

# Salient Regions Detection for Indoor Robots using RGB-D Data

Lixing Jiang, Artur Koch and Andreas Zell

**Abstract**—The goal of saliency detection is to highlight objects in image data that stand out relative to their surrounding. Therefore, saliency detection aims to capture regions that are perceived as important. The most recent bottom-up approaches for saliency detection measure contrast based on visual features in 2D scenes, ignoring depth value. This work presents an effective method to measure saliency by mapping pixels into foreground and background regions in RGB-D images. Namely, we first segment an image into regions to evaluate the object uniqueness and consistency using graph-based segmentation. Then, we utilize the region color, depth, layout and boundary information to produce robust foreground and background saliency measures. Finally, we combine the two saliency maps based on Gaussian weights. As a result, our approach produces high-quality saliency maps, which may be used for further processing like object detection or recognition. Experimental results on two datasets compare our method with the state of the art and highlight its effectiveness.

## I. INTRODUCTION

### A. Motivation

For an intelligent robot, as with a human, salient region detection plays a vital role in identifying and filtering information in unknown and complex environments. Visual saliency maps can competently guide the attention of an agent to potentially relevant candidates and locations in a scene, which is beneficial for many applications like object detection and recognition. Current methods estimate the visual saliency based on global or local contrasts of colors or textures in an image [1]–[3]. However, such methods have difficulties in handling variations in lighting and homogeneous color distributions between foreground and background. As a consequence, saliency detection, albeit considered practically useful, is still a technologically challenging problem in the field of computer vision.

With the advent of low-cost RGB-D sensors like the Microsoft Kinect, the suitability of RGB-D-based methods has become more universal. By utilizing additional depth information and derived features, the view of saliency becomes more precise and thus more feasible for practical applications [4], [5]. Depth data makes it possible to separate objects which are similar in appearance. Inspired by those advances, we incorporate depth values with visual features to estimate salient regions. A simple practical scenario is a mobile service robot as an object recognition system in a real-world indoor environment.

The major focus of this work, therefore, is the development of saliency detection measures to meet the requirements of

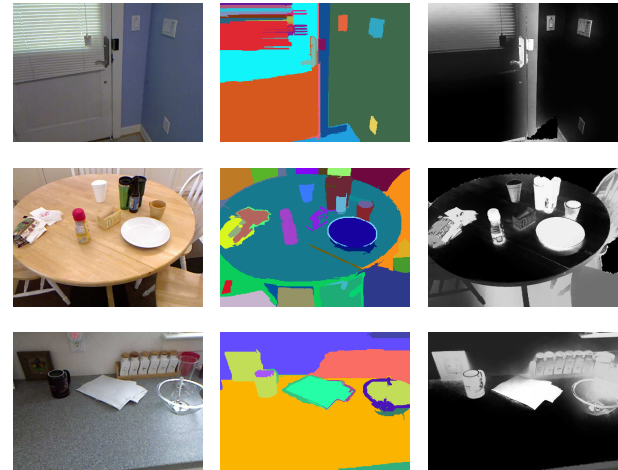


Fig. 1. Saliency map examples. RGB image samples (left column) from the public dataset in [4], graph-based RGB-D segmentation (middle column) and the resulting saliency map (right column).

indoor mobile robots. In addition to considering contrast of images, we combine depth cues with measures of distance, color, spatial layout and boundary connectivity to calculate a saliency map. Fig. 1(a) shows three scenes captured by a service robot in an indoor environment from [4]. Multiple foreground objects of potential interest can be seen in these scenes. The segmented candidate regions are visualized with different colors in Fig.1(b). Instead of attending to entire segmented regions, we expect that the robot can identify the most visually noticeable foreground objects through the saliency map (Fig.1(c)).

For this purpose, we first introduce a segmentation method which applies a graph-based algorithm for color and depth. The graph-based segmentation is designed to identify homogeneous regions based on color as well as depth cues. The algorithm clusters pixels in regard to similar properties but retains the uniqueness and consistency of different objects. This clustering in the segmentation stage largely decreases the computational complexity since widespread areas may usually be filtered as being visually unimportant due to low variance. After discussing different saliency methods in use today, we will propose a new saliency estimation approach that integrates color, depth, spatial layout, and boundary connectivity. Through the fusion of RGB and depth data, the proposed method provides good results despite the presence of homogeneous color distributions between foreground and background areas. To measure the performance of the approach, we evaluate it on two datasets against different state-of-the-art alternatives. The results show that combining

L. Jiang, A. Koch are with the Chair of Cognitive Systems, headed by Prof. A. Zell, Computer Science Department, University of Tuebingen, Sand 1, D-72076 Tuebingen, Germany {lixing.jiang, artur.koch, andreas.zell}@uni-tuebingen.de

visual color, depth value and spatial layout significantly improves the accuracy of object detection algorithms.

The contributions and advantages of the proposed method are as follows:

- The graph-based segmentation fuses RGB with depth data to measure inconsistencies in the image. The algorithm achieves superior detection of textured or inhomogeneous colored objects compared to our previous work, where we were assuming mostly uniformly colored fruits [6] or closest objects [7] and were thus able to successfully use the homogeneity and shortest distance in the segmentation.
- Given the assumption of rarity in regard to the scene representation for objects of interest, we employ an adapted and compact estimation of salient regions with essential RGB-D characteristics using color luminance, depth, spatial layout and boundary information. *Foregroundness* is estimated by difference of color, depth and position. *Backgroundness* is produced by area, boundary connectivity and the relative distance from the center of one region to the image corners instead of defining a narrow border region as an indicator for background [8].
- The final saliency map is acquired by combining foreground and background measures based on a *Gaussian* filter.

The remainder of the paper is structured as follows: a brief summary of relevant concepts in saliency algorithms is presented in Section I-B. Section II describes our proposed method, including the segmentation approach and saliency algorithm. Moreover, visual samples of saliency results are presented for the comparison of different saliency methods. Finally, in Section III we present the experimental results based on two datasets and conclude the paper in Section IV.

### B. Related Work

In recent years, the development of methods for object recognition and detection has been rapidly advancing. Many researchers have studied the effects of saliency detection [1]–[3], [8]–[10]. In general, saliency detection algorithms can be roughly classified into two categories: top-down and bottom-up. The top-down methods [11] obtain a saliency map by learning visual knowledge. In other words, top-down saliency methods require a large amount of annotated images for training. In contrast, the bottom-up approaches [1]–[3], [8], [9] focus on a low-level algorithm by determining contrast of image regions relative to their surrounding, intensity, color and orientation. These approaches do not require prior training. Itti *et al.* [12] were the first to advocate a bottom-up approach in visual attention. They utilized local contrast and visual low-level features to acquire saliency. Subsequently, Achanta *et al.* [13] acquired a saliency map by computing the difference between the image and a *Gaussian* blurred version of the original image. Though being simple and computationally efficient, the method failed when the saliency region occupied more than half the pixels of the image, or in the presence of complex backgrounds. Achanta presented a

revised approach based on the idea of maximum symmetric surround [14], which is derived from the assumption of a relation between scale and position of the candidate object in the image. Cheng *et al.* [2] proposed a global histogram-based contrast for saliency detection. The dissimilarity of a pair of patches is determined by comparing their color histograms. *Saliency filters* were presented by Perazzi *et al.* [3] relying on estimating an element uniqueness and distribution as a function of image contrast. Inspired by recent advances in contrast analysis, Zhu *et al.* [9] proposed a saliency optimization from background detection. They utilize a measure describing the connectivity between region and image boundaries.

While there is a wealth of research on visual saliency maps, few attempts have been made to combine depth values to form a saliency map. Maki *et al.* [15] presented a computational model for attention by using depth cues. In this depth-based model, closer targets were mapped to higher priority in an attentional scheme. Ouerhani *et al.* [16] proposed an alternative method which utilizes depth as an additional channel that is combined with color and other features. Lang *et al.* [5] collected a human eye fixation database in both 2D and 3D scenes by the Kinect sensor. They derive depth priors that may be applied to saliency maps aiming to predict visual attention areas of humans. Another method for incorporating visual saliency and depth information was proposed by Ciptadi *et al.* [4]. This method used 3D layout and shape features from depth measurements to generate a saliency map. They presented promising results by saliency-based segmentation using a superpixel Markov Random Field (*MRF*). Our work follows the paradigm of bottom-up approaches incorporating depth cues.

## II. METHOD

In this section, the segmentation algorithm and salient measures under RGB-D data are described. The main task of this work is to let the robot automatically detect salient objects in a scene. Hence, we focus on the need for capturing salient objects. The algorithm proposed in this section includes three steps to address and optimize this problem. At first, we use a graph-based RGB-D segmentation to cluster pixels in an image. This process minimizes the search space and integrates common colors, textures and depth in a region. Then, we present a method which combines salient foreground and background regions to model the corresponding saliency map. In a final step, we compute the desired saliency map by a weighted combination of saliency sub-maps.

### A. Graph-Based RGB-D Segmentation

The goal of the segmentation process is to select possible salient region candidates from an intricate environment represented by the RGB-D data stream. In this paper, we apply the graph-based approach from [17] to label different elements in an RGB-D image. First, we treat an RGB-D image as a fully-connected directed graph  $G = (V, E)$  with vertices  $v_i \in V$  and a set of edges  $(v_i, v_j) \in E$ . Each edge

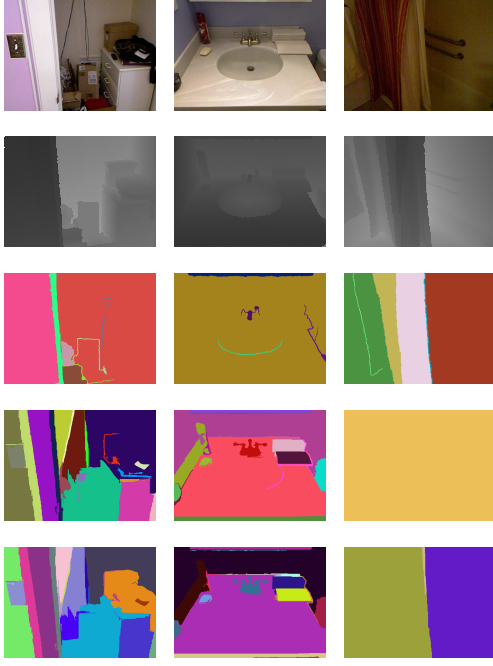


Fig. 2. Segmentation examples. Raw RGB and gray-scale depth samples (top two rows), graph-based depth segmentation (the third row), graph-based color segmentation (the fourth row) and graph-based RGB-D segmentation (bottom row)

$(v_i, v_j) \in E$  has a weight value  $w(v_i, v_j)$  to measure the dissimilarity between neighbouring vertices  $v_i$  and  $v_j$ . The similarity between vertices in homogeneous regions is higher than in discrepant regions. In our case, the weight  $w(p_i, p_j)$  refers to the distance between adjacent pixels in RGB and the gray-scale depth map simply defined as:

$$w(p_i, p_j) = \frac{\|RGBD(p_i) - RGBD(p_j)\|}{\sqrt{\|RGB(p_i) - RGB(p_j)\|^2 + (D(p_i) - D(p_j))^2}} \quad (1)$$

Here,  $RGB(p_i)$  represents the 3-dimensional vector of red, green and blue values of the pixel  $p_i$  in RGB color space. Respectively,  $D(p_i)$  is the gray-scale depth representing the distance of the pixel  $p_i$ . We firstly normalize four channels to the range  $[0...255]$ . To reduce influence of noise artifacts, we then apply a *Gaussian filter* to smooth each of the four channels before calculating the edge-weights. Finally, we construct a minimum spanning tree (*MST*) to merge similar regions using the minimum weighted edge between the regions.

Fig. 2 shows the graph-based segmentation results using the RGB, depth and the combined color and depth-based segmentation, respectively. Results indicate that fusion of RGB and depth provides the best segmentation quality. This is reasonable since color is an informative but sensitive feature, while the depth value is very well able to capture the compactness of objects, and thus may be used to remove over-segmentation using solely color.

### B. Foreground Saliency Measure

After segmentation, we obtain multiple different regions, each of which has similar homogeneous inner properties.

We expect to determine a saliency map from these regions. Based on segmented regions, the saliency measurements are calculated by foreground saliency features and background information, due to uniqueness of the foreground and consistency of the background.

We assume that foreground salience is significant and distinctive in an image. The distinctiveness refers to a region with high difference from its neighborhood. In this section, we propose a new foreground salience measure based on distribution of color, depth, position and area as an extension to the approach by [3].

In an image, the foreground salience value  $FS$  of each segmented region is calculated as:

$$FS(r_i) = ar(r_i) \cdot \sum_{j=1}^N \|cd(r_i) - cd(r_j)\|^2 \cdot w_{fs}(r_i, r_j) \quad (2)$$

$ar(r_i)$  is the area ratio of the region  $r_i$  to the entire image and is used as an additional weighting factor as opposed to [3].  $N$  represents the number of segmented regions, and  $cd(r)$  again is extended to use depth information and represents average color in *CIE Lab*-space combined with depth in region  $r$ . The Gaussian weight  $w_{fs}(r_i, r_j)$  is a local contrast term of foreground, as introduced in [3]:

$$w_{fs}(r_i, r_j) = \exp\left(-\frac{1}{\delta_{fs}^2} \|p_i - p_j\|^2\right)$$

$FS(r_i)$  effectively represents the rarity of a region  $r_i$  with color, depth  $cd_i$  and area compared to all other regions  $r_j$ .

### C. Background Measure

To be able to efficiently filter out true negative as well as false positive saliency candidates, it is important to define adequate measures that are able to identify the respective regions. Since image background usually features some nice properties like high area ratio and a wide spread close to the image borders, those characteristics are often modelled in literature for saliency filtering. While salient regions may be regarded as local region candidates, with a high variance in small areas, background regions may be interpreted as their counterpart featuring global homogeneity and high depth values, and as thus being highly suitable for broad-phase filtering. We follow this well known paradigm and wish to find a background representation that, combined with the foreground representation, is able to remove or weaken false saliency candidates and thus deliver better estimates.

Taking these considerations into account, we propose a measure  $CDis(r_i)$  to quantify the position of a region  $r_i$  in an image. It is defined as

$$CDis(r_i) = 1 - \frac{\min_j \|\bar{p}_i, c_j\|}{d_c} \quad (3)$$

where  $\bar{p}_i$  represents the center position of region  $r_i$ ,  $c_j$  represents the corner positions of the image and  $d_c$  is the distance of the image center to its corners. Assuming, that background is far away from the image center,  $CDis(r_i)$  may be interpreted geometrically as the distance from one

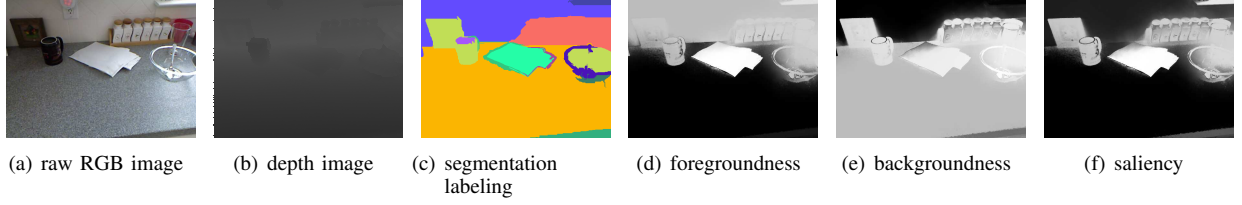


Fig. 3. Overview of our saliency mapping algorithm. Raw RGB-D data (a-b) is used to segment the scene into homogeneous regions, the resulting regions are shown in (c). In the next step, foreground and background measures are computed from color, depth and the spatial distribution of the regions (d-e). The final saliency map is obtained by combining the foreground and background maps (f).

region to the center of the image. It will be large for objects being close to corners and small for objects around the center of the image.

The *Boundary Connectivity* measure proposed by Zhu *et al.* [9] is used to quantify how heavily a region  $r_i$  is connected to the image boundaries:

$$BCon(r_i) = \frac{len(r_i)}{\sqrt{area(r_i)}} \quad (4)$$

where  $len(r_i)$  is a region perimeter on the boundary,  $area(r_i)$  refers to its area. The salient region has a small  $BCon(r_i)$  value, compared to the background region.

Finally, we define a dissimilarity measure for background saliency between regions as:

$$BS(r_i) = \sum_{j=1}^N \| area(r_i) - area(r_j) \| \cdot w_{bs}(r_i, r_j) \quad (5)$$

The background saliency  $BS(r_i)$  may be interpreted as the difference of the scale of region  $r_i$  compared to all other regions  $r_j$ . Similarly to  $FS$ , the Gaussian weight  $w_{bs}(r_i, r_j)$  is defined as

$$w_{bs}(r_i, r_j) = 1 - \exp\left(-\frac{(BCon(r_i) + CDis(r_i))^2}{\delta_{bs}^2}\right) \quad (6)$$

which effectively describes the position of the region, with values close to zero indicating that the region is far from corners and boundaries. The background usually has a larger value than other regions. The parameter  $\delta_{bs}$  controls the range of the background measure.

#### D. Final Saliency Map

Because of the complementarity between foreground and background saliency, we combine these two kinds of saliency maps together with different weights. We normalize foreground  $FS$  as well as background saliency  $BS$  to the range  $[0, 1]$  and assume both estimations to be independent. Hence we combine two measures as follows to compute the final saliency map  $Sal$ ,

$$Sal_i = BS_i \cdot \exp(-t \cdot FS_i) \quad (7)$$

The weight  $t$  is determined according to the information contained in the corresponding map. In our case, we set  $t = 3$  as the scaling factor throughout all experiments.

As can be seen in Fig. 3, the input *RGB-D* images (Fig.3(a) and Fig. 3(b)) are first merged into homogeneous regions

as depicted in Fig. 3(c). Two samples of foreground and background saliency maps are shown in Fig. 3(d) and Fig. 3(e). Finally, the resulting saliency map is shown in Fig. 3(f).

### III. EXPERIMENTAL RESULT

In this section, we evaluate our saliency detection method by using two different datasets: the RGB-D dataset provided by Ciptadi *et al.* [4], which we in the following refer to as *DSD* (depth salient data), and the *MSRA* dataset from Liu *et al.* [19]. Both datasets include images with complex background and low contrast objects, as well as manually labelled ground truth masks (*GT*) for salient object candidates. Several state-of-the-art saliency detection algorithms are chosen for comparison and are in the following referred to as SF [3], MSS [14], IG [13], AC [18], IT [12], MZ [20] and SR [10], respectively. To be more precise, we compare the results of different approaches on the introduced datasets if saliency maps are available for the current benchmark dataset. So for example, RGB based algorithms are not evaluated on the RGB-D dataset since they are not tailored to make use of the additional depth information (e.g. AC, SR). On the other hand, we extended some of the algorithms, e.g. IG [13] to additionally utilize depth information if this was possible in a straightforward fashion.

#### A. Evaluation on *DSD* dataset

The *DSD* dataset is an RGB-D dataset, comprising 80 RGB-D images using a mobile robot in a real-world indoor environment. For performance evaluation, we first give a visual comparison of different methods on this dataset. Three image sample results on the *DSD* dataset are shown in the top three rows of Fig. 4. The first column represents the input RGB image samples and the last column depicts the binary ground truth masks. Visually, our method (FBS) performs best compared to other methods in regard to the *GT* masks and delivers best results in regard to our desired saliency.

Since performance of saliency is highly dependent on the desired properties and the interpretation of important objects in scenes, it is generally not easy to compare different methods. Therefore, to be able to quantify the different results over the whole database and complying with the related work (e.g [3]), we choose the *mean absolute error (MAE)* as a measure, which simply describes the difference between the obtained saliency map  $S$  and the *GT*.

Conforming to the visual impression, Fig.5 shows that our method also outperforms the other approaches in regard to



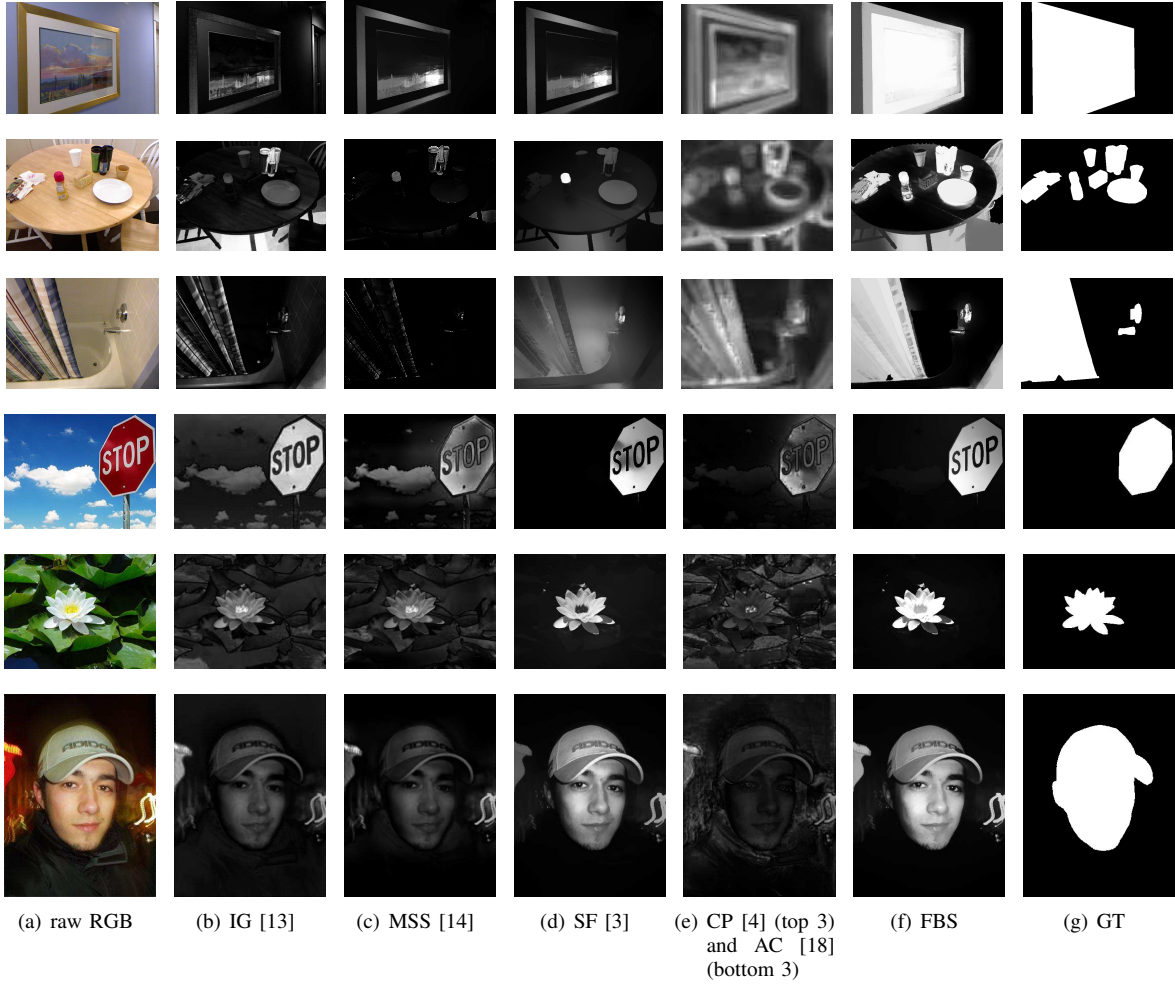


Fig. 4. Visual comparison of saliency maps obtained by different state-of-the-art algorithms. Our method produces saliency maps that highlight the whole object region and outline the foreground better than other methods. The top three image samples are taken from the RGB-D dataset published in [4], the bottom three rows are RGB samples without depth from the *MSRA* dataset in [19].

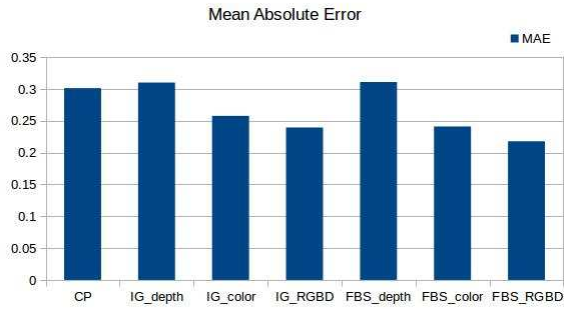


Fig. 5. Mean absolute errors of different algorithms on the DSD dataset.

MAE, and provides a better estimate of the desired saliency maps. Since the MAE is designed to capture pixel intensity differences, it gives a good approximation of the performance of the compared algorithms.

Next we analyze the performance of our algorithm in regard to the depth extensions. Our proposed approach as expected delivers the best results when we use the combined RGB-D version, although very good results are already

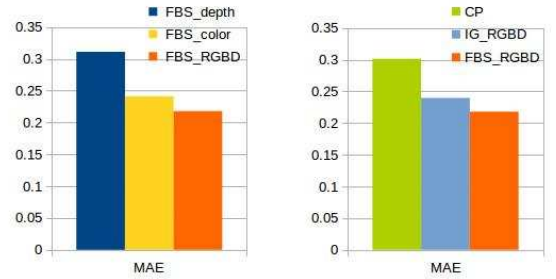


Fig. 6. Mean absolute errors on RGB-D dataset (DSD [4]). Results obtained with our proposed method (FBS) using color, depth and combined color and depth information (left). And comparison to alternative algorithms IG [13] extended to use depth (IG\_RGBD) and CP [4] (right)

achieved with color only as can be seen in Fig.6. Furthermore we compared the IG algorithm, that we extended to use additional depth information, with the CP algorithm, which naturally is tailored for depth data. The minor extensions to the RGB-D version of the IG algorithm make it almost as powerful as the proposed method.

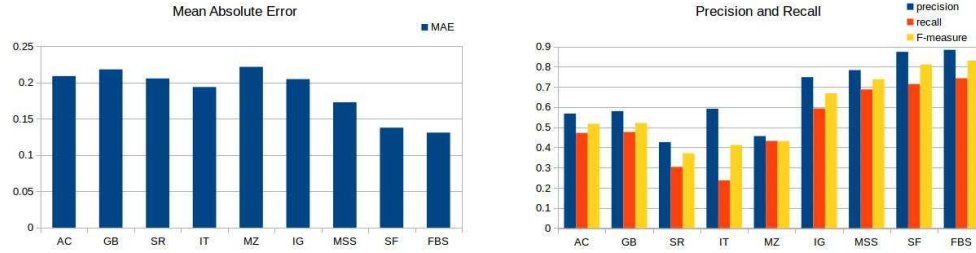


Fig. 7. Comparison of mean absolute errors and precision, recall and F-measure bars for different algorithms on the MSRA-1000 RGB database.

## B. Evaluation on MSRA dataset

To further evaluate our algorithm, we also perform experiments on the *MSRA* dataset. *MSRA* [19] is a standard benchmark dataset comprising 1000 instances with accurate human-labelled masks (*GT*) for salient objects. Visual samples are again presented the bottom three rows in Fig.4. Since this dataset does not provide depth image samples, we plug in a SLIC based segmentation into the proposed method which does not depend on depth data. For the other stages of our pipeline, we create dummy depth images with all depth values initialized to zero. Furthermore, we evaluate and compare to the RGB-only based AC algorithm instead of the previously evaluated CP algorithm on RGB-D. The vision samples again demonstrate, that our saliency approach is able to compete with other state of the art methods, even when only color is available. In our final analysis we additionally evaluated *precision and recall* bars on the whole *MSRA* dataset as depicted in Fig.7. The obtained results once again emphasize the robustness and accuracy of our proposed method, even on color-only image samples.

## IV. CONCLUSION

In this paper, we proposed a saliency detection method which is able to successfully detect multiple objects on a mobile robot. To solve the defined task of detecting salient objects in the scene, we propose an improved RGB-D based segmentation method to be able to automatically cluster major regions of interest. The clustering maintains the uniqueness and consistency of different objects and decreases computational complexity for the later stages, which rely on color and depth information. Furthermore, we proposed a new method to effectively generate a saliency map by combining foreground and background saliency maps. We utilized contrast-based saliency measures by color, depth and spatial layout information. Depth was used to recoup complementary object information for segmentation as well as saliency tasks. The proposed approach is very well suited to compete with state-of-the-art saliency algorithms and even outperformed them on both evaluated datasets.

## REFERENCES

- [1] P. Wang, L. Zhou, G. Wei, and H. Qiao, "salient region detection based on local and global saliency," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, (Hongkong, China), pp. 1546–1551, May 2014.
- [2] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, pp. 433–449, 2014.
- [3] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 733–740, 2012.
- [4] A. Ciptadi, T. Hermans, and J. M. Rehg, "An In Depth View of Saliency," in *British Machine Vision Conference (BMVC)*, (Bristol, United Kingdom), September 2013.
- [5] C. Lang, T. V. Nguyen, H. Katti, K. Yadati, M. S. Kankanalli, and S. Yan, "Depth matters: Influence of depth cues on visual saliency," in *IEEE Pro. European Conf. Computer Vision (ECCV)*, vol. 7573, (Berlin, Heidelberg), pp. 101–115, 2012.
- [6] L. Jiang, A. Koch, S. A. Scherer, and A. Zell, "Multi-class fruit classification using RGB-D data for indoor robots," in *IEEE Int. Conf. Robotics and Biomimetics (ROBIO)*, (Shenzhen), Dec. 2013.
- [7] L. Jiang, A. Koch, and A. Zell, "Object recognition and tracking for indoor robots using an rgb-d sensor," in *Int. Conf. Intelligent Autonomous Systems (IAS-13)*, (Padova, Italy), Jul. 2014.
- [8] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, (Portland), pp. 2083–2090, Jun. 2013.
- [9] W. Zhu, S. Liang, Y. W. Wei, and J. Sun, "Saliency optimization from robust background detection," in *IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, (Columbus, Ohio), Jun. 2014.
- [10] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2007.
- [11] L. Mai, Y. Niu, and F. Liu, "Saliency aggregation: A data-driven approach," in *IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 1131–1138, June 2013.
- [12] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 20, pp. 1254–1259, nov. 1998.
- [13] R. Achanta, S. S. Hemami, F. J. Estrada, and S. Ssstrunk, "Frequency-tuned salient region detection," in *IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 1597–1604, 2009.
- [14] R. Achanta and S. Ssstrunk, "Saliency Detection using Maximum Symmetric Surround," in *IEEE Int. Conf. Image Processing (ICIP)*, (Hong Kong), Sep. 2010.
- [15] A. Maki, P. Nordlund, and J.-O. Eklundh, "A computational model of depth-based attention," in *IEEE Int. Conf. Pattern Recognition (ICPR)*, vol. 4, pp. 734–739, Aug 1996.
- [16] N. Ouerhani and H. Hugli, "Computing visual attention from scene depth," in *IEEE Int. Conf. Pattern Recognition (ICPR)*, vol. 1, pp. 375–378 vol.1, Sep. 2000.
- [17] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph based image segmentation," *Int. J. Comput. Vision*, vol. 59, pp. 167–181, sep. 2004.
- [18] R. Achanta, F. Estrada, P. Wils, and S. Ssstrunk, "Salient region detection and segmentation," in *Pro. Int. Conf. Computer Vision Systems*, (Berlin, Heidelberg), pp. 66–75, 2008.
- [19] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 33, pp. 353–367, feb. 2011.
- [20] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Pro. Conf. Multimedia*, (New York, NY, USA), pp. 374–381, ACM, 2003.