Bio-Inspired Visual Attention Model and Saliency Guided Object Segmentation

Lijuan Duan[1], Jili Gu[1], Zhen Yang[1], Jun Miao[2,*], Wei Ma[1], Chunpeng Wu[3]

[1]College of Computer Science and Technology, Beijing University of Technology, Beijing, China
[2]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[*]jmiao@ict.ac.cn
[3]Fujitsu Research & Development Center Co. Ltd., Beijing, China

Abstract. In this paper, we present a saliency guided image object segment method. We suppose that saliency maps can indicate informative regions, and filter out background in images. To produce perceptual satisfactory salient objects, we use our bio-inspired saliency measure which integrating three factors: dissimilarity, spatial distance and central bias to compute saliency map. Then the saliency map is used as the importance map in the salient object segment method. Experimental results demonstrate that our method outperforms previous saliency detection method, yielding higher precision (0.7669) and better recall rates (0.825), F-Measure (0.7545), when evaluated using one of the largest publicly available data sets.

Keywords: visual attention, dissimilarity, spatial distance, central bias, salient object detection

Introduction

Visual attention has been studied by researchers in physiology, psychology, neural systems, and computer vision for a long time. It is useful for many computer vision tasks such as content-based image retrieval, segmentation, and object detection. Computationally modeling such mechanism has been widely studied in order to identify which part of image is more useful when analyzing an image in recent years [1], [2], [3]. Applications of the models such as image classification [4], image segmentation [5] and object detection [6] have become a popular research topic. We study visual saliency by detecting a salient object in an input image. The automatic detection of visually salient regions in images is useful for image segmentation, adaptive region-of-interest based image compression, object recognition, and content aware image resizing.

Salient object detection is defined as an image segmentation problem, where the salient object is separated from the image background [7]. It is supported by research on human vision system that the human brain and visual system pay more attention to

adfa, p. 1, 2013.

some parts of an image. The recognition and localization of searching targets in complex visual scenes is still a challenge problem for computer vision systems. However, this task is performed by humans in a more intuitive and efficient manner by selecting only a few regions to focus on. Observers never form a complete and detailed representation of their surroundings [8]. For example, in Figure 1, the tomato, dog, and woman attract the most visual attention in each respective image. Therefore, salient object detection is formulated as a binary labeling problem that separates a salient object from the background. Like face detection, we learn to detect a familiar object; unlike face detection, we detect a familiar yet unknown object in an image [7]. There are many traditional methods of saliency detection which are used in unsupervised object segmentation. To produce perceptual satisfactory salient object segmented images, we use the saliency map as the importance map in the salient object detection method. Han et al. [9] use low-level features of color, texture, and edges in a Markov random field framework to grow salient object regions from seed values in the saliency maps. Salient regions are selected by Ko and Nam [10] using a support vector machine trained on image segment features to select the salient regions of interest using Itti's maps, which are then clustered to extract the salient objects. Ma and Zhang [11] utilize fuzzy region growing on saliency maps to confine salient regions within a rectangular region. Achanta et al. [12] average saliency values within image segments produced by mean-shift segmentation, and then find salient objects by identifying image segments that have average saliency above a threshold that is set to be twice the mean saliency value of the entire image. More recently, Cheng Mingming et al. [13] propose a region-based contrast saliency detection method (RC) to show that segmentation-based method is better than their pixel-based method (HC) and then iteratively use GrabCut to refine the segmentation result initially obtained by thresholding the saliency map.

The paper is organized as follows: Review of our saliency detection is in Section 2. In Section 3, we introduce a salient object segment method. In Section 4, we compare the performance of salient object methods based on different saliency measures. The conclusions are given in Section 5.

Review of Our Saliency Measure

In this paper, we use our biologically inspired saliency measure proposed in [14] to detect the salient object. The saliency measure integrates three factors: dissimilarity, spatial distance and central bias, and these three factors are supported by research on human vision system. The dissimilarity is evaluated by a center-surround operator simulating the visual receptive field [1], and this structure is a general computational principle in the retina and primary visual cortex [15]. The spatial distance is supported by the research [16] on foveation of human vision system. Human samples the visual field by a variable resolution, and the resolution is highest at center (fovea) and drops rapidly toward the periphery [17]. In addition, according to previous studies on the distribution of human fixations [18], human tend to look at the center of images. This fact is also known as central bias which reflects that photographer tent to center objects of interest [19]. There are five main steps in our saliency detection method: changing an image to YCbCr color space, splitting image into patches, reducing dimensionality, evaluating the spatially-weighted dissimilarity. First Non-overlapping patches drawn from an image are represented as vectors of pixels, and all patches are mapped into a reduced dimensional space. The saliency of image patch p_i is calculated as

$$\text{Saliency}(p_i) = w_1(p_i).GD(p_i) \tag{1}$$

Where $w_1(p_i)$ represents central bias and $GD(p_i)$ is the global dissimilarity. $w_1(p_i)$ and $GD(p_i)$ are computed as

$$w_1(p_i) = 1 - \frac{DistToCenter(p_i)}{D} \tag{2}$$

$$\text{GD}(p_i) = \sum_{j=1}^{L}\{\, w_2(p_i,p_j).Dissimilarity(p_i,p_j)\,\} \tag{3}$$

In Eq. 2, $DistToCenter(p_i)$ is the spatial distance between patch p_i and center of the original image, and $\text{D} = max_j\{(DistToCenter(p_j)\,\}$ is a normalization factor. In

Eq. 3, L is total number of patches, $w_2(p_i,p_j)$ is inverse of spatial distance, and $Dissimilarity(p_i,p_j)$ is dissimilarity of feature response between patches. $w_2(p_i,p_j)$ and $Dissimilarity(p_i,p_j)$ are computed as

$$w_2(p_i,p_j) = \frac{1}{1+Dist(p_i,p_j)} \tag{4}$$

$$Dissimilarity(p_i,p_j) = ||\, f_i - f_j|| \tag{5}$$

In Eq. 4, $Dist(p_i,p_j)$ is the spatial distance between patch p_i and patch p_j in the image. In Eq. 5, feature vectors f_i and f_j correspond to patch p_i and patch p_j respectively. Finally, the saliency map is normalized and resized to the scale of the original image, and then is smoothed with a Gaussian filter ($\sigma = 3$).

Saliency Guided Object Segmentation

The true usefulness of a saliency map is determined by the application. In this paper we consider the use of saliency maps in salient object segmentation. To produce perceptual satisfactory salient objects, we use our bio-inspired saliency measure which integrating three factors: dissimilarity, spatial distance and central bias to compute saliency map. Then the saliency map is used as the importance map in the salient object segment method. To segment a salient object, we need to binarize the saliency map such that ones (white pixels) correspond to salient object pixels while zeros (black pixels) correspond to the background [1]. Fixed parameters of 7, 10, and 20 for sigmaS, sigmaR, and minRegion area used respectively, for all the images (see [20]). In the experiment we use the image dependent adaptive threshold proposed by [12], which average saliency values within image segments produced by mean-shift segmentation, and then find salient objects by identifying image segments that have average saliency above a threshold that is set to be twice the mean saliency value of the entire image:

$$T_a = \frac{2}{W*H} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} S(x, y) \quad (6)$$

where W and H are the width and height of the saliency map in pixels, respectively, and $S(x, y)$ is the saliency value of the pixel at position (x, y). Using this approach, we obtain binarized maps of salient object from each of the saliency algorithms.

Experimental Validation

Experiment is conducted on the publicly available database provided by Achanta et al. [12] to evaluate our performance. The ground truths of this database are binary images in which salient objects are accurately marked by human. Performance of salient object detection based on different saliency detection methods including ours is compared. The same parameters of our method will be used across all images. According to our previous parameter settings [14], the color space is YCbCr, the size of image patch is 14x14 and the dimensions to which each feature vector reduced is 11. Our saliency detection method with above parameters outperforms some state-of-the-art [1], [21], [22], [23] on predicting human fixations, please see [14] for details.

Qualitative comparison between Saliency Measures

We provide an exhaustive comparison of our approach to other six state-of-art methods on a database of 1000 images [12] with binary ground truth [2]. Saliency maps of previous works are provided by [14]. Comparison of salient object detection results between our method and other six saliency detection method in more images are shown in Fig. 1. The first row are original images, the second row are ground truth images according with the first row. From the third row, the five images on each row are segmented results by Itti et al. [1], Ma and Zhang [11], Harel et al. [22], Hou et al. [21], R. Achanta [23], R. Achanta [12]. Experiment show that our ultimate saliency maps are superior to the other saliency maps produced with a segmented result which prove the effectiveness of our saliency map evaluation method. The saliency method in [14] highlights the woman and dragon in image with well-defined border and suppresses background regions efficiently. In all experiments, our approach consistently produces results closest to ground truth. However, the image in fifth column, its ground truth image of other subjects may be same with our detected one. The key objective of attention detection should be to locate position of a salient object as accurately as possible, i.e. with high precision, recall, and F-Measure. Because background regions are successfully suppressed in our saliency map, the binary mask generated from our saliency map is more accurate than that from other methods (see also Fig. 2).

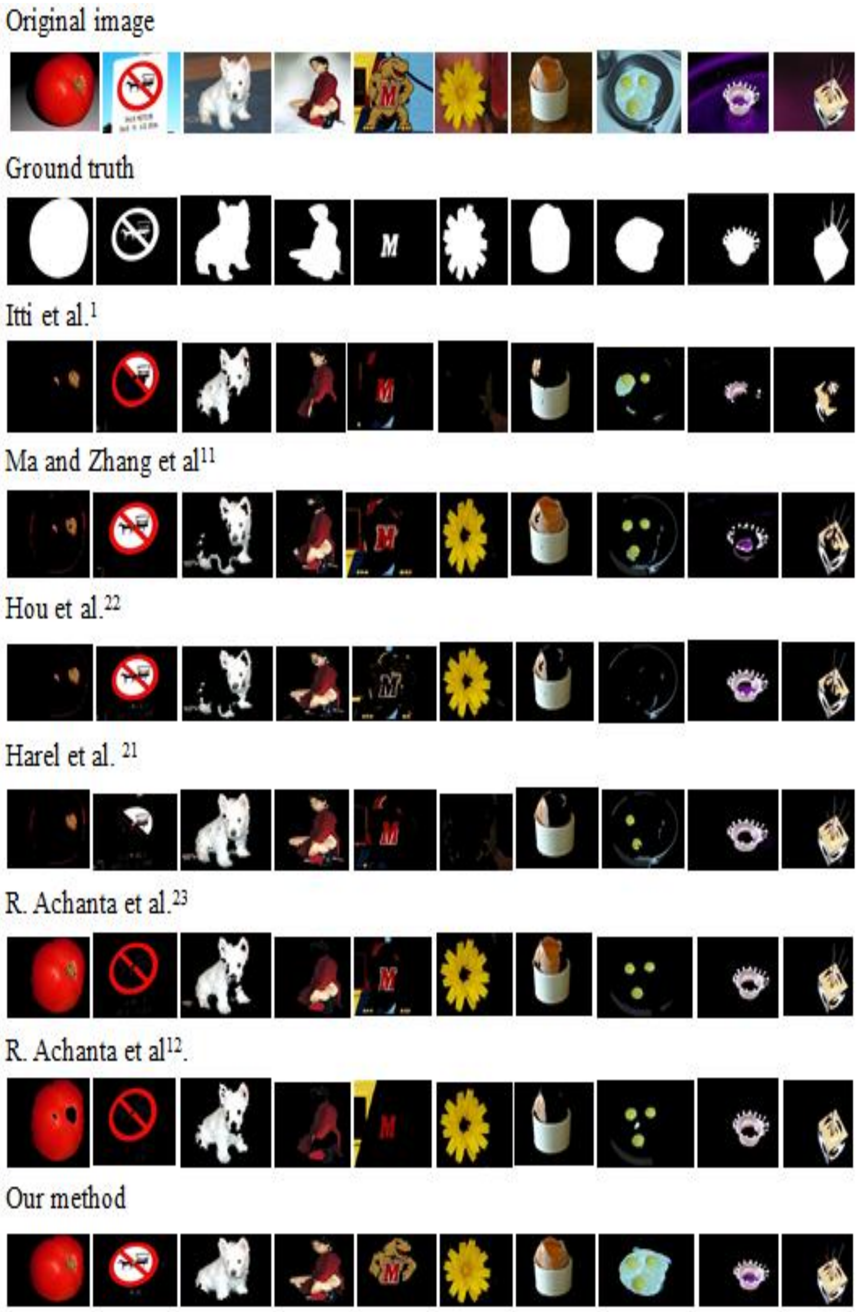

Fig. 1. Comparison between salient object detection based on seven different saliency measures as follows: Itti et al. [1], Ma and Zhang [11], Harel et al. [22], Hou et al. [21], R. Achanta [23], R. Achanta [12] and our method. The first row are original images, the second row are ground truth images according with the first row.

Quantitative comparison between Saliency Measures

We evaluate the performance of our algorithm measuring its precision and recall rate. Precision corresponds to the percentage of salient pixels correctly assigned, while recall corresponds to the fraction of detected salient pixels in relation to the ground truth number of salient pixels. High recall can be achieved at the expense of reducing the precision and vice-versa so it is important to evaluate both measures together. With a ground-truth saliency map G, for any detected salient region mask A, we use following measurements:

$$\text{Precision} = \frac{\sum_x g_x a_x}{\sum_x a_x} \tag{7}$$

$$\text{Recall} = \frac{\sum_x g_x a_x}{\sum_x g_x} \tag{8}$$

F-Measure is the weighted harmonic mean of precision and recall, with a non-negative β :

$$F_\beta = \frac{(1+\beta^2)\text{Precision}*\text{Recall}}{\beta^2 * Precision + Recall} \tag{9}$$

F_β (Eq. 9) are obtained over the same ground-truth database by Achanta et al. [12]. β^2= 0.3 is used in our work to weigh precision more than recall. The comparison is shown in Table1 and Fig. 2 which are according with [12]. Itti's saliency detection method has a high precision (0.7919) but very poor recall (0.4643). Among all the methods, our method shows the highest recall value, third precision and second F-Measure. Compared with Achanta [12], our method has a higher recall but a lower precision. However, like all the other saliency detection methods, it can fail when the object of interest is not distinct from the background.

Table 1. Comparison between salient object detection precision, recall and F- Measure based on seven different saliency measures as follows: Itti et al. [1], Ma and Zhang [11], Harel et al. [22], Hou et al. [21], R. Achanta [23], R. Achanta [12].

	Precision	Recall	F-Measure
Itti [1]	0.7919	0.4643	0.6336
Ma and Zhang [11]	0.675	0.6613	0.6459
Harel [22]	0.7321	0.7519	0.7104
Hou [21]	0.6581	0.5573	0.5998
R.Achanta [23]	0.7543	0.6983	0.7152
R.Achanta [12]	0.8363	0.7936	0.8048
Our Method	0.7669	0.825	0.7545

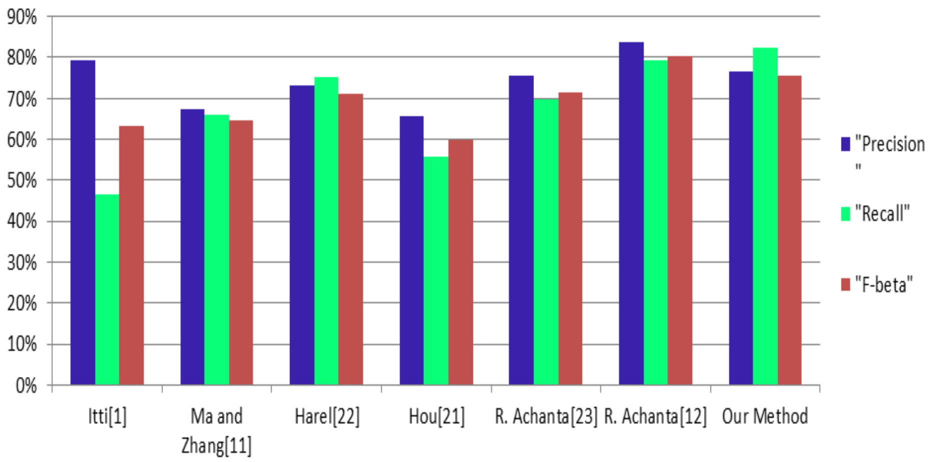

Fig. 2. Comparison between salient object detection precision, recall and F-Measure based on seven different saliency measures as follows: Itti et al. [1], Ma and Zhang [11], Harel et al. [22], Hou et al. [21], R. Achanta [23], R. Achanta [12].

Conclusion and Discussion

We use the saliency map as the importance map in the salient object method. Experimental results show that our model could generate high quality saliency maps that highlight the whole salient object with well-defined boundary, meanwhile successfully suppress the background regions. The resulting saliency maps of our method on Achanta's dataset of 1000 images are better suited to salient object segmentation, demonstrating highest recall value, third precision and second F-Measure values. Salient object detection has wider applications. For example, a more semantic, object-based image similarity can be defined with salient object detection for content-based image retrieval.

Acknowledgment

This research is partially sponsored by National Basic Research Program of China (No.2009CB320900), Natural Science Foundation of China (61175115, 61003105, 61070116, 61272320，61070149 and 61001108), and the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions （CIT&TCD201304035).

References

1. Itti, L., Koch, C., Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE TPAMI, vol. 20, pp. 1254-1259, 1998.

2. Gao, D., Vasconcelos, N.: Bottom-Up Saliency is a Discriminant Process. IEEE ICCV, pp. 1-6 ,2007.
3. Murray, N., Vanrell, M., Otazu, X., Parraga, C., A.: Saliency Estimation Using A Non-Parametric Low-Level Vision Model. IEEE CVPR, pp. 433-440, 2011.
4. Kanan, C., Cottrell, G.: Robust Classification of Objects, Faces, and Flowers Using Natural Image Statistics. IEEE CVPR, pp. 2472-2479, 2010.
5. Yu, H., Li, J., Tian, Y., Huang, H.: Automatic Interesting Object Extraction from Images Using Complementary Saliency Maps. ACM Multimedia, pp. 891-894, 2010.
6. Navalpakkam, V., Itti, L.: An Intergrated Model of Top-Down and Bottom-Up Attention for Optimizing Detection Speed. IEEE CVPR, pp. 2049-2056, 2006.
7. Liu, Tie; Yuan, Zejian; Sun, Jian; Wang, Jingdong; Zheng, Nanning; Tang, Xiaoou; Shum, Heung-Yeung Source. Learning to detect a salient object. IEEE Transactions on Pattern Analysis and Machine Intelligence, v 33, n 2, p 353-367, 2011.
8. R. Rensink, K. O'Regan, J. Clark. To see or not to see: The need for attention to perceive changes in scenes. In: Psychological Sciences,1997.
9. J. Han, K. Ngan, M. Li, and H. Zhang. Unsupervised extraction of visual attention objects in color images. IEEE Transactions on Circuits and Systems for Video Technology,16(1):141–145, 2006.
10. B. C. Ko and J.-Y. Nam. Object-of-interest image segmentation based on human attention and semantic region clustering. Journal of Optical Society of America A, 23(10):2462–2470, 2006.
11. Y.-F. Ma and H.-J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In ACM International Conference.
12. R. Achanta, S. Hemami, F. Estrada, and S. S ̈usstrunk. Frequency-tuned salient region detection. In CVPR, pages 1597–1604, 2009. 409, 410, 412, 413, 414, 415 on Multimedia, 2003.
13. M.M. Cheng, G.X. Zhang, N.J. Mitra, X.L. Huang and S. M. Hu, "Global Contrast based salient region detection, " in Proceedings of CVPR, pp. 409-416, Providence, RI,2011.
14. Duan, L., Wu, C., Miao, J., Qing, L., Fu, Y.: Visual Saliency Detection by Spatially Weighted Dissimilarity. IEEE CVPR, pp. 473-480 ,2011.
15. Levelthal, A., G.: The Neural Basis of Visual Function: Vision and Visual Dysfunction. Fla.: CRC Press ,1991.
16. Rajashekar, U., van der Linde, I., Bovik, A., C., Cormack, L., K.: Foveated Analysis of Image Features at Fixations. Vision Research, vol. 47, pp. 3160-3172 ,2007.
17. Wandell, B. A.: Foundations of vision. Sinauer Associates ,1995.
18. Tatler, B., W.: The Central Fixation Bias in Scene Viewing: Selecting an Optimal Viewing Position Independently of Motor Biased and Image Feature Distributions. J. Vision, vol. 7, no. 4, pp. 1-17 ,2007.
19. Zhao, Q., Koch, C.: Learning A Saliency Map Using Fixated Locations in Natural Scenes. J. Vision, vol. 11, no. 9, pp. 1-15 ,2011.
20. C. Christoudias, B. Georgescu, and P. Meer. Synergism in low level vision. IEEE Conference on Pattern Recognition,2002.Bruce, N., D., B., Tsotsos, J., K.: Saliency Based on Information Maximization. NIPS, pp. 155-162 , 2005.
21. Hou, X., Zhang, L.: Dynamic Visual Attention: Searching for Coding Length Increments.NIPS, pp. 681-688 ,2008.
22. Harel, J., Koch, C., Perona, P.: Graph-Based Visual Saliency. NIPS, pp. 545-552 ,2006.
23. R. Achanta, F. Estrada, P. Wils, and S. S ̈usstrunk. Salient region detection and segmentation. International Conference on Computer Vision Systems, 2008.