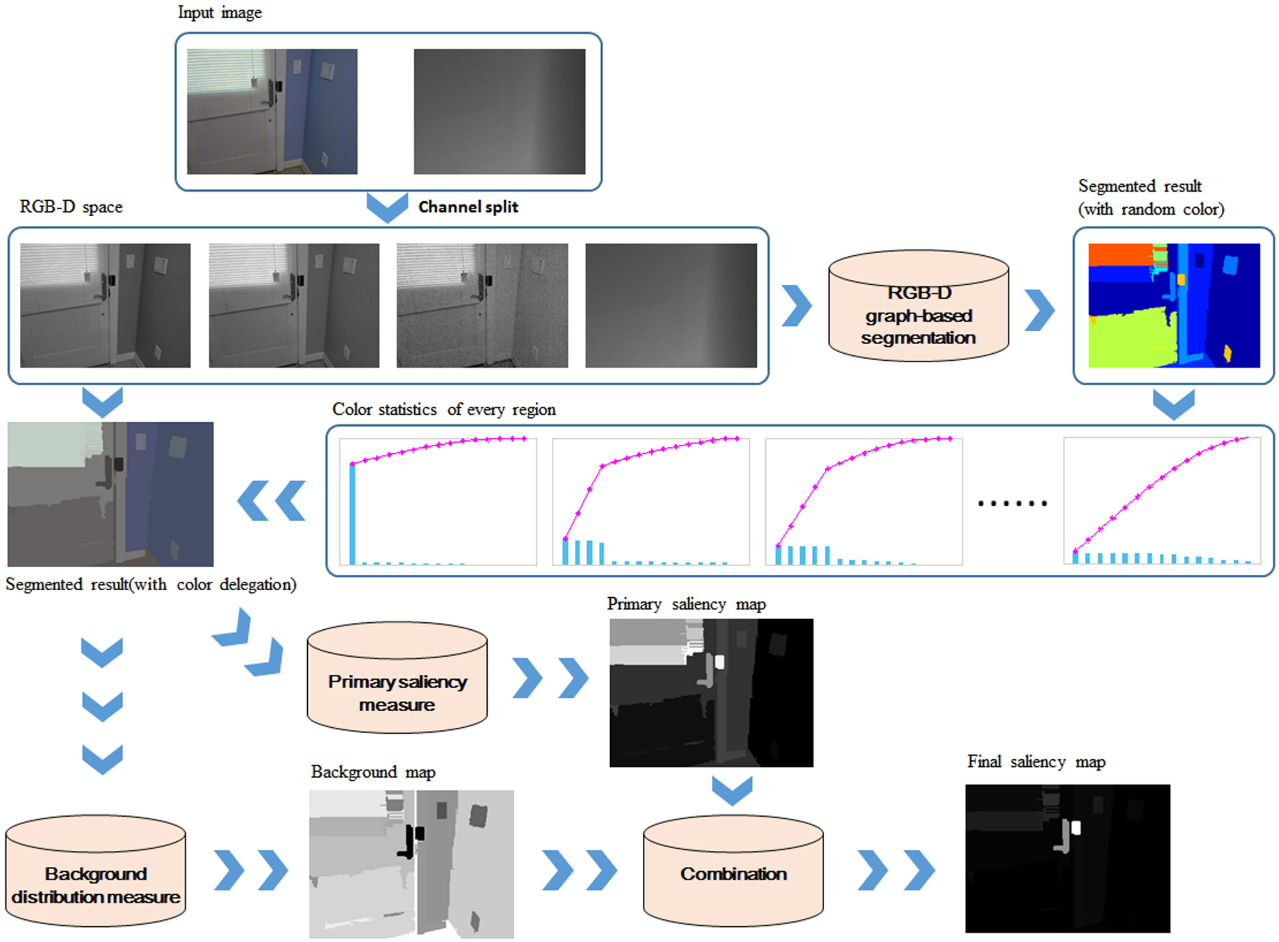
Input image
RGB-D space
Channel split
RGB-D graph-based segmentation
Segmented result (with random color)
Color statistics of every region
……
Segmented result(with color delegation)
Primary saliency measure
Primary saliency map
Background map
Background distribution measure
Combination
Final saliency map

Fig 1. The framework of the proposed method. Firstly, to keep the completeness and compactness of salient region candidates, an input image is divided into regions using color(RGB) and depth cues. Secondly, a color delegation are adaptively selected for each region and used to measure the primary saliency. Thirdly, the background distribution is assigned by the spatial distribution and the compactness of each region. Finally, the primary saliency map and the background distribution map are combined to reduce the false positive of background regions and false negative of salient regions.

https://doi.org/10.1371/journal.pone.0180519.g001

Related work

In this section, related work of saliency object detection are introduced in two aspects: computer vision and robot application.

Currently, salient region detection methods proposed in the field of computer vision have sprung up and gained widespread popularity in many applications including robots. They utilize some different features (such as color, intensity, edge, face and person) to determine the possibility to be attended of a pixel [20, 21, 23–25], superpixel [26, 27] or region [20, 22, 28, 29]. The spatial attention model by Zhai [23] computes a pixel-level saliency map by the intensity histogram of input image, whose attended regions are detected by the region growth technique and marked with bounding boxes. Following in Zhai [23], two models proposed by Cheng [20], the histogram-based contrast (HC) and the region-based contrast(RC), perform well in natural images. In HC, saliency is measured for each pixel by color contrast to all other pixels. In RC, after graph-based segmentation [30] a region's saliency is computed by color contrast and spatial distance to others in the image. The frequency-tuned (FT) saliency detection approach [25] uses band-pass filtering to produce full-resolution saliency maps. Its low computational complexity makes it highly suitable for real-time applications on a mobile robot. In Federico's model [26], the input image is abstracted into elements using the SLIC superpixel technique and the saliency estimation is obtained from the uniqueness and the spatial distribution of these elements. Developed from Itti model, the model proposed by Wang [29] assigns saliency based local and global saliency information of each segmented region using color and orientation feature. An optimization framework presented by Zhu [22] combines background detection, and boundary connectivity is proposed to quantify how heavily a region is connected to image boundaries to ensure high accuracy background detection. An method by Jiang [21] is proposed to detect salient regions by mapping pixels into foreground and background regions in RGB-D images.

Visual saliency models for robots could be generally divided into two categories: overt visual attention models and covert visual attention models. For the overt visual attention model, the research works concentrate on the camera maneuvering mechanism based on the principle of overt visual attention [31]. In [14], a fast approximation to a Bayesian model of visual saliency is proposed to orient a camera as quickly as possible toward human faces. Results show that it can provide saliency maps in about 10ms per 160 ×120 pixel video frame. For the real-time implementation, they only use image intensity channel, not color channels As described in detailed previously [32], a saliency-based "lazy" approach is proposed for scene exploration of newly entered rooms that reduces the amount of necessary head movement while it strongly favors to attend the most salient proto-objects as soon as possible. For the covert attention model, they give emphasis on a sensory stimulus mentally without changing gaze direction. In the intelligent system described in [19], vision saliency is used to restrict the total number of possible gazes to a smaller set that still contains salient objects. In [17] a biologically inspired vision system is presented for human-robot interaction. In order to reduce clutter and improve gesture recognition rates, visual saliency, computed by color, motion and disparity cues, is used to segregate hands from the background. A real-time parallel model for saliency maps in [18] is put forward to predict gaze directions. IMAPCAR2, an image recognition processor with advanced parallel processing capabilities, is exploited to improve the operating efficiency. In [21] an effective method is proposed to detect salient regions by mapping pixels into foreground and background regions in RGB-D images.

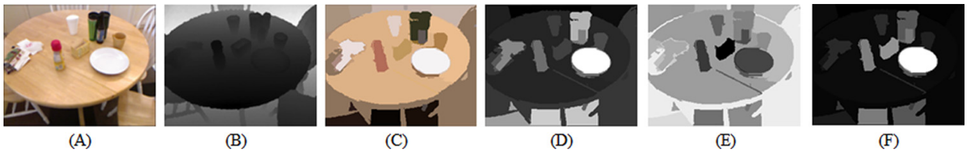

Fig 2. An illustrative example of our method. A: RGB image. B: depth map. C: the segmented result. For display, each region is shown in the mean color. Note that in our model, each region is assigned with a color delegation that is obtained by picking more frequently occurring colors. D: the primary saliency map. E: the background distribution map. F: the final saliency map.

https://doi.org/10.1371/journal.pone.0180519.g002

Methods

In this section, we propose a new approach for salient region detection with the combination of primary saliency and background distribution. The illustration of the proposed method is shown in Fig 2. In the following subsections, each step will be presented in detail.

Segmentation

It is essential for indoor robots to perceive the surrounding environment, localize desired objects, and know their shapes and sizes. Therefore, keeping the completeness of salient objects is the premise and guarantee to accomplish tasks for indoor robots. Obviously, Pre-segmentation is an effective approach to meet the requirement. Depth information is demonstrated to well hold the integrality of foreground objects and strong contours in the input image even in an intricate scenario. Therefore, the graph-based RGB-D segmentation [21, 30] is introduced taking advantage of depth and color information. Thanks to the segmentation, salient region candidates tend to be complete.

The segmentation example is shown in Fig 2C. Notice that the segmented result in mean color is just for display. It can be seen that each object on the table is segmented into one single region nearly. For example, the glass bottle with a red lid and a white label is separated into one region. It implies that graph-based RGB-D segmentation can capture completeness of objects excellently. Completeness of objects and homogeneous color property of regions is inherently a pair of contradiction and it is difficult to balance them, therefore we discard general representation which uses mean color values to characterize each region, and adopt a new depiction that some representative colors of each region are picked adaptively. Details of the representative colors selectivity and primary saliency measure are described in the next subsection.

Primary saliency measure

The commonly employed assumption, regions which stand out from other regions in the image tend to catch human attention and are regarded as salient, supports the implementation of recent contrast based saliency measure [20–24, 26, 29]. Saliency maps of these approaches are calculated based on some visual features, such as color, depth, orientation and contour, or spatial distribution. Here a primary saliency map and a background distribution map are calculated using two kinds of visual features (color and depth) and spatial layout of each region, respectively.

In order to keep the comprehensiveness of color information and balance it with the computational efficiency, we build a color histogram for each region and pick more frequently occurring colors as the color delegation. Specifically, each color channel in RGB color space is first quantized to N bins [20] so that the total color number drops from 256^3 to N^3, and a

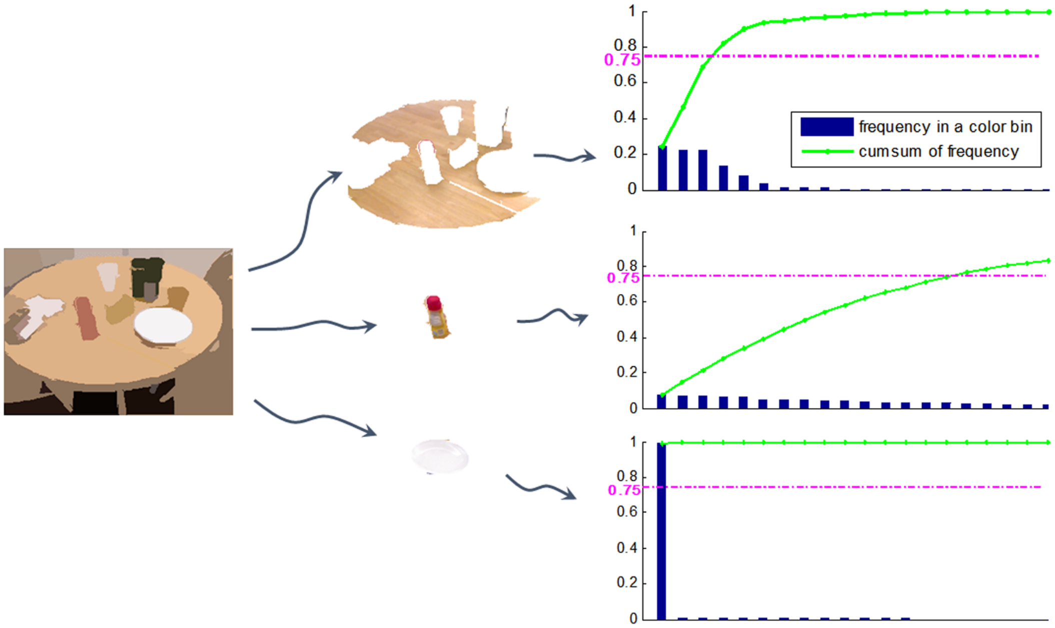

Fig 3. An example of the color delegation. The segmented result (first column) presented by mean color values, three segmented regions (second column) by original RGB information and color statistics (third column) of respective regions where the blue histogram is sorted in the non-increasing order of color occurrence frequency, and the green line with asterisk mark represents the cumulative sum of the corresponding histogram. The color delegation of a region consists of more frequently occurring colors which cover more than a certain proportion of this region area. We introduce a threshold for the color delegation selection, and it is set to 0.75 as shown by the magenta dashed line. For each region, intuitively, color bins below the magenta dashed line and the next one compose its color delegation.

https://doi.org/10.1371/journal.pone.0180519.g003

N^3-bin histogram is built to denote color properties for each region. After these bins are sorted in the non-increasing order, more frequently occurring colors which cover more than φ of this region area will be selected as its color delegation. An example result is shown in Fig 3 where we set $N = 8$ and $\varphi = 0.75$. The example shows that the glass bottle can be described by 15 color values, while the white plate can be depicted by only one.

Thanks to the region representation based on the color delegation, the primary saliency can be measured next. Salient regions should be distinctive and high-contrast compared with other regions in the image. We evaluate the primary saliency by contrasting colors of a region with all other regions,

$$RS(r_i) = \sum_{i\neq j} \omega_j * D_c(i,j), \tag{1}$$

in which $D_c(i, j)$ is the color distance of region r_i and r_j with the color delegation $c_i = \{c_i^1, c_i^2, \ldots\ldots, c_i^{n_i}\}$ and $c_j = \{c_j^1, c_j^2, \ldots\ldots, c_j^{n_j}\}$ respectively. Region $r_i(r_j)$ has $n_i(n_j)$ representative colors and each representative color value is obtained by the mean value of pixels dropping into the relevant bin.

$$D_c(i,j) = \sum_{p=1}^{n_i}\sum_{q=1}^{n_j} f(c_i^p) * f(c_j^q) * \parallel c_i^p - c_j^q \parallel, \tag{2}$$

in which c_i^p is the p-th representative color of region r_i and $f(c_i^p)$ is the occurrence frequency of c_i^p in region r_i. Here we utilize occurrence frequencies as the weight of color distance between two representative colors to emphasize high-frequency ones.

$$\omega_j = AR(r_j) * e^{-\alpha * D_s(r_i, r_j)}. \tag{3}$$

In Eq (1), ω_j is the weight of region r_i and influenced by two factors $AR(r_j)$ and $D_s(r_i, r_j)$. $AR(r_j)$ is the area ratio of region r_j in the input image and $D_s(r_i, r_j)$ is the distance between centroids of region r_i and r_j. α is the scaling factor to control the strength of $D_s(r_i, r_j)$.

Background distribution measure

In indoor environments with complex backgrounds, several salient objects and illumination variations, the global contrast method based on saliency measure often causes some undesired outcomes. For instance, in Fig 2D the floor intuitively regarded as region of background is labeled with relatively high saliency values, while the carton close to the table in color domain isn't underscored.

We discover that, generally, salient object regions have the property of compact sizes while background ones distribute widely and near image boundaries. Based on the hypothesis, background distribution $BD(r_i)$ is introduced to reduce and even exclude false negative and false positive from the previous step,

$$BD(r_i) = \omega_{BD}(r_i) * e^{-\beta * rd(r_i)}, \tag{4}$$

in which $\omega_{BD}(r_i)$ is the Gaussian weight of region r_i, and β controls the strength of the region roundness. $rd(r_i)$, called region roundness, is a new proposed measure to describe the compactness of a region. It is defined as:

$$rd(r_i) = \frac{4\pi * A(r_i)}{L_c(r_i)^2}, \tag{5}$$

in which $A(r_i)$ and $L_c(r_i)$ are area and contour length of region r_i respectively. The more this value is, the more compact relevant region is. Region roundness is usually large for salient object regions and small for background regions. Fig 4 can support our inference where the blue region which we believe the most compact has the maximum of region roundness and the

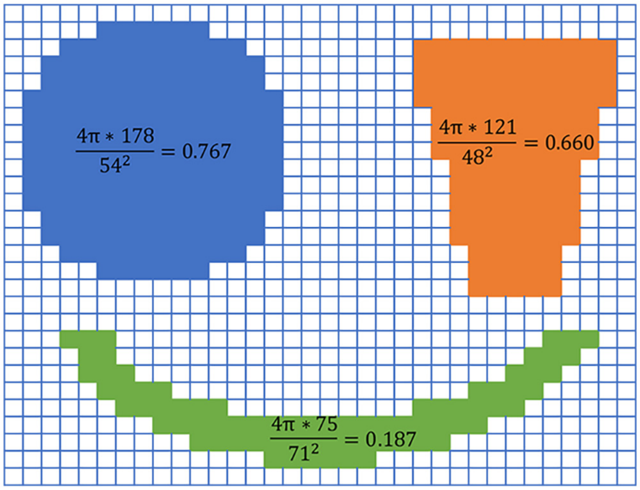
$\frac{4\pi * 178}{54^2} = 0.767$
$\frac{4\pi * 121}{48^2} = 0.660$
$\frac{4\pi * 75}{71^2} = 0.187$

Fig 4. An illustration of region roundness. The synthetic image comprises three regions with different compactness. Each region roundness value is displayed on it.

https://doi.org/10.1371/journal.pone.0180519.g004

green one which is dispersed apparently is allocated a smallest value. Therefore, the region roundness measure is available to quantify region characteristics.

Spatial layout in images have the universality that background regions can be easily connected to image boundaries while foreground objects cannot. Utilizing boundary connectivity [22] and distance to the image center of each region, we set a valid weight $\omega_{BD}(r_i)$ to influence background distribution measure.

$$\omega_{BD}(r_i) = 1 - e^{-\left\{\frac{BC(r_i)}{\delta_{BC}}+\frac{DC(r_i)}{\delta_{DC}}\right\}}, \tag{6}$$

in which $DC(r_i)$ is the distance between the image center and region r_i, and $BC(r_i)$ is the boundary connectivity of region r_i to quantify how heavily a region is connected to image boundaries. δ_{BC} and δ_{DC} control the strength of weighting of $BC(r_i)$ and $DC(r_i)$ respectively. Boundary connectivity is first defined in Zhu's model [22] as:

$$BC(r_i) = \frac{L_b(r_i)}{\sqrt{A(r_i)}}. \tag{7}$$

For simplicity and efficiency, $L_b(r_i)$ is set to be the pixel numbers along image boundaries in region r_i.

Combination

By now, we have already obtained a primary saliency map and a background distribution map. To inhibit false positive of background regions and raise false negative of salient regions, we combine the two maps in a way of an exponential function:

$$Sal(r_i) = RS(r_i) * e^{-\gamma * BD(r_i)}. \tag{8}$$

Note that primary saliency $RS(r_i)$ and background distribution $BD(r_i)$ are both normalized. The parameter γ is the scaling factor for the exponential to control the range of the background distribution measure.

As can be seen in Fig 2, input images ((A) and (B)) are first merged into homogeneous regions as in (C). Two samples of primary saliency map and background distribution map are shown in (D) and (E). The resulting saliency map is shown in (F).

Experiments on datasets

Datasets

The proposed model is aimed at salient object detection in the complicated indoor environment. To evaluate the proposed model, we select two datasets with complex scenes: DSD [33] and ECSSD [28].

DSD dataset [33] comprises 80 color and depth image pairs with associated pixel-wise ground truth segmentation masks. The dataset is obtained in a real-world indoor environment and used for studies of depth-based salient detection (such as [21]). Scenes in this dataset are assigned to multiple foreground objects of potential interest and complex background structure.

ECSSD [28] dataset consists of 1000 color images with associated pixel-wise binary masks. It includes a large number of semantically meaningful but structurally complex natural images [34]. Note that ECSSD don't cover any depth images, the reason we select this dataset is just to check the performance and universality of our model in some complex scenes. Depth information is ignored in the experiments with ECSSD, namely the graph-based RGB-D segmentation turns into the ordinary graph-based RGB segmentation.

Evaluation criteria

The precision-recall curve (PR curve), F-measure and the mean absolute error (MAE) [34], three universally-agreed, standard and easy-to-understand measures, are selected to evaluate the detection accuracy of the proposed model.

The PR curve is based on the overlapping area between saliency prediction and ground truth segmentation masks. For a saliency map, a fixed threshold which changes from 0 to 255 is set to convert *S* to a binary mask *M*, and *Precision* and *Recall* are computed by comparing *M* and ground truth *G*,

$$Precision = \frac{\| M \cap G \|}{\| M \|}, \tag{9}$$

$$Recall = \frac{\| M \cap G \|}{\| G \|}, \tag{10}$$

where $\|\bullet\|$ represents the number of non-zero entries in the mask. On each threshold, a pair of precision/recall scores are computed, and are finally combined to form a PR curve. The PR curve is commonly used to reliably compare how well various saliency maps highlight salient regions in images.

Besides, proposed by Achanta [25], the another way to partition a saliency map *S* is to use an image-dependent adaptive threshold, which is computed as twice as the mean saliency of *S*,

$$T_a = \frac{2}{W * H} \sum_{x=1}^{W} \sum_{y=1}^{H} \| S(x,y) \| . \tag{11}$$

where *W* and *H* are the width and the height of the respective saliency map. And this mode of calculating *Precision* and *Recall* will be used by $F_{measure}$.

Usually, neither *Precision* nor *Recall* can comprehensively evaluate the quality of a saliency map. To this end, the *F-measure* is proposed as a weighted harmonic mean of them with a non-negative weight ζ [35]:

$$F_{measure} = \frac{(1 + \zeta^2) * Precision * Recall}{\zeta^2 * Precision + Recall}. \tag{12}$$

As suggested by many salient region detection works [20, 25, 26], ζ^2 is set to 0.3 to increase the importance of the *Precision* value. The reason for weighting precision more than recall is that recall rate is not as important as precision [24]. Here, the saliency map *S* can be binarized with the threshold from the Eq (11), and the presion and recall values can be calculated by the Eq (9) and the Eq (10).

The overlap-based evaluation measures introduced above do not consider the true negative assignments of the saliency map, i.e., pixels correctly marked as non-salient. For a more comprehensive comparison, the MAE is introduced as the second evaluation criteria to compensate the PR curve, and is used in recent methods [21, 22, 26, 34]. The *MAE* aims to measure how close a saliency map *S* is to the ground truth *G*,

$$MAE = \frac{1}{W * H} \sum_{x=1}^{W} \sum_{y=1}^{H} \| S(x,y) - G(x,y) \|, \tag{13}$$

where *S* and *G* are both normalized to the range [0, 1].

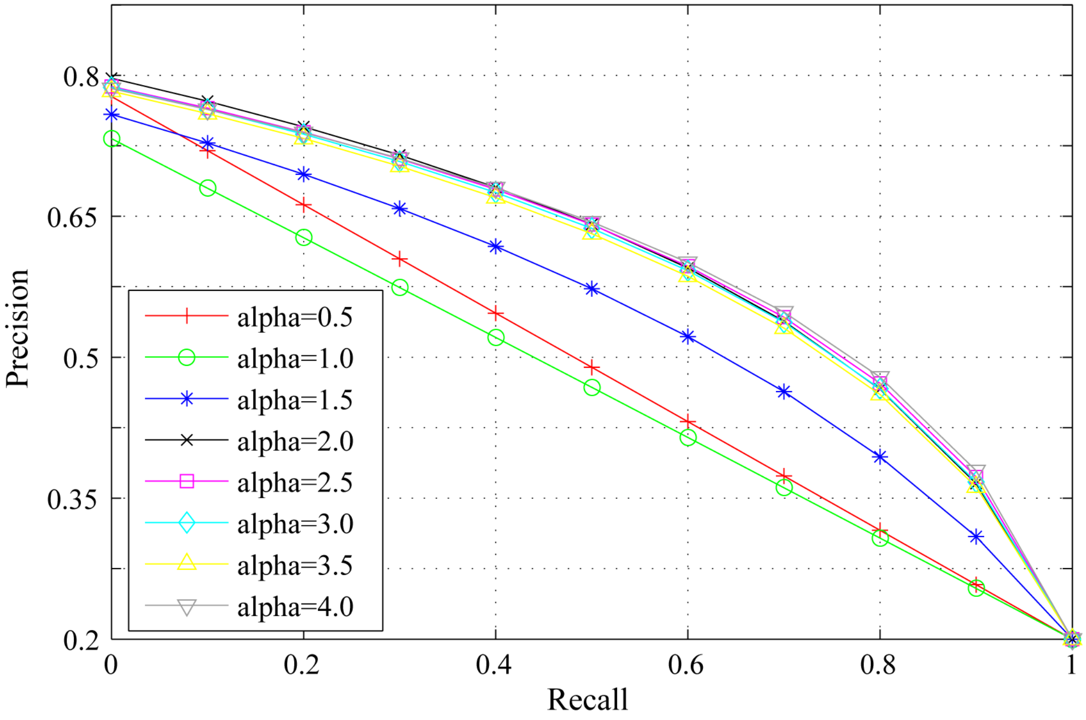

Fig 5. PR curves for different α on the DSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g005

Parameter evaluation

The proposed model brings in several parameters that are α in the Eq (3), β in the Eq (4), γ in the Eq (8) and δ_{BC} and δ_{DC} in the Eq (6). To investigate the effect of these parameters in the model [36], we conduct some experiments using DSD dataset.

α is the scaling factor to control the strength of spatial distance between regions. If α is too small, the weight of spatial distance between regions($e^{-\alpha * D_s(r_i, r_j)}$ in the Eq (3)) will increase(close to 1), which will lead to expand the influence of spatial features during the primary saliency measure. If α is too large, the influence of spatial features will reduce during the primary saliency measure. Figs 5, 6 and 7 show the PR curves, F-measure and MAE for different α on the DSD dataset respectively. When α is greater than 2.0, the performance of different PR curves is similar. According to bar charts of F-measure and MAE, the performance reaches its peak when $\alpha = 3.0$. The performance improves gradually with α when α is smaller than 3.0 and reduces speedily with α when α is larger than 3.0. Therefore, we set $\alpha = 3.0$ through the experiment process.

β is the scaling factor to control the strength of region roundness. The influence of β on region roundness is the same as that of α on spatial distance between regions. Figs 8, 9 and 10 show the PR curves, F-measure and MAE for different values of β on the DSD dataset respectively. In terms of PR curves, F-measure and MAE simultaneously, the performance reaches its peak as $\beta = 1.5$. When β is smaller than 1.5, the performance improves steadily with β. When β is larger than 1.5, the performance reduces speedily as β. Therefore, we set $\beta = 1.5$ in all experiments.

γ is the scaling factor to control the range of background distribution measure. The influence of γ on background distribution measure is the same as that of α on spatial distance between regions. Figs 11, 12 and 13 show the PR curves, F-measure and MAE for different

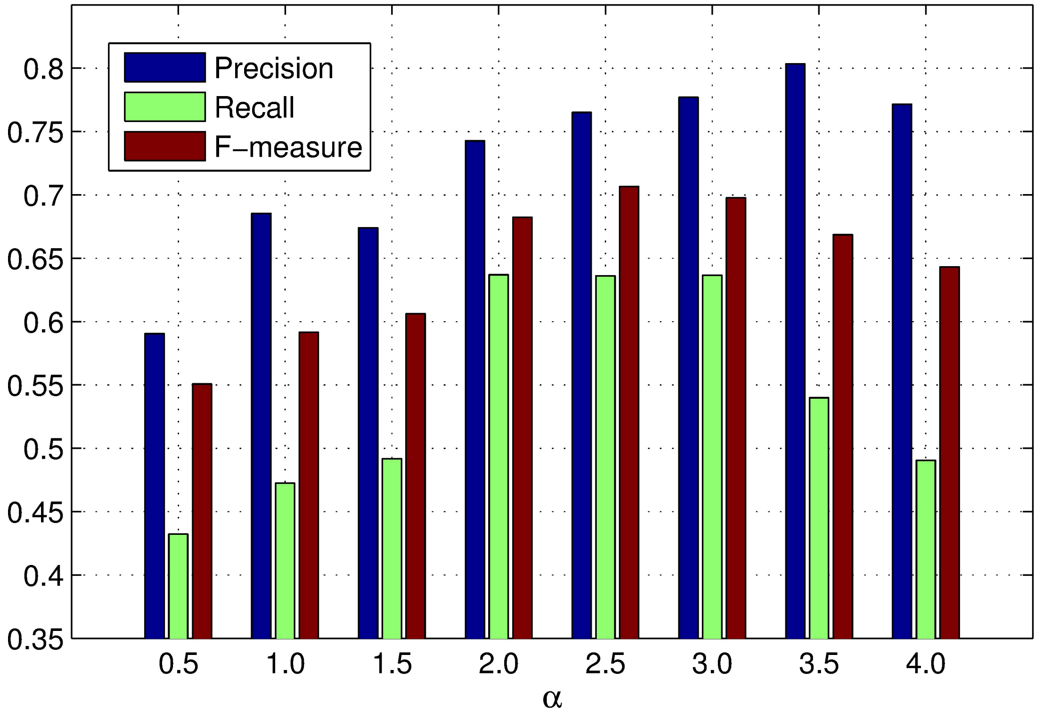

Fig 6. F-measure for different α on the DSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g006

values of γ on the DSD dataset respectively. In terms of PR curves, F-measure and MAE simultaneously, the performance reaches its peak when $\gamma = 2.5$. When γ is smaller than 2.5, the performance increases steadily with γ. When γ is larger than 2.5, the performance reduces gradually with γ. Therefore, we set $\gamma = 2.5$ in all experiments.

In addition, δ_{BC} and δ_{DC} control the strength of boundary connectivity $BC(\bullet)$ and distance between the image center and a region $DC(\bullet)$. This two parameters jointly constitute a Gaussian weight $\omega_{BD}(\bullet)$ to impact background distribution measure. To obtain reasonable value of $\omega_{BD}(\bullet)$, δ_{BC} and δ_{DC} are adaptively set as the maximum of $BC(\bullet)$ and $DC(\bullet)$ for every image. Let us consider two extreme situations: When a region is far away from image center and connected to image boundaries, $BC(\bullet)$ and $DC(\bullet)$ of the region both approach the corresponding maximums, which will lead to approximately the largest Gaussian weight of background distribution measure in the region($\omega_{BD} \approx 1 - e^{-(1+1)} \approx 0.865$). Consequently, background distribution $BD(\bullet)$ of the region tends to obtain a bigger value, which means that the region may belong to background areas. When a region is close to image center and separated from image boundaries, $BC(\bullet)$ and $DC(\bullet)$ of the region are both close to 0, which will lead to approximately the smallest Gaussian weight of background distribution measure in the region ($\omega_{BD} \approx 1 - e^{-(0+0)} \approx 0$). Accordingly, background distribution $BD(\bullet)$ of the region also approaches 0, which means that the region tends to be a salient region.

Results on datasets

Herein, we first give a visual comparison with eleven different methods (IT [37], GBVS [38], AC [39], FT [25], CA [40], SF [26], RBD [22], RC [20], MB [41], MST [42] and GP [43]) on two dataset(DSD and ECSSD). Fig 14 shows saliency maps and ground truths of five examples on DSD dataset. In the first case, there are several objects on the burlywood table that is placed

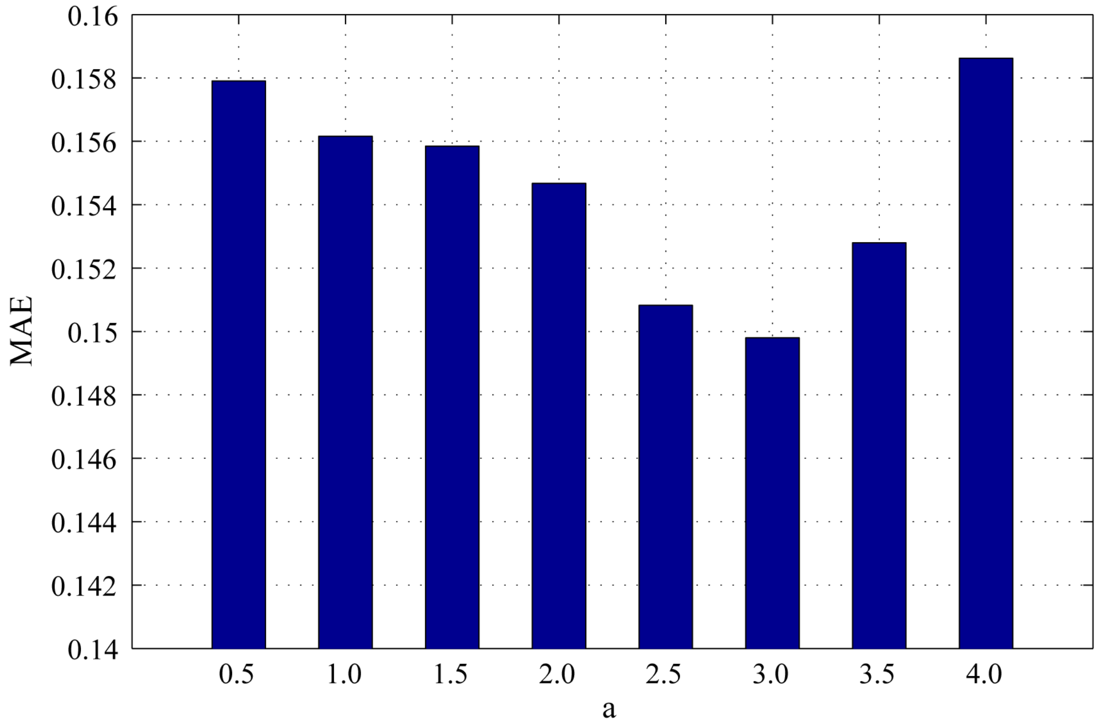

Fig 7. MAE for different *α* on the DSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g007

on the dark floor. We can catch these foreground objects effortlessly and rapidly, but it is a relatively hard task for artificial computational models. For instance, the dark floor belonging to the background area is allocated with the highest saliency value in four models of GBVS, FT, RC, SF and GP, which indicates that the floor is regarded as a salient region. Besides, the

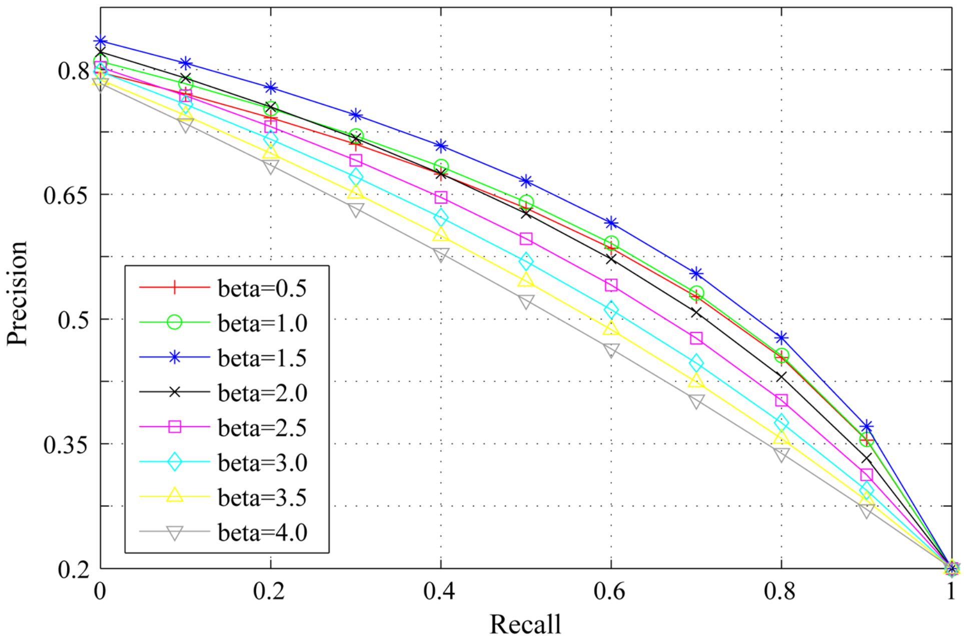

Fig 8. PR curves for different *β* on the DSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g008

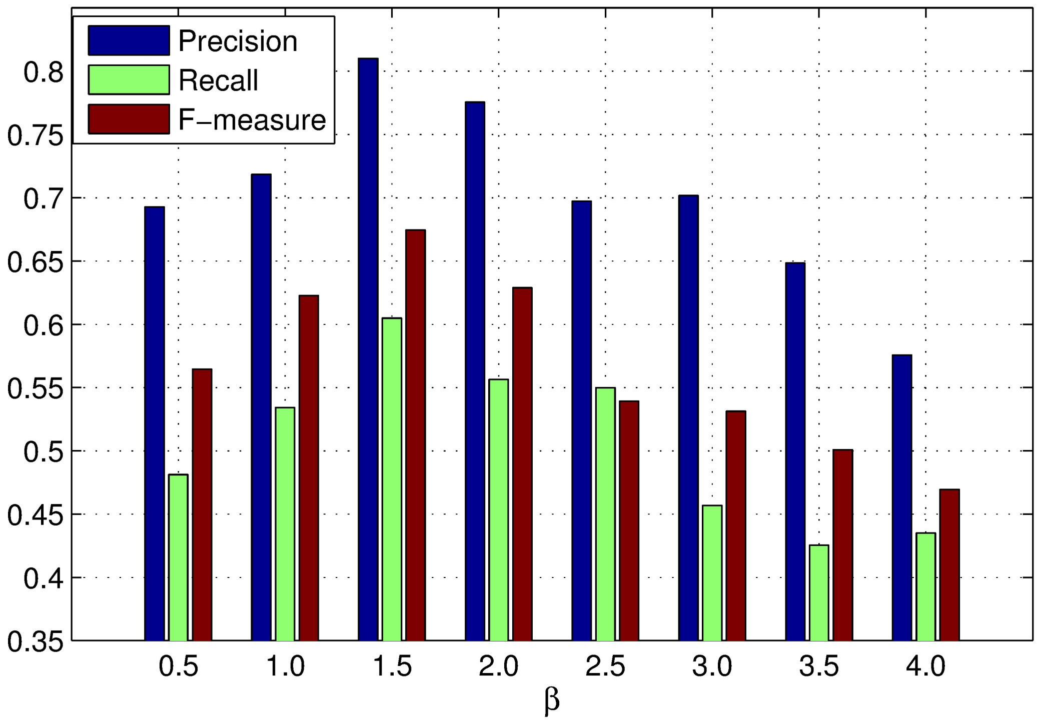

Fig 9. F-measure for different *β* on the DSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g009

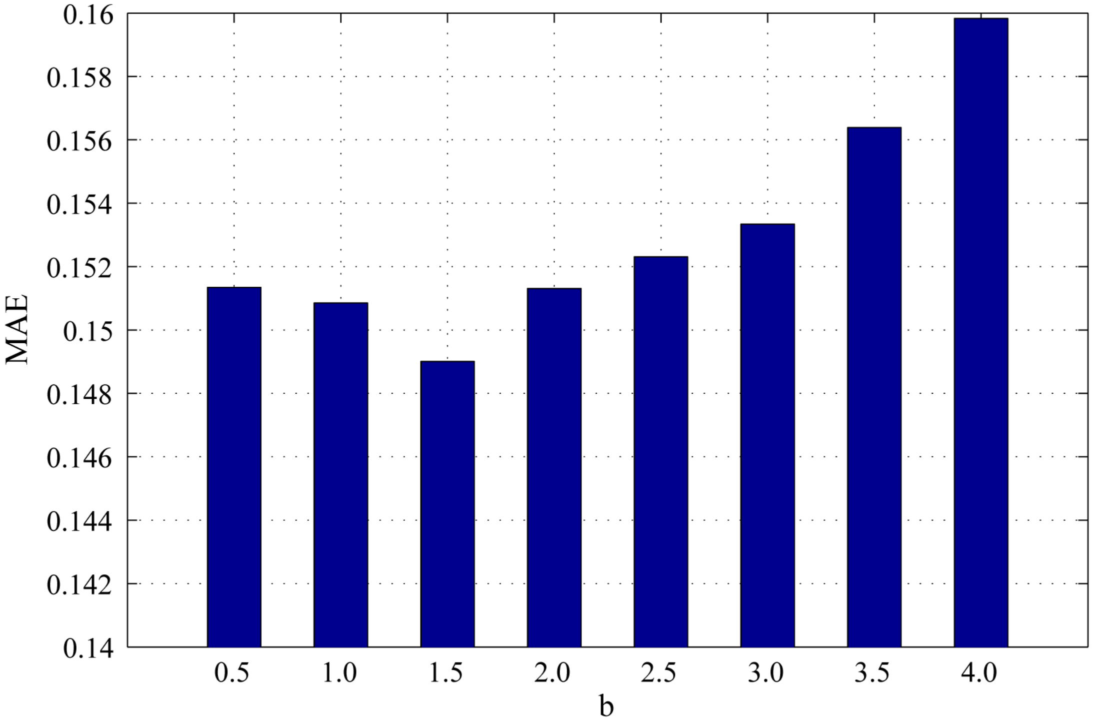

Fig 10. MAE for different *β* on the DSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g010

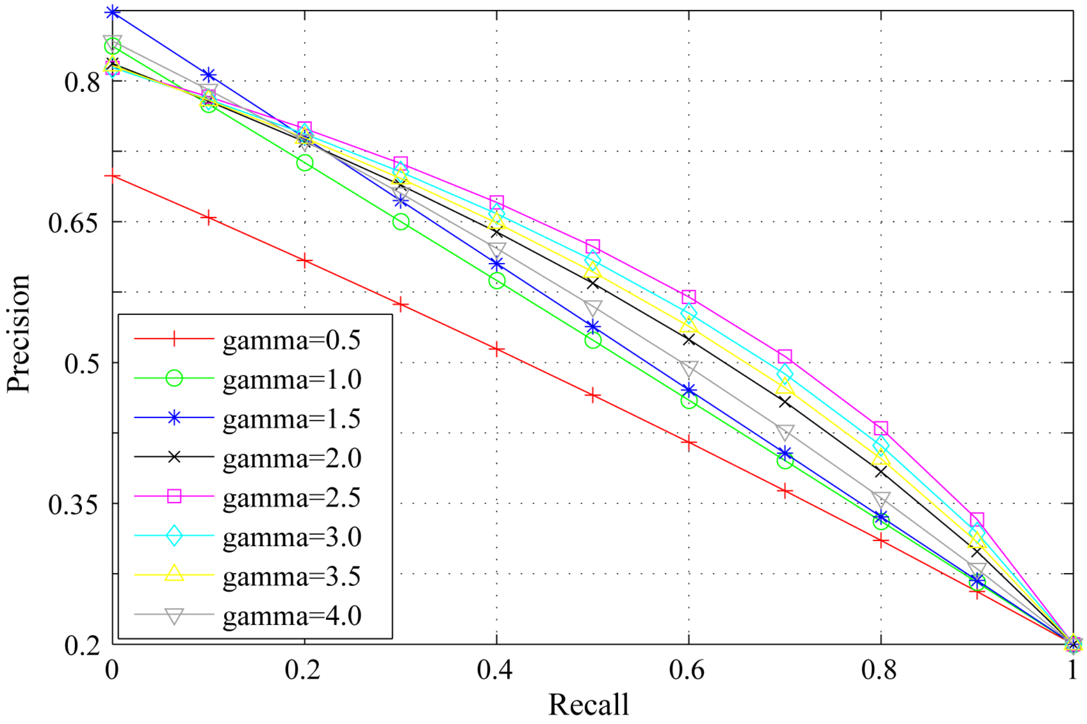

Fig 11. PR curves for different γ on the DSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g011

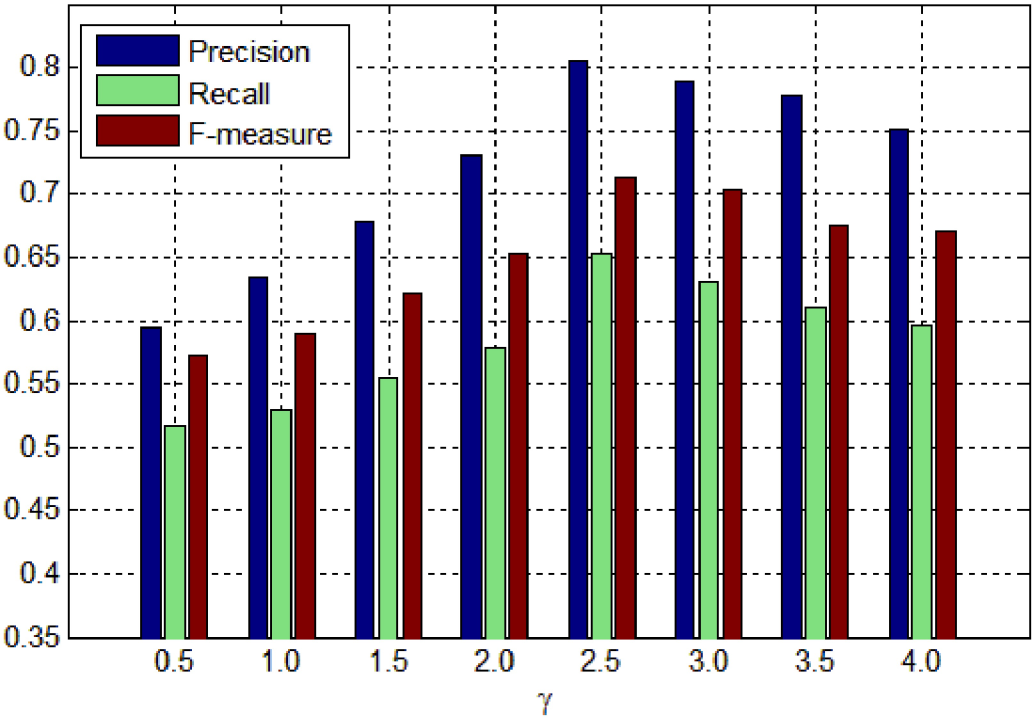

Fig 12. F-measure for different γ on the DSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g012

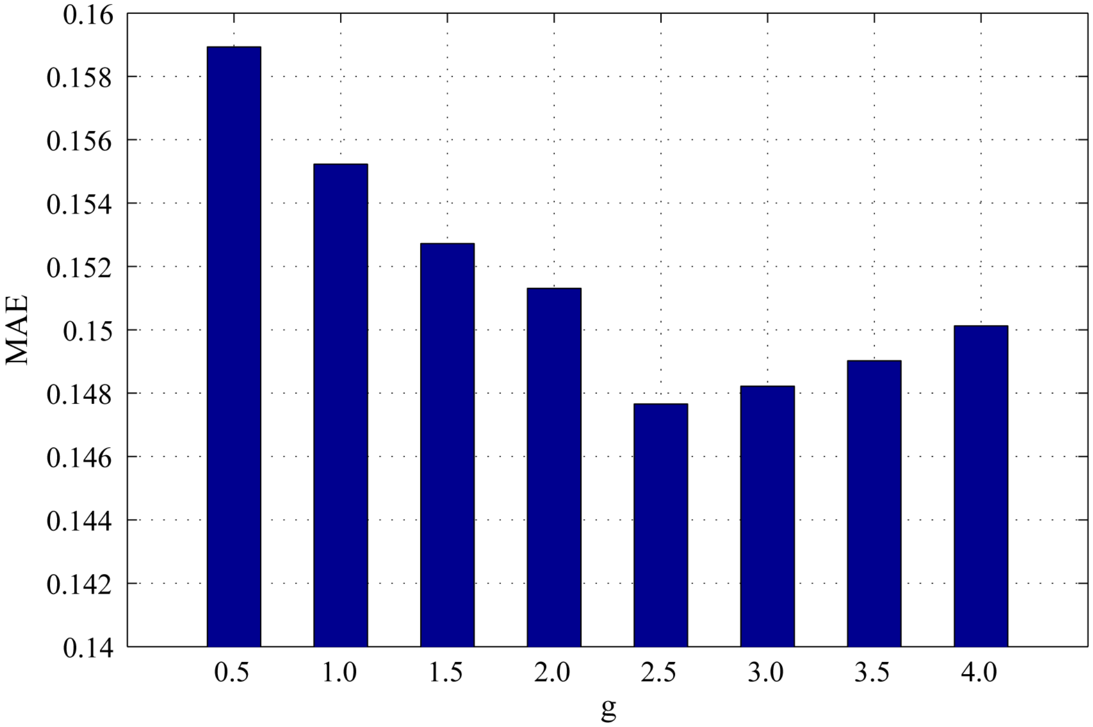

Fig 13. MAE for different γ on the DSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g013

middle carton and the right cup against the table of a similar color are difficult to be detected that can be found from all eleven models. Taking advantage of not only color but also depth, spatial layout, boundary connectivity and region roundness, the middle carton is assigned with a medium saliency value in the proposed model. In particular, that the glass bottle with red cap is divided into a whole region profits from the graph-based RGB-D segmentation. MB and MST perform well in the first case, but they sometimes mistake background areas for salient regions, such as the second and forth cases. In summary, our algorithm visually outperforms others in regard to ground truth.

The PR curves, F-measure and MAE are obtained to quantitatively evaluate the performance of our method in regard to others. Fig 15 shows PR curves of all the competing approaches on the DSD dataset and it clearly demonstrates that our method performs favorably against other eight. Particularly, the intersecting point of precision 0.2 and recall 1.0, where all pixels are retained as positives, means salient regions occupy 20% of an entire image on average [20]. From Fig 16, it is observed that the proposed approach achieves the best F-measure performance. Consistently, the MAE result of our model has the smallest value on the DSD dataset as shown in Fig 17 which also supports the excellence of our method. Figs 18, 19 and 20 shows PR curves, F-measure and MAE of ten different approaches on the ECSSD dataset, respectively. Note that we don't consider GP algorithm in the experiments on the ECSSD dataset, because GP algorithm requires depth cues such as depth images and point clouds. From the quantitative comparison with ten models, the proposed model get moderately good performance colosd to MB and MST. As shown in Fig 18, four comparable approaches including RBD, MB, MST and our model outperform the other seven. In Figs 19 and 20, it can be seen that RBD performs poorly compared with MB, MST and our model, and our model underperforms slightly MB and MST. In the ECSSD experiment process, we discover that

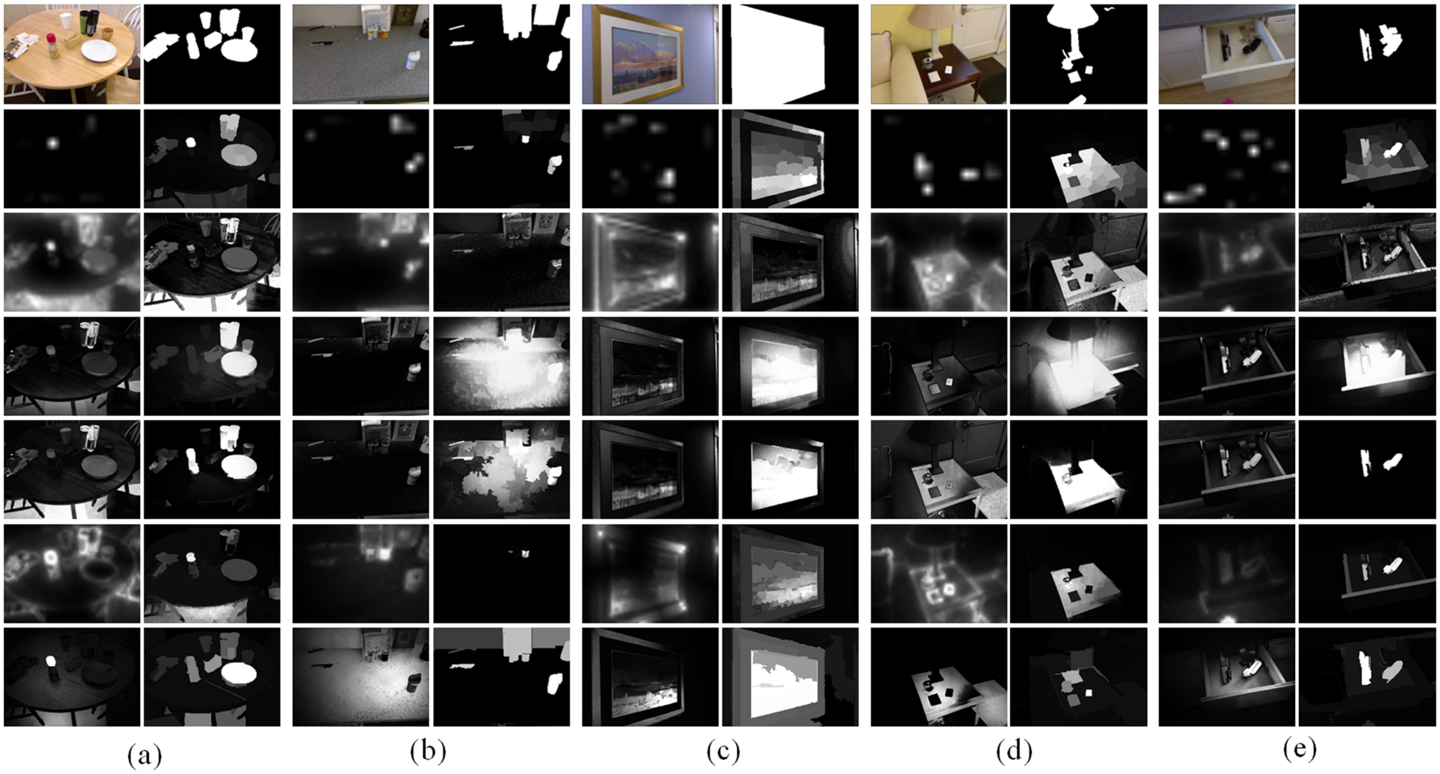

Fig 14. Visual comparison of saliency maps by eleven state-of-the-art algorithms on five different scenes. First column: input RGB image [33], saliency maps by IT [37], GBVS [38], AC [39], FT [25], CA [40], SF [26]; Second column: ground truth [33], RBD [22], RC [20], MB [41], MST [42], GP [43] and ours.

https://doi.org/10.1371/journal.pone.0180519.g014

sometimes our model may not hold the intergrality of foreground objects when they occupy a large area in terms of an image. In summary, our model is suitable for indoor scene with complex background, multi salient objects, while MB and MST algorithm are applicable to natural images with simple background and single salient object.

Application on the robot platform

The robot platform

To explore the indoor robot application of our method, a mobile robot platform is built that is equipped with a PartolBot robot, a Bumblubee-2 stereo camera that can capture images at the maximum resolution of 1024*768 pixels, a miniature pan-tilt unit PTU-46-17.5 that can provide accurate real-time positioning of a camera, and an Anmite touch screen that is used as human-machine interface(HMI) providing users with visual inputs and corresponding perception results intuitively. Note that depth cues is only used to assist the pre-segmentation, in spite of the inconsistency of depth generation between the Kinect in the DSD dataset and the stereo camera here. The PatrolBot robot is a programmable autonomous general purpose service robot rover built by MobileRobots Inc, which embeds an onboard computer and also can be connected with an offboard computer by wireless network. For these experiments, we choose the former approach to run the program of our algorithm. A structure made of aluminum alloy is mounted on the robot in order to give it more height, providing the point of view

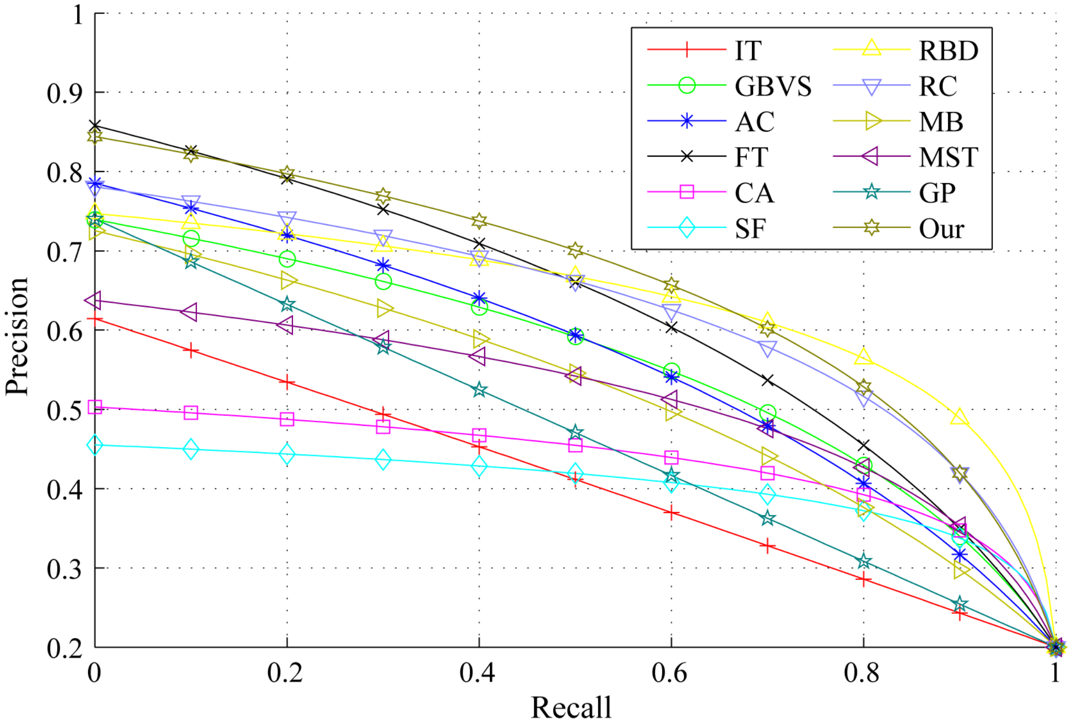

Fig 15. PR curves for different salient detection methods on the DSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g015

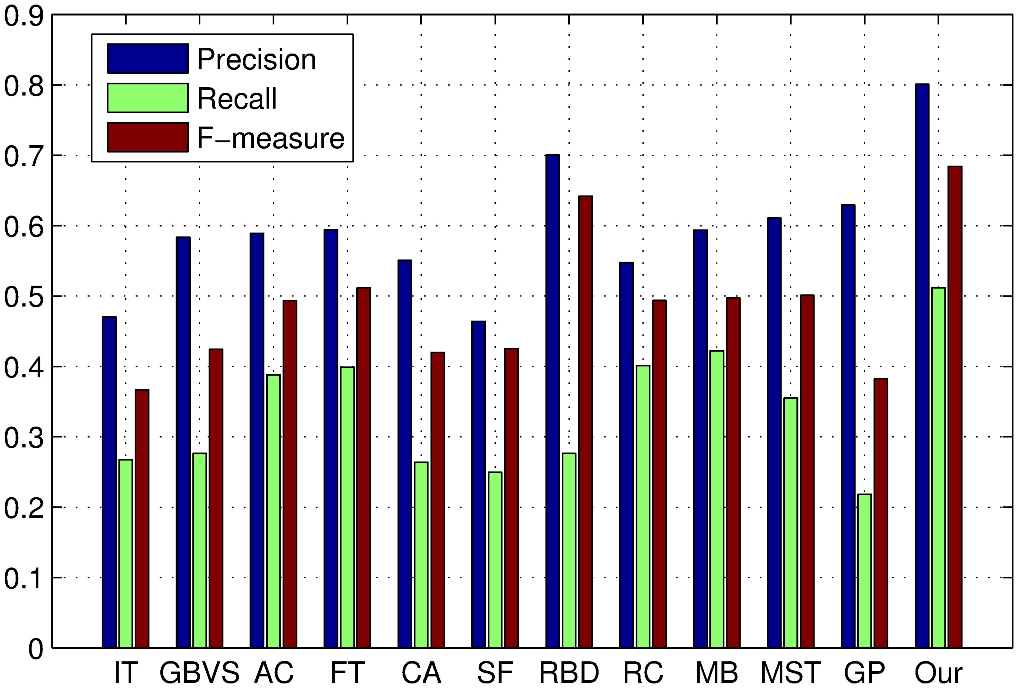

Fig 16. F-measure for different salient detection methods on the DSD dataset. Precision, Recall and F-measure using an adaptive threshold.

https://doi.org/10.1371/journal.pone.0180519.g016

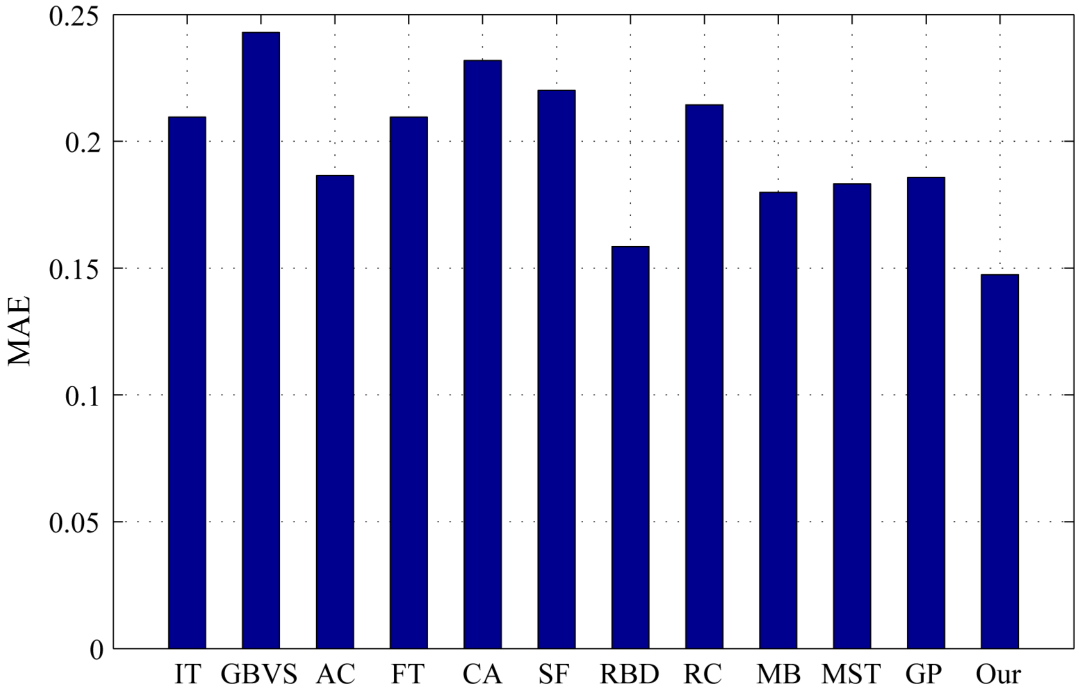

Fig 17. MAE for different salient detection methods on the DSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g017

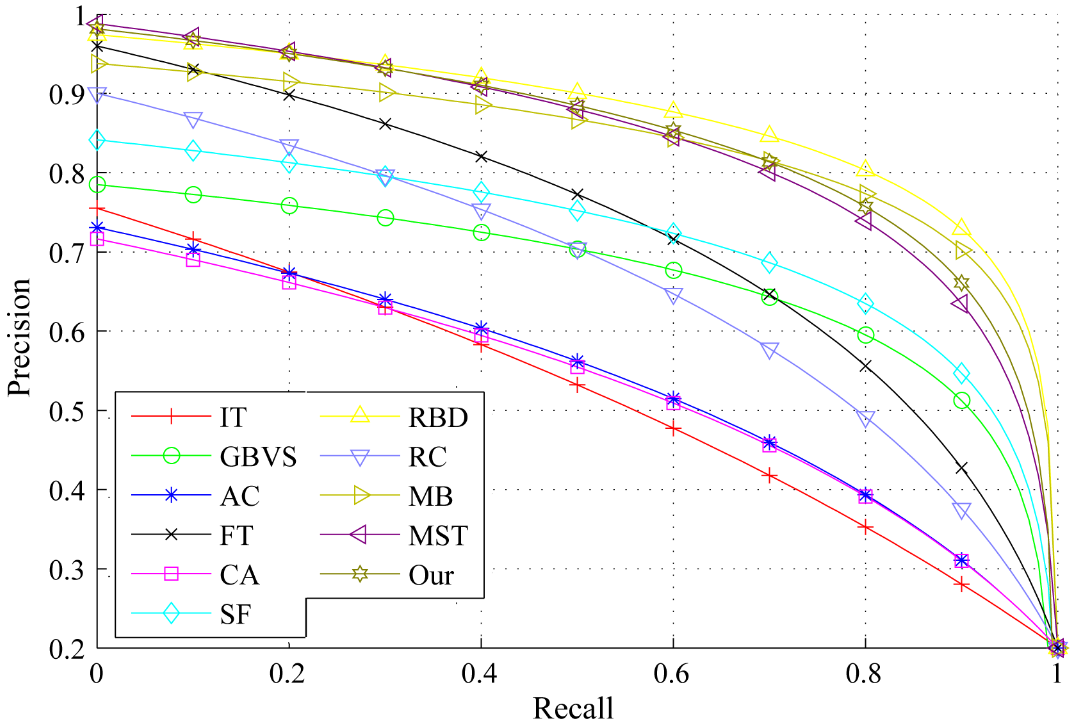

Fig 18. PR curves for different salient detection methods on the ECSSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g018

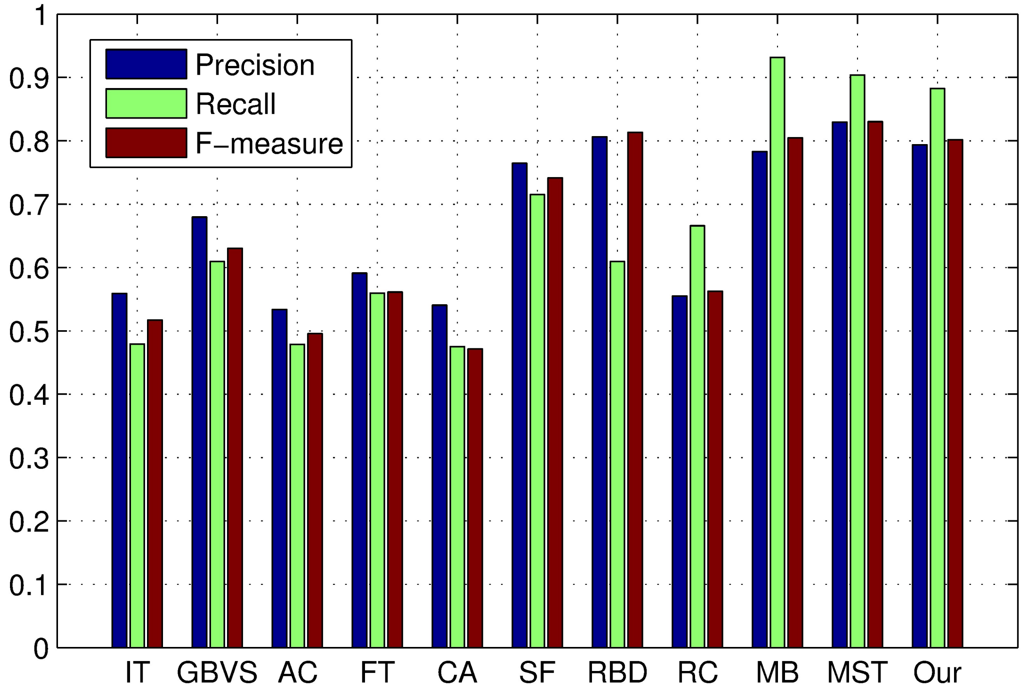

Fig 19. F-measure for different salient detection methods on the ECSSD dataset. Precision, Recall and F-measure using an adaptive threshold.

https://doi.org/10.1371/journal.pone.0180519.g019

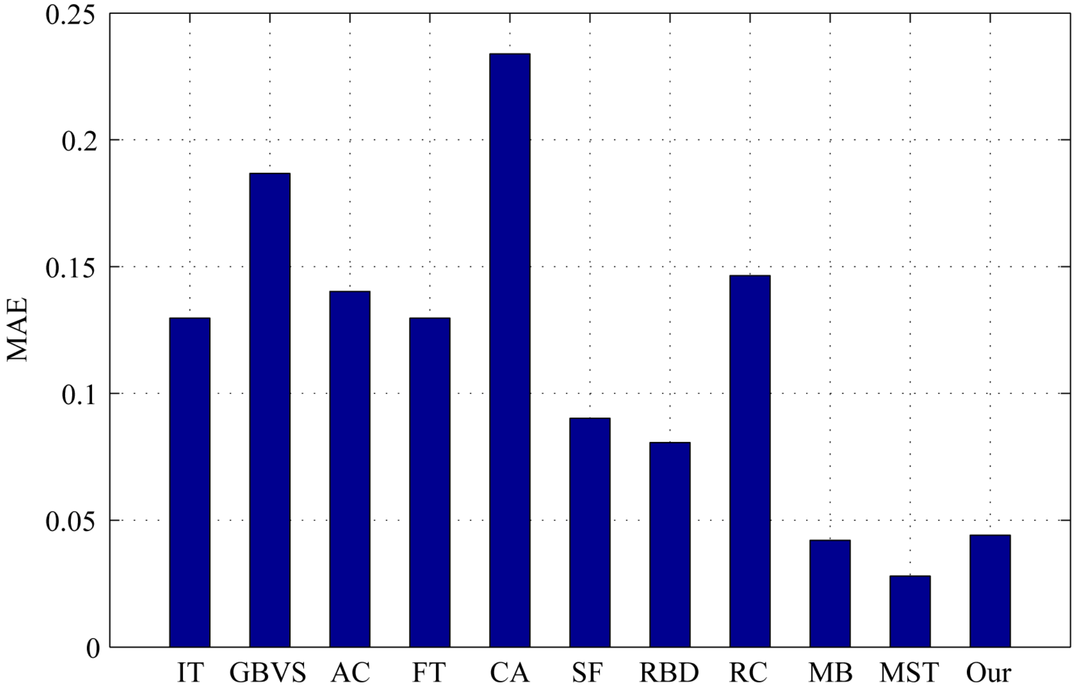

Fig 20. MAE for different salient detection methods on the ECSSD dataset.

https://doi.org/10.1371/journal.pone.0180519.g020

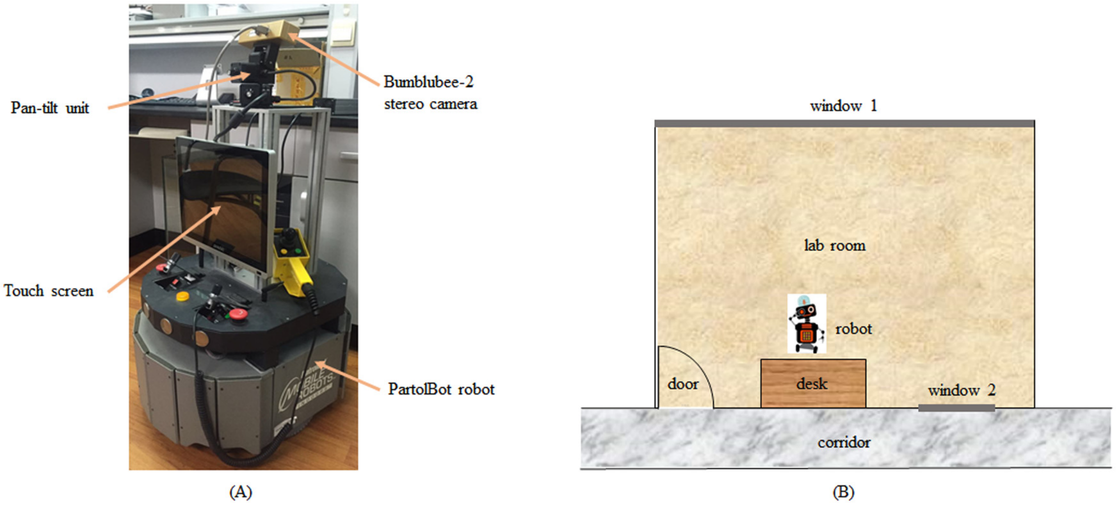

Fig 21. The robot platform and layout of the lab room. A: the robot platform. B: layout of the lab room.

https://doi.org/10.1371/journal.pone.0180519.g021

of a child with a height of 1.2m. The software is implemented in C++ using Visual Studio 2008. The robot platform can be seen in Fig 21A.

Results on the robot platform

For application in an actual environment and performance evaluation of our proposed algorithm, a lot of experiments are conducted on a real robot platform mentioned above.The situation of different viewpoints is challenging for object detection in that appearances of object may differ significantly when the viewpoint changes. Another important property of the visual system is the robustness under illumination variations, and this is a problem that frequently induces difficulties in robotic applications since images often look very different if illumination changes. Moreover, partial occlusions between objects are common cases in daily life which is an essential problem for object detection to be solved. Therefore, we design three different kinds of experimental constructions that are different viewpoints, illumination variations and partial occlusions, referring to [44].

The experiments are conducted in our lab room whose layout is depicted in Fig 21B. The robot is placed in front of a desk on which we arrange several objects such as book, mouse, cup, and so on. These objects' positions will be changed during the partial occlusion test. There are two windows as shown in Fig 21B: window 1 is a large French window with curtains, while window 2 is a high and small window without curtains along the corridor. So daylight entering from window 2 is rather fewer. Because the room faces the north and lighting is not good, so fluorescent lamps are on throughout the whole phase of robotic application. In consequence, illumination conditions will be controlled by means of opening and closing curtains of window 1. An instance of salient region detection is shown in Fig 22. In this case, the raw disparity map in Fig 22B generated by the Triclops Stereo Vision Software Development Kit (SDK) of Bumblubee-2 is not fully populated, and the missing data is labeled with aqua color. Curtains of window 1 are opening which means the current illumination contains artificial light and sunlight. Since the illumination is not equally distributed over the scene and shadows are present, background areas are segmented into some regions as shown in Fig 22D. Fortunately, five objects are extracted from the image successfully and allocated with high

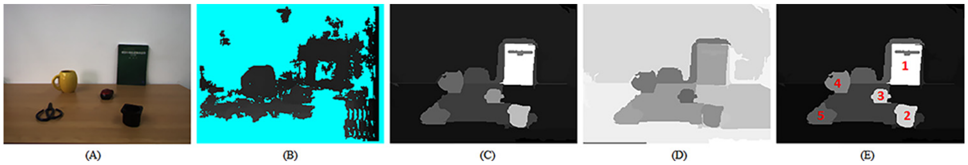

Fig 22. An instance running on the robot platform. A: left camera image captured by the Triclops stereo vision SDK of the Bumblubee-2 stereo camera. B: raw disparity map, computed automatically by the Bumblubee-2, in which the aqua color parts show the areas where the disparities are not available. C: primary saliency map. D: background distribution map. E: final saliency map with manual annotation for the attention shift sequence.

https://doi.org/10.1371/journal.pone.0180519.g022

saliency values on account of the excellent primary saliency measure. The attention shift sequence, salient regions in descending order, is 1–2–3–4–5 that is labeled by red number in Fig 22E.

The first experiment is a robustness test in regard to different viewpoints, which is implemented by moving wheels and the pan-tilt unit of our robot platform slightly. Herein two separate groups of viewpoint change experiment are displayed in Fig 23(a) and 23(b), whose difference is the illumination conditions that will be discussed in the next experiment. For the Fig 23(a), the left notebook is detected with the first focus in three of four cases. In the third image, the first focus diverts to the yellow cup. This happens probably because centroid of the cup is closer to the image center and its region roundness gives a more weight to background distribution measure. For the Fig 23(b), the left notebook is detected with the first focus in all of four cases. As we can see from these figures, the several objects are detected with high saliency values generally, and the sequence of attention shift oscillates as the viewpoint changes. Therefore, the algorithm has a good detection performance despite slight different viewpoints.

The second experiment is a robustness test in regard to illumination variations, also depicted in Fig 23(a) and 23(b). Two different situations are set up by closing(Fig 23(a)) and opening(Fig 23(a) and 23(b)) curtains of window 1, corresponding to illumination condition 1 and 2 respectively. Note that daylight is unavoidable for window 2 has no curtains, thus illumination condition 1 denotes some artificial light and few daylight while illumination condition 2 denotes the same artificial light and lots of daylight. Segmentation results of illumination condition 2 are inferior to that of illumination condition 1, such as the last instance in Fig 23 (a) and 23(b), in that the situation introduces more light interference. Nevertheless, five objects are detected with relatively high saliency values and the first focus attaches to the left notebook generally.

The third experiment is a robustness test in regard to partial occlusion. As shown in Fig 23 (a) and 23(c), we give four instances occluding different objects. For the first three instances, the curtains of window 1 are all opening and input images are captured from the same viewpoint. We can see from the first three columns of Fig 23(a) and 23(c) that the saliency of occluded objects is influenced faintly. The forth instance is designed for comprehensive evaluation of the reliability of our model in robotic applications, whose three experimental variables are all different from those three cases. And its first focus is stable on the left notebook although it is occluded by the yellow cup. Overall, the algorithm can allocate high saliency values for dominant objects regardless of occlusion. However, background regions are segmented into splintery ones so that saliency maps are to some extent cluttered in the background regions.

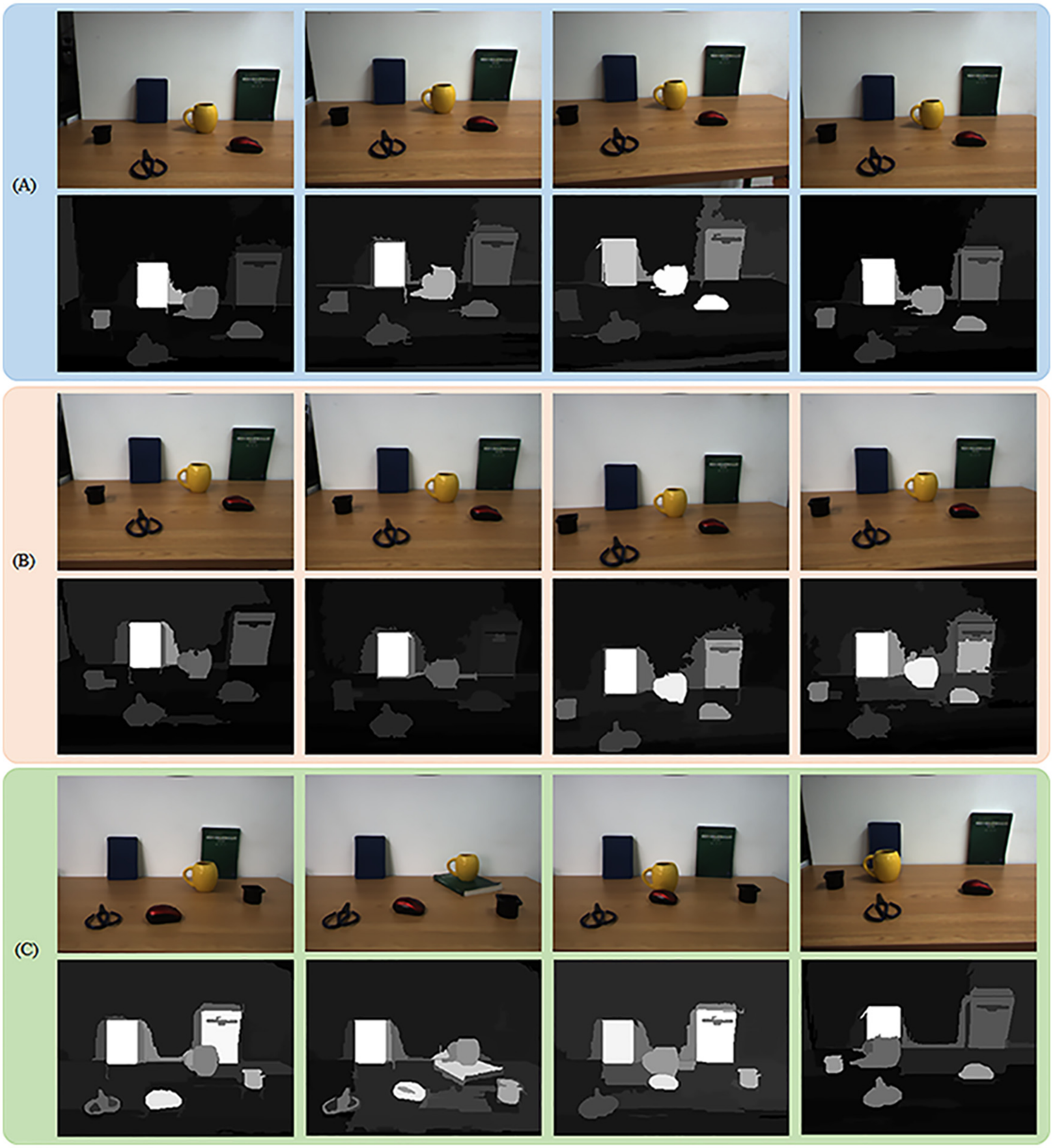

Fig 23. Experiment results of three experimental constructions that are different viewpoints, illumination variations and partial occlusions. The first row of (a): input RGB images from different viewpoints under illumination condition 1. The second row of (a): final saliency maps corresponding to the first row of (a). The first row of (b): input RGB images from different viewpoints under illumination condition 2. The second row of (b): final saliency maps corresponding to the first row of (b). The first row of (c): input RGB images in the first three cases are captured from the same viewpoint under illumination condition 1 while that of the last case is captured from another viewpoint under illumination condition 2. The second row of (c): final saliency maps corresponding to the first row of (c). illumination condition 1: some artificial light and few daylight when curtains of window 1 is closed. illumination condition 2: some artificial light and lots of daylight when curtains of window 1 is open.

https://doi.org/10.1371/journal.pone.0180519.g023

Conclusions

This paper presents an effective method for indoor robot application to detect salient regions or objects in complex environment. Firstly, to keep the completeness of salient object candidates, a given image is divided into regions by the graph-based RGB-D segmentation using color and depth feature. Secondly, based on the segmentation results, the primary saliency map is measured by utilizing color(distance in RGB color space and color delegation of a region), region area and spatial layout of regions, while the background distribution map is calculated by region roundness, boundary connectivity and spatial layout between image center and region. During this stage, the color delegation of a region is enumerated by building a color histogram and picking more frequently occurring colors. Besides, region roundness is proposed to describe the compactness of a region to produce robust background distribution measure. Thirdly, the final saliency map is calculated by combining the above two maps in an exponential function way, where the regions with the relatively high saliency values are considered as salient in our model. To validate the proposed approach, two kinds of experiments are conducted. The first experiment is carried out by the comparison with eleven significant models on the public DSD and ECSSD dataset, whose results show that our approach outperforms these existing saliency detection approaches in indoor secnes. The second experiment on the self-made robot platform shows that the algorithm is robust to different viewpoints, illumination variations and partial occlusion.

The contributions and advantages of the work are summed up as follows. Firstly, in order to keep the completeness and compactness of salient region candidates, depth cues are utilized during the graph-based RGB-D segmentation stage, which is important for robots to perceive the location and size of desired objects. Besides, assuming that salient regions possess the attributes of compact sizes while background ones tend to distribute widely and near image boundaries, we put forward the concept of region roundness, the representation of how compact a region is. Background distribution measure is more robust when region roundness is applied. Moreover, a principled framework which combines the primary saliency and background distribution is built, and it is applied on the indoor robot platform.

Honstly, there are still some insufficiencies in our model. For example, the graph-based RGB-D segmentation is to some extent susceptible to illumination variations, particularly to shadows, which can be seen from Fig 23. Specifically, segmented salient region candidates are prone to carry rough edges, which weakens values of region roundness, same with the final saliency values. This work is concentrated on the bottom-up, stimulus-driven and involuntary stage of attention, but it is not enough for the robot application. As the further work, we would like to introduce top-down, goal-driven and voluntary stage of attention into the method so that our indoor robot has the ability of "active vision".

Acknowledgments

The authors acknowledge the support of the National Natural Science Foundation of China (NSFC, Grant No. 61305118 and U1509202) and the National Natural Science Foundation of Jiangsu (JSNSF, Grant No. BK20130329).

Author Contributions

Conceptualization: NL HX ZHW.

Data curation: NL HX.

Formal analysis: NL HX ZHW LNS GDC.

Funding acquisition: GDC.

Investigation: NL HX ZHW.

Methodology: NL HX ZHW LNS GDC.

Project administration: LNS GDC.

Software: NL HX ZHW.

Supervision: LNS GDC.

Validation: NL HX ZHW LNS GDC.

Visualization: NL HX ZHW.

Writing – original draft: NL HX.

Writing – review & editing: NL HX ZHW LNS GDC.

References

1. Domhof J, Chandarr A, Rudinac M, Jonker P. Multimodal joint visual attention model for natural human-robot interaction in domestic environments. IROS 2015: the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2015 Sept 28-Oct 02; Hamburg, Germany. p.2406–2412.
2. Begum M, Karray F. Visual Attention for Robotic Cognition: A Survey. IEEE Trans Auton Ment Dev. 2011; 3:92–105. https://doi.org/10.1109/TAMD.2010.2096505
3. Lukic L, Billard A, Santos-Victor J. Motor-Primed Visual Attention for Humanoid Robots. IEEE Trans Auton Ment Dev. 2015; 7:76–91. https://doi.org/10.1109/TAMD.2015.2417353
4. Rutishauser U, Walther D, Koch C, Perona P. Is bottom-up attention useful for object recognition? CVPR 2004: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2004 June 27-July 2; Washington, DC, USA. 2: p.II–37–II–44.
5. Walther D, Rutishauser U, Koch C, Perona P. Selective visual attention enables learning and recognition of multiple objects in cluttered scenes. Comput Vis Image Underst. 2005; 100(1–2):41–63. https://doi.org/10.1016/j.cviu.2004.09.004
6. Ren Z, Gao S, Chia LT, Tsang IWH. Region-Based Saliency Detection and Its Application in Object Recognition. IEEE Trans Circuits Syst Video Technol. 2014; 24(5):769–779. https://doi.org/10.1109/TCSVT.2013.2280096
7. Donoser M, Urschler M, Hirzer M, Bischof H. Saliency driven total variation segmentation. ICCV 2009: the 2009 IEEE International Conference on Computer Vision; 2009 June 20-25; Miami, Florida, USA. p.817–824.
8. Li H, Wu W, Wu E. Robust Salient Object Detection and Segmentation. ICIG 2015: the 8th International Conference on Image and Graphics; 2015 Aug 13-16; Tianjin, China. p.271–284.
9. Feng S, Xu D, Yang X. Attention-driven salient edge(s) and region(s) extraction with application to {CBIR}. Signal Processing. 2010; 90(1):1–15. https://doi.org/10.1016/j.sigpro.2009.05.017
10. Simone F, Markus K. Most salient region tracking. ICRA 2009: the 2009 IEEE International Conference on Robotics and Automation; 2009 May 12-17; Kobe, Japan. p.1869–1874.
11. Zhang G, Yuan Z, Zheng N, Sheng X, Liu T. In: Zha H, Taniguchi Ri, Maybank S, editors. Visual Saliency Based Object Tracking. Computer Vision– ACCV 2009: the 9th Asian Conference on Computer Vision; 2009 Sept 23-27; Xi'an, China. Berlin: Springer; 2010.p.193–203.
12. Borji A, Frintrop S, Sihite DN, Itti L. Adaptive object tracking by learning background context. CVPR 2012: the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops; 2012 June 16-21; Providence, Rhode Island, USA. p.23–30.
13. Madokoro H, Ishioka Y, Takahashi S, Sato K, Shimoi N. Visual Saliency Based Multiple Objects Segmentation and its Parallel Implementation for Real-Time Vision Processing. Computer Science and Information Technology. 2015; 5(3):188–198.
14. Butko NJ, Zhang L, Cottrell GW, Movellan JR. Visual saliency model for robot cameras. ICRA 2008: the 2008 IEEE International Conference on Robotics and Automation; 2008 May 19-23; Pasadena, California, USA. p.2398–2403.
15. Wang X. Active Vision for Humanoid Robots. M.Sc. Thesis, Northwestern Polytechnical University. 2015.
16. Cynthia B, Aaron E, Paul F, Brian S. Active vision for sociable robots. IEEE Trans Syst Man Cybern—Part A: Systems and Humans. 2001; 31(5):443–453. https://doi.org/10.1109/3468.952718
17. Saleiro M, Farrajota M, Terzi x0107; K, Krishna S, Rodrigues JMF, Buf JMH. In: Antona M, Stephanidis C, editors. Biologically Inspired Vision for Human-Robot Interaction. UAHCI 2015: the 9th International

Conference Universal Access in Human-Computer Interaction; 2015 Aug 2-7; Los Angeles, CA, USA. Cham: Springer International Publishing;2015. p.505–517.

18. Shirai K, Madokoro H, Takahashi S, Sato K. Parallel implementation of saliency maps for real-time robot vision. ICCAS 2014: the 14th International Conference on Control, Automation and Systems; 2014 Oct 22-25; Seoul, Korea. p.1046–1051.

19. Meger D, Forssén PE, Lai K, Helmer S, McCann S, Southey T, et al. Curious George: An attentive semantic robot. Rob Auton Syst. 2008; 56(6):503–511. https://doi.org/10.1016/j.robot.2008.03.008

20. Cheng MM, Zhang GX, Mitra NJ, Huang X, Hu SM. Global contrast based salient region detection. CVPR 2011: the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2011 June 7-12; Colorado Springs, CO, USA. p.409–416.

21. Jiang L, Koch A, Zell A. Salient regions detection for indoor robots using RGB-D data. ICRA 2015: the 2015 IEEE International Conference on Robotics and Automation; 2015 May 26-30; Washington, DC, USA. p.1323–1328.

22. Zhu W, Liang S, Wei Y, Sun J. Saliency Optimization from Robust Background Detection. CVPR 2014: the 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2014 June 23-28; Columbus, OH, USA. p.2814–2821.

23. Zhai Y, Shah M. Visual attention detection in video sequences using spatiotemporal cues. Proceedings of the 14th ACM international conference on Multimedia; 2006 Oct 23-27; Santa Barbara, CA, USA. p.815-24.

24. Liu T, Sun J, Zheng NN, Tang X. Learning to Detect A Salient Object. CVPR 2007: the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2007 June 18-23; Minneapolis, Minnesota, USA. p.1–8.

25. Achanta R, Hemami S, Estrada F, Susstrunk S. Frequency-tuned salient region detection. CVPR 2009: the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2009 June 20-25; Miami, Florida, USA. p.1597–1604.

26. Perazzi F. Saliency filters: Contrast based filtering for salient region detection. CVPR 2012: the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2012 June 16-21; Providence, Rhode Island, USA. p.733–740.

27. Yang C, Zhang L, Lu H, Xiang R. Saliency Detection via Graph-Based Manifold Ranking. CVPR 2013: the 2013 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2013 June 23-28; Portland, Oregon, USA. p.3166–3173.

28. Yan Q, Xu L, Shi J, Jia J. Hierarchical Saliency Detection. CVPR 2013: the 2013 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2013 June 23-28; Portland, Oregon, USA. p.1155–1162.

29. Wang P, Zhou Z, Liu W, Qiao H. Salient region detection based on local and global saliency. ICRA 2014: the 2014 IEEE International Conference on Robotics and Automation; 2014 May 31-June 7; Hong Kong, China. p.1546–1551.

30. Felzenszwalb PF, Huttenlocher DP. Efficient Graph-Based Image Segmentation. Int J Comput Vis. 2004; 59(2):167–181. https://doi.org/10.1023/B:VISI.0000022288.19776.77

31. Ferreira JF, Dias J. Attentional Mechanisms for Socially Interactive Robots—A Survey. IEEE Trans Auton Ment Dev. 2014; 6(2):110–125. https://doi.org/10.1109/TAMD.2014.2303072

32. B K, Schauerte B, Kroschel K, Stiefelhagen R. Multimodal saliency-based attention: A lazy robot's approach. IROS 2012: the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2012 Oct 7-12; Vilamoura, Algarve, Portugal. p.807–814.

33. Ciptadi A, Hermans T, Rehg J. An In Depth View of Saliency. BMVC 2013: the 24th British Machine Vision Conference; 2013 Sept 9-13; Bristol, UK. p.112.1–112.11.

34. Borji A, Sihite DN, Itti L. Salient Object Detection: A Benchmark. Computer Vision—ECCV 2012: the 12th European Conference on Computer Vision; 2012 Oct 7-13; Florence, Italy. Berlin: Springer; 2012. p.414–429.

35. Scharfenberger C, Waslander SL, Zelek JS, Clausi DA. Existence Detection of Objects in Images for Robot Vision Using Saliency Histogram Features. In: 2013 International Conference on Computer and Robot Vision; 2013. p. 75–82.

36. Zhou L, Yang Z, Yuan Q, Zhou Z, Hu D. Salient Region Detection via Integrating Diffusion-Based Compactness and Local Contrast. IEEE Transactions on Image Processing. 2015; 24(11):3308–3320. https://doi.org/10.1109/TIP.2015.2438546 PMID: 26080382

37. Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell. 1998; 20(11):1254–1259. https://doi.org/10.1109/34.730558

38. Jonathan H, Christof K, Pietro P. Graph-Based Visual Saliency. NIPS 2006: Proceedings of the 20th Conference on Neural Information Processing Systems; 2006 Dec 4-9; Vancouver, British Columbia, Canada. p.545–552.

39. Achanta R, Estrada F, Wils P, Süsstrunk S. In: Gasteratos A, Vincze M, Tsotsos JK, editors. Salient Region Detection and Segmentation. Berlin: Springer; 2008. p.66–75.

40. Goferman S, Zelnik-Manor L, Tal A. Context-Aware Saliency Detection. IEEE Trans Pattern Anal Mach Intell. 2012; 34(10):1915–1926. https://doi.org/10.1109/TPAMI.2011.272 PMID: 22201056

41. Zhang J, Sclaroff S, Lin Z, Shen X, Price B, MeÆch R. Minimum Barrier Salient Object Detection at 80 FPS. In: IEEE International Conference on Computer Vision(ICCV); 2015. p.1404–1412.

42. Tu WC, He S, Yang Q, Chien SY. Real-Time Salient Object Detection with a Minimum Spanning Tree. In: IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 2334–2342.

43. Ren J, Gong X, Yu L, Zhou W, Yang MY. Exploiting global priors for RGB-D saliency detection. In: Computer Vision and Pattern Recognition Workshops; 2015. p. 25–32.

44. Frintrop S. VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search. Springer Berlin Heidelberg; 2006.