Salient Obstacle Avoidance for Robotic Systems

Anonymous Submission

*Anonymous Afﬁliation*

Abstract

Salient object detection is a very active area of interest, based on human selective attention which shifts our focus across prominent areas in a scene. Many models have been created and achieved impressive accuracies, however, this often comes at a high computational efficiency cost making them less attractive for real-time robotic systems. We develop a novel salient object detection algorithm and investigate the trade-off between accuracy and computation time by comparing pixel-based and region-based processing. Salient object detection has been used in a variety of applications, however, this work is focused on developing an algorithm considering obstacle avoidance for a robot operating in a cluttered and dynamic environment. The strategy is evaluated on the publicly available MSRA10K salient object dataset, against three current state-of-the-art methods.

**Keywords:** Saliency, SLIC, Obstacle Avoidance.

# 1. Introduction

Selective attention is a characteristic of visual perception contained within the human visual system. With the volume of visual stimuli entering the eye; approximated between 108 – 109 bits per second [Borji et al., 2013], the selective attention mechanism guides our focus across the conspicuous regions within a scene. Derived from this ability of filtering non-interesting information, salient object detection became a popular research topic, which is the process of detecting and segmenting the salient object/region within a scene. Many models have been proposed

[Fu et al., 2018; Ishikura et al., 2018; Aytekin et al., 2018; Zhang et al., 2017; Frintrop et al., 2007], however, saliency detection is still a research challenge due to the vast spectrum of objects, varying environments, complex backgrounds, lack of representation of top-down feature cues and computational complexity. While many algorithms achieve impressive detection results, this often is obtained using computationally expensive algorithmic approaches. In aiming to replicate visual attention, some developed models have glanced over the observation that humans do not process every detail within a scene, but rather process regions to locate the salient object.

Saliency detection models can be partitioned into two classes, top-down and bottom-up [Itti et al., 1998]. Top-down models are slow, task-dependent and based on prior information of the scene or object, whereas bottom-up approaches focus on detecting regions or points that focus people’s attention, determined by low-level stimuli within a scene. Some significant saliency algorithms that have been proposed include [Yan et al., 2013; Niu et al., 2016; Yang et al., 2017]. [Yan et al., 2013] propose a hierarchical saliency model. Feature cues are calculated on three image layers extracted from the input image. The final saliency map is computed using a hierarchical inference method, which is based on a tree-structure graph model. This approach outputs accurate saliency maps, however, the number of steps within the algorithm suggest it is not feasible for a real-time application. [Niu et al., 2016] present a contrast saliency model using a patch-based approach, incorporating global statistics and surrounding contrast operators. A salient object detection method exploiting edges for super-pixel segmentation is presented in [Yang et al., 2017].

Within an image, a region or super-pixel can be described as a cluster of pixels with similar colour values and proximity. In this paper, we propose a novel algorithm that will detect the most salient object within a scene using a region-based approach. We also analyse the process of how saliency algorithms are applied to images, in terms of processing an image at a pixel-level in comparison with region-based processing. Pixel-based and region-based processing in terms of saliency detection can be defined as the estimation of whether the respective pixel or region belongs to the salient object.

Salient object detection has been applied to multiple applications such as: object recognition [Rutishauser et al., 2004], image segmentation [Ko et al., 2006] and context-aware image editing [Wang et al., 2008], however, only recently has saliency object detection been utilized within the area of robotics [Scharfenberger et al., 2013]. This may be due to many existing salient object detection algorithms being quite computationally expensive and unsuitable for real-time robotic applications. We address this by developing an efficient and accurate approach to saliency object detection for the task of obstacle avoidance for robots operating within a cluttered or dynamic environment. A real-time salient obstacle avoidance algorithm will enable robotic platforms to detect and avoid any immediate obstacle within the scene, whether it be for the goal of avoiding obstacles while traversing a specific path or avoiding obstacles while completing manipulation tasks. Proposing a solution to real-time obstacle avoidance using vision offers additional resilience within a real-world environment, while providing a framework for introducing more biologically-inspired solutions to robotic systems.

In this paper, the proposed model is compared with three state-of-the-art saliency approaches. Firstly, we use the Salient Region Detection and Segmentation (SRDAS) approach in [Achanta et al., 2008], which implements a pixel-level approach that varies the filter sizes of compared pixels. SRDAS computes saliency using two feature cues: luminance and colour. Secondly, we compare with the Fast and Efficient Saliency Detection Using Sparse Sampling and Kernel Density Estimation (FES) [Tavakoli et al., 2011], which divides an image into windows and computes saliency using a center-surround approach. Saliency maps are produced using sparse sampling and kernel density estimation. Finally, we compare with the Saliency Detection in the Compressed Domain for Adaptive Image Retargeting (DCT) [Fang et al., 2012]. This approach processes images in blocks, and considers intensity, colour and texture as feature cues in computing saliency.

The remainder of the paper is structured as follows. Section 2 outlines the methodology and the proposed algorithm, including the approaches used for evaluation in this paper, specifically, Salient Region Detection and Segmentation (SRDAS) [Achanta et al., 2008], Fast and Efficient Saliency Detection using Sparse Sampling and Kernel density Estimation (FES) [Tavakoli et al., 2011] and Saliency Detection in the Compressed Domain for Adaptive Image Retargeting (DCT) [Fang et al., 2012]. Experimental results are detailed in Section 4, with Section 5 concluding the paper.

|  |
| --- |
|  |
| Figure 1. Top-row: Overview of the proposed saliency algorithm. We compute saliency using global colour contrast and gradient contrast. Bottom-row: Adaptation of algorithm to process regions |

# 2. Saliency Detection Methodology

With the aim of creating an efficient, yet accurate approach to salient object detection, a saliency model was designed and implemented (outlined in Section 3.1) using a pixel-based approach and also a region-based approach that uses super-pixels generated via the simple linear iterative plane space, to efficiently generate almost uniform, compact super-pixels. A graphical representation of both approaches is represented in Figure 1.

## 2.1 Proposed pixel-based approach

A new model for detecting the most salient object within a scene is presented. The proposed model considers two features cues: global colour contrast () and gradient contrast (). Saliency, measured in each of the feature channels is computed per-pixel. To determine the salient object, two saliency maps are computed from each feature cue. After this, the computed saliency maps are combined to produce the final saliency map. Originally, the algorithm also considered centre-bias as a feature cue. However, when considering obstacle avoidance, there are a number of considerations that resulted in not implementing this including the fact that a mobile robot will be navigating an environment and therefore the salient object (obstacle) may not always be in the centre of the field of view.

Colour is one of the main features for attracting human attention within an image. Pixels that have a high contrast to their surroundings when considering colour are considered to be salient [Zhang et al., 2017]. Many models compare pixels/regions to their neighbours which is known as local contrast. Feature maps from local colour contrast tend to be noisy and mainly highlight the edges of the salient area/object, whereas global contrast computes the contrast of a pixel/region in relation to the colour value of all of the remaining pixels/regions within an image. The calculation to compute the colour contrast () of a pixel against the average colour value of the remaining pixels using the Euclidean distance, is defined as:

|  |  |
| --- | --- |
|  | (1) |

where is the total number of pixels, and represent the average L\* a\* b\* values of pixels and respectively.

Gradient contrast (GC) can play an integral part in approximating the salient object within a scene. While retaining important information about the background of an image, it also helps to preserve the contour of the salient object. GC measures the magnitude of local grayscale changes in an image.

We calculate the gradient magnitude per pixel using a Sobel gradient operator. From the obtained gradient map , we define gradient contrast as:

|  |  |
| --- | --- |
|  | (2) |

where and are the gradient values of the pixels and respectively.

To obtain the final saliency map, each of the feature cue outputs must be fused together. As suggested in [Cheng et al., 2014], we investigated the most appropriate approach for fusing multiple feature cues. We tested multiple integration schemes (e.g. max, \* and +) and normalised the saliency map from each feature cue, before amalgamating these. The resultant saliency value per pixel is defined by:

|  |  |
| --- | --- |
|  | (3) |

where and are the respective feature cues of global color contrast and gradient contrast, and defines the current pixel. The pixel in one feature channel is fused with its corresponding pixel within the other feature channel using addition as the integration scheme.

## 2.2 Proposed region-based approach

The first step within the region-based model is to over-segment the image into regions/super-pixels using Simple Linear Iterative Clustering (SLIC) [Achanta et al., 2012]. SLIC groups pixels together with similar colour values in close proximity to form super-pixels. Images can then be processed in a region-based manner resulting in improved computational efficiency, in comparison with the time taken to process images using a pixel-based approach. The challenge with processing regions is achieving the correct balance between computation time and accuracy. With the super-pixel approach, the input RGB image is converted to L\* a\* b\* colour space, in preparation for clustering. The input image is segmented into super-pixelswhere denotes the total number of super-pixels. Experimentation was completed, measuring the average run-time and accuracy when varying the number of super-pixels within a scene. The results from this are discussed in Section 4, and outlined in Figure 3.

Global colour contrast and gradient contrast are computed per super-pixel within each image. Equation 1 can be adapted for defining the computation of global colour contrast () within the region-based approach:

|  |  |
| --- | --- |
|  | (4) |

where denotes the total number of super-pixels, and signify the average L\* a\* b\* values of super-pixels and respectively. Gradient contrast () is calculated per super-pixel using the following equation:

|  |  |
| --- | --- |
|  | (5) |

with defined as the total super-pixels in image and denote the respective average gradient values of super-pixels and . The resultant saliency map for the proposed region-based approach can be obtained by:

|  |  |
| --- | --- |
|  | (6) |

Feature maps are amalgamated per super-pixel. Super-pixel within the global colour contrast feature channel is combined with the corresponding super-pixel from the gradient contrast feature channel.

# 3. Experimental Results

In order to validate the performance of the proposed approach, we used the MSRA10K salient object dataset [Cheng et al., 2015], which consists of structurally complex natural images, each accompanied by a ground-truth binary mask highlighting the salient object. 100 images were selected with varying salient objects and scenes, and each algorithm was evaluated on this set of images with respect to computational efficiency and accuracy. We evaluated the proposed Gradient and Colour Contrast Saliency (GCCS) approach against three current state-of-the-art models. To obtain a fair comparison, the SRDAS approach was adapted to process super-pixels (referred to as SRDAS\_SLIC) as well as the original pixel- based approach, which compares each pixel with neighbours at different filter scales, to formulate a saliency map.

Firstly, the proposed GCCS model was evaluated against SRDAS; their average run-times and accuracy measures are presented in Table 1. To measure efficiency, the average time taken to apply a saliency algorithm to the scene and obtain the output saliency map was used. Algorithmic accuracy (A) was calculated by comparing the output saliency map with the associated ground-truth mask and determining how well each algorithm segmented the salient object using the following equation:

|  |  |
| --- | --- |
|  | (7) |

where , , and are true positives, true negatives, false positives and false negatives respectively.

**Table 1. Results when comparing pixels-based model SRDAS with GCCS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **Avg. Runtime** | **Avg. Accuracy** | **Max Accuracy** |
| SRDAS | 32.73 secs | 74.19% | 96.12% |
| GCCS | 34 secs | 76.69% | 99.08% |

In computing accuracy, each saliency map was transformed into a binary mask image format using the Matlab implementation of Otsu’s method [Otsu et al., 1979], which computes a global image threshold to minimize the variance of thresholded black and white pixels/regions. As seen in Table 1, the proposed algorithm (GCCS) had a higher average accuracy across the dataset of 76.69%, compared with 74.19% achieved by SRDAS. While this result was encouraging, the recorded average run-time of 34 seconds could not be considered for implementation on a robotic platform within a dynamic environment. With this observation, the algorithm was adapted to process regions rather than pixels. To obtain the best accuracy in real-time, we varied the number of super-pixels per image across the dataset, and recorded the average processing time alongside the achieved accuracy. Figure 2 shows the results of the completed study, with the selected number of super-pixels being 100, 250, 500, 750 and 1000 respectively. The recorded accuracies range from 70.5% to 81.1%, however, increase in accuracy comes at a cost to efficiency. Based on the results presented in Figure 3, each image was segmented into 500 super-pixels as this yields a very respectable average accuracy of 78.3% at a small cost to processing time.

The implemented region-based algorithm was evaluated against two state-of-the-art saliency detection methods. For a fair comparison, SRDAS was also implemented with SLIC. As can be seen in Table 2, the proposed approach incorporating SLIC (GCCS\_SLIC) performs very well when considering the trade-off between accuracy and computational efficiency. FES recorded the fastest average runtime at 0.4 seconds, with the GCCS\_SLIC algorithm closely following at 0.75 seconds. What should be noted is that GCCS\_SLIC outperforms FES when considering average accuracy, with an increase of 13.8%. DCT recorded the highest average accuracy across the dataset, at an average cost of 1.74 seconds per image. When calculating the maximum accuracy, i.e. the percentage of each salient object that was correctly segmented by the algorithm, both proposed approaches, GCCS (pixel-level) and GCCS\_SLIC (region- level) performed best obtaining maximum accuracies of 99.08% and 98.99% respectively.

**Table 2. Results when comparing region-based models FES, DCT, SRDAS\_SLIC with GCCS\_SLIC**

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **Avg. Runtime** | **Avg. Accuracy** | **Max Accuracy** |
| FES | 0.4 secs. | 64.5% | 95.18% |
| DCT | 1.74 secs. | 81% | 97.9% |
| SRDAS\_SLIC | 2.23 secs | 54.29% | 95.10% |
| GCCS\_SLIC | 0.75 secs | 78.3% | 98.99% |

|  |
| --- |
|  |
| **Figure 2. Results showing accuracy vs. average run-time with a varying number of super-pixels** |

# 4. Conclusions

We have proposed a novel salient object detection model incorporating two feature cues. This approach was adapted to process at both pixel-level and region-based level. An evaluation of these implementations found that region-based approaches can achieve comparable accuracy, whilst vastly outperforming pixel-based approaches in terms of computational efficiency. We presented experimental results using the well-known MSRA10K salient object dataset, showing a significant improvement in computation time; taking less than one second to process a single image and output a saliency map. While a pixel-level approach captures more detail in the final saliency map, current pixel-level approaches are not feasible for real-time applications, such as obstacle avoidance for a mobile robot. However, the GCCS\_SLIC approach provides a clear solution for this.

Future work will investigate other feature cues that will aid in detecting salient obstacles within a scene or environment, such as orientation and depth. We also plan to implement a multi-scale approach to salient object detection to capture more information at different scales, and therefore increase accuracy and implement this using a Kinect and Summit XL robotic system.

# 5. References

[Achanta et al., 2008] R. Achanta, F. Estrada, P. Wils, and S. Sabine, “Salient Region Detection and Segmentation,” *Int. Conf. Comput. Vis. Syst. ICVS, 2008*, vol. 5008, pp. 66–75, 2008.

[Achanta et al., 2012] R. Achanta, A. Shaji, K. Smith, and A. Lucchi, “SLIC Superpixels Compared to State-of-the-Art Superpixel Methods,” *IEEE Trans. PATTERN Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, 2012.

[Aytekin et al., 2018] C. Aytekin, A. Iosifidis, and M. Gabbouj, “Probabilistic saliency estimation,” *Pattern Recognit.*, vol. 74, pp. 359–372, 2018.

[Borji et al., 2013] A. Borji and L. Itti, “State-of-the-Art in Visual Attention Modeling,” *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 35, no. 1, pp. 185–207, 2013.

[Cheng et al., 2014] Y. Cheng, H. Fu, X. Wei, J. Xiao, and X. Cao, “Depth Enhanced Saliency Detection Method,” in *Proceedings of International Conference on Internet Multimedia Computing and Service*, 2014, no. 1.

[Cheng et al., 2015] M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu, “Global Contrast Based Salient Region Detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, 2015.

[Fang et al., 2012] Y. Fang, Z. Chen, W. Lin, S. Member, C. Lin, and S. Member, “Saliency Detection in the Compressed Domain for Adaptive Image Retargeting,” *IEEE Trans. IMAGE Process.*, vol. 21, no. 9, pp. 3888–3901, 2012.

[Frintrop et al., 2007] S. Frintrop, M. Klodt, and E. Rome, “A Real-time Visual Attention System Using Integral Images,” in *International Conference on Computer Vision Systems*, 2007, no. ICVS.

[Fu et al., 2018] K. Fu, I. Y. Gu, and J. Yang, “Spectral salient object detection,” *Neurocomputing*, vol. 275, pp. 788–803, 2018.

[Ishikura et al., 2018] K. Ishikura, N. Kurita, D. M. Chandler, S. Member, and G. Ohashi, “Saliency Detection Based on Multiscale Extrema of Local Perceptual Color Differences,” *IEEE Trans. IMAGE Process.*, vol. 27, no. 2, pp. 703–717, 2018.

[Itti et al., 1998] L. Itti, C. Koch, and E. Niebur, “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis,” in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998, vol. 20, no. 11, pp. 1254–1259.

[Ko et al., 2006] B. C. Ko and J. Nam, “Object-of-interest image segmentation based on human attention and semantic region,” Opt. Soc. Am. A, vol. 23, no. 10, pp. 2462–2470, 2006.

[Niu et al., 2016] J. Niu, X. Bu, and K. Qian, “Exploiting contrast cues for salient region detection,” in *Multimedia Tools and Applications*, 2016, no. March 2016, pp. 1–15.

[Otsu et al., 1979] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, 1979.

[Rutishauser et al., 2004] U. Rutishauser, D. Walther, C. Koch, and P. Perona, “Is bottom-up attention useful for object recognition?,” in *IEEE Comp. Soc. Conf. CVPR*, 2004, p. II-37-II-44.

[Scharfenberger et al., 2013] C. Scharfenberger, S. L. Waslander, J. S. Zelek, and D. A. Clausi, “Existence Detection of Objects in Images for Robot Vision Using Saliency Histogram Features,” in *International Conference on Computer and Robot Vision*, 2013.

[Tavakoli et al., 2011] H. R. Tavakoli, E. Rahtu, and J. Heikkil, “Fast and Efficient Saliency Detection Using,” *Image Anal. SCIA*, vol. 6688, pp. 666–675, 2011.

[Wang et al., 2008] Y. Wang, C. Tai, O. Sorkine, and T. Lee, “Optimized Scale-and-Stretch for Image Resizing,” in *ACM Trans. Graph.*, 2008, vol. 27, no. 5, pp. 1–8.

[Yan et al., 2013]. Q. Yan, L. Xu, J. Shi, and J. Jia, “Hierarchical Saliency Detection,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[Yang et al., 2017] B. Yang, X. Zhang, L. Chen, H. Yang, and Z. Gao, “Edge guided salient object detection,” *Neurocomputing*, vol. 221, no. August 2016, pp. 60–71, 2017.

[Zhang et al., 2017] Q. Zhang, J. Lin, Y. Tao, W. Li, and Y. Shi, “Salient object detection via color and texture cues,” *Neurocomputing*, vol. 243, pp. 35–48, 2017.

|  |
| --- |
|  |
| **Figure 3. Visual comparison examples of methods on the MRSA10K dataset: Column (a) is the input image; Columns (b) – (g) are output saliency maps produced by (b) SRDAS [7], (c) SLIC\_SRDAS, (d) FES [13], (e) DCT [14], (f) GCCS (pixel-level), (g) GCCS (super-pixels); (h) is the ground truth binary mask.** |