Saliency density maximization for object detection and localization

Luo, Ye; Yuan, Junsong; Xue, Ping; Tian, Qi

2011

Luo, Y., Yuan, J., Xue, P., & Tian, Q. (2011) Saliency Density Maximization for Object Detection and Localization. Proceedings of the 10th Asian Conference on Computer Vision (ACCV 2010), 396-408.

<https://hdl.handle.net/10356/81409>

https://doi.org/10.1007/978-3-642-19318-7_31

"© 2011 Springer Berlin Heidelberg This is the author created version of a work that has been peer reviewed and accepted for publication by Saliency Density Maximization for Object Detection and Localizatio, Springer Berlin Heidelberg. It incorporates referee' s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: http://dx.doi.org/10.1007/978-3-642-19318-7_31 ."

Downloaded on 26 Apr 2023 16:58:29 SGT

Saliency Density Maximization for Object Detection and Localization

Ye Luo, Junsong Yuan, Ping Xue and ^{*}Qi Tian

School of EEE, Nanyang Technological University, Singapore

^{*}Institute for Infocomm Research, Singapore

Abstract. Accurate localization of the salient object from an image is a difficult problem when the saliency map is noisy and incomplete. A fast approach to detect salient objects from images is proposed in this paper. To well balance the size of the object and the saliency it contains, the salient object detection is first formulated with the maximum saliency density on the saliency map. To obtain the global optimal solution, a branch-and-bound search algorithm is developed to speed up the detection process. Without any prior knowledge provided, the proposed method can effectively and efficiently detect salient objects from images. Extensive results on different types of saliency maps with a public dataset of five thousand images show the advantages of our approach as compared to some state-of-the-art methods.

1 Introduction

Detection of the salient object from an image has many applications in object recognition [1], image/video retargeting [2], compression [3], retrieval etc. To find the salient object, a saliency map of the image is firstly generated, where each pixel is associated with a value that indicates the importance of the pixel. Then the salient object can be detected or segmented from the saliency map.

A lot of efforts have been reported in saliency detection. However, accurate localization of the salient region or salient object from an image is still a challenging and non-trivial problem. First of all, it is not uncommon that the obtained saliency map is noisy and incomplete. As shown in Fig. 1, only several salient parts of the flower are highlighted, while the rest are missing. Due to the distraction from the cluttered background, it is not easy to find the salient region and accurately crop it out. Moreover, most existing methods apply exhaustive search for the smallest region that covers a fixed amount of fixation points [4–6], e.g. 95 % of the total salient points [4]. The major limitation is that it is difficult to predefine the amount of saliency the salient region should contain, as it depends on the size and shape of the salient object, as well as how cluttered the background is. Ideally, the salient region should be adapted to the shape of the salient object.

To address the mentioned problems, we propose a novel method to efficiently detect salient object from the saliency map. Given the saliency map, the goal is to locate a bounding box from the image that has a small size and contains most of

Fig. 1. Our salient object detection result based on the saliency map proposed in [7] (The first image is the saliency map by [7]. The second image is the binarized result of the saliency map. The third image is the result of salient object detection by searching method in [8]. The red bold rectangle is the detected result while the blue plain one is the ground truth from [4]. The last one is the result by our MSD detection method.).

the salient parts of the image. We formulate the problem as region localization with the maximum saliency density (MSD). As a new formulation of salient object detection, it balances the size of the object and the saliency it contains, and can tolerate the noise and incompleteness in the saliency map. As shown in Fig. 1, even though the salient pixels distribute sparsely, the detected saliency region with highest saliency density accurately crops the object out. Our method does not require any prior knowledge of the salient object and can automatically adapt to its size and shape through bounding box search. To avoid an exhaustive search of all possible bounding boxes of various sizes and at different locations, a branch-and-bound (B&B) search algorithm is proposed to efficiently find the global optimal bounding box.

There are several advantages of our method. First of all, it can automatically adapt to the size and shape of the salient object, despite the cluttered background. There is no need to find salient object with fixed fraction of saliency and it does not require the binary mask of the saliency map, where pixels need to be classified into salient and non-salient ones. Instead, it directly finds the bounding box of maximum saliency density from the original saliency map. Moreover, by using the branch-and-bound search, it is fast to find the optimal bounding box, e.g. in tens of milliseconds. Last but not least, our new formulation and search algorithm can be well applied to different types of saliency maps. A better performance is achieved when performed on a fused map of different types of saliency maps.

2 Related Work

2.1 Saliency map

Literally, there are two categories of computational saliency map models: local or edge/corner based [7, 9, 10, 8] and global or region based [11, 4, 12]. The first row in Fig. 3 shows several examples. The 1st and 4th columns are based on

Hou’s method [7]. The 3rd and 6th columns are from Bruce’s method [9]. And the 2nd and 5th columns are from Achanta’s [12], which generates larger visually consistent object regions than that of the previous two. In this paper, saliency map generation method is not our focus. Our main work is to detect salient objects from various saliency maps.

2.2 From saliency map to salient object

The simplest method to obtain the salient object region is by thresholding the saliency map to get a binary mask. Methods to threshold saliency map are intensively discussed in [5, 12–14]. This method is restricted on the selection of threshold and detection accuracy. In order to accurately detect salient objects from saliency maps, image segmentation result is combined with the saliency map [10, 8, 12]. However, the performance heavily relies on the accuracy of image segmentation results. Some heuristic methods [4, 15–17] are proposed to improve the performance of salient object detection. For example, exhaustive search is adopted in [4] to find the smallest rectangle window containing 95% salient pixels. Liu et al. [5] noticed the disadvantages of exhaustive search and proposed to use dynamic threshold and greedy algorithm to improve the search efficiency. However, their method is still based on thresholds and not solved by a standard optimization method. In [8], the search of the rectangle subwindow is speeded up by applying the efficient subwindow search (ESS). ESS is a recently proposed branch-and-bound search method for sliding window search [18]. It has many applications in image/video analysis [8, 18, 19].

3 The Proposed Method

Given an image I and its associated saliency map S , where $S(x, y)$ indicates the saliency value of the pixel at (x, y) , our goal is to accurately locate the salient object, i.e. to locate a salient region $W \subseteq I$. We first review existing methods then propose our new approach.

3.1 Existing schemes

Exhaustive search (ES). Some previous approaches proposed to obtain salient object regions with fixed fraction of saliency by exhaustive search from saliency maps [5, 12] or binary saliency maps [4]. We take the binary saliency map as in [4] and the formulation can be written as in Eq. 1:

$$W^* = \arg \min_{W \subseteq I} \text{area}(h(W)) \quad (1)$$

$$h(W) = \{W \mid \sum_{(x,y) \in W} S_b(x, y) \geq \lambda \sum_{(x,y) \in I} S_b(x, y)\}.$$

where $S_b(x, y)$ is the binary image of $S(x, y)$. $S_b(x, y) = 1$ when $S(x, y) \geq \tau$ and $S_b(x, y) = 0$ when $S(x, y) < \tau$. τ is the threshold and W is the subwindow of

the whole image region I . λ is the fraction threshold. The brute force method works, however, it is not time efficient and λ is heuristically decided.

Maximum saliency region (MSR). Other approaches proposed to detect the salient object by efficient subwindow search [8]. Since efficient subwindow search is based on the maximum subarray problem [18, 20], the idea of salient object detection in [8] can be formulated as in Eq. 2:

$$W^* = \arg \max_{W \subseteq I} h(W) \quad (2)$$

$$h(W) = \sum_{S_b(x,y) \in W} S_b(x,y).$$

where $S_b(x,y)$ is obtained in the same way in Eq. 1 with a slight difference that $S_b(x,y) = -1$ when $S(x,y) \leq \tau$. From Eq. 2, the salient object is located with the region W^* that contains the maximum of saliency. We call this method as the maximum saliency region (**MSR**). However, there are two major limitations of this method: (1) it highly relies on the selection of threshold τ , which is difficult to optimize; (2) when the binary saliency map is sparse, it prefers to detect a small region as shown in Fig. 1.

3.2 Our new formulation

Before giving our new formulation, we first introduce the concept of sparse and dense saliency map. Fig. 2 shows two examples. Since different saliency map generation method emphasizes different aspect, edges or corners of the salient object in Fig. 2(b) are highlighted while in Fig. 2(c) the whole salient object is popped out with uniform highlighted intensities. Therefore, we can say that the salient object in Fig. 2(b) is sparsely represented by Hou’s saliency map and it is densely represented by Achanta’s saliency map in Fig. 2(c). Sparse

Fig. 2. (a)Original image (b) Sparse saliency map example by [7] (c) Dense saliency map example by [12].

saliency map accurately detects the salient parts of the object but the boundary of the salient object is not well defined. Dense saliency map represents the salient object completely but some cluttered background is also included in the detection result. However, one thing is in common: the averaged density of the salient object region is much larger than that of any other regions on the saliency map.

To address the above characteristics and the problems in section 3.1, we propose to find the salient region W^* with the maximum saliency density from the raw saliency map $S(x, y)$. Thus we do not need to select the threshold τ and the fraction ratio λ . Moreover, it balances the size of the salient object when the saliency map is sparse. We formulate our objective function $f(W)$ as:

$$W^* = \arg \max_{W \subseteq I} f(W) \quad (3)$$

$$f(W) = \frac{\sum_{(x,y) \in W} S(x,y)}{\sum_{(x,y) \in I} S(x,y)} + \frac{\sum_{(x,y) \in W} S(x,y)}{C + \text{Area}(W)}.$$

where C is a positive constant to balance the size of $\text{Area}(W)$. The first term in $f(W)$ prefers that W contains more salient points, while the second term ensures that the detected region W is of high quality in terms of the saliency density. Therefore, by maximizing the two terms together in $f(W)$, we balance the size of the object and the saliency it contains. We call our new formulation as the maximum saliency density (**MSD**).

4 Our Algorithm

Exhaustive search of W^* from Eq. 3 is time consuming. $W^* = [T, B, L, R]$ contains four parameters, where T, B, L, R are the top, bottom, left, and right position of W^* , respectively. Suppose the frame is of size $m \times n$, the original hypotheses space is $[0, n-1] \times [0, n-1] \times [0, m-1] \times [0, m-1]$, where we need to pick up T, B, L, R from each dimension respectively. To solve this combinatorial problem, an exhaustive search is of complexity $O(m^2n^2)$. A branch-and-bound search method is proposed in [18] to accelerate the search by recursively partitioning the parameter space until it reaches the optimal solution. It shows that under certain conditions, such a branch-and-bound search can lead to the exact solution as the exhaustive search, while with a practical complexity of only $O(mn)$. The details of the branch-and-bound search can be referred to [18].

The original branch-and-bound only works for the saliency map having both positive and negative pixel values. However, in our case, the saliency map only contains positive elements and we do not want to deliberately introduce negative pixels. Therefore we need to derive our own branch-and-bound search method. Considering the efficiency of branch-and-bound search depends on the upper bound estimation, we derive the upper bound of our $f(W)$ first. Denote the set of regions by $\mathbb{W} = \{W_1, \dots, W_i\}$, where each $W_i \subseteq I$. Suppose there exists two regions W_{min} ($W_{min} \in \mathbb{W}$) and W_{max} ($W_{max} \in \mathbb{W}$), such that for any ($W \in \mathbb{W}$), $W_{min} \subseteq W \subseteq W_{max}$. Given the set $\mathbb{W}$, we denote by $\hat{f}(\mathbb{W})$ the upper bound estimation of the best solution that can find from $\mathbb{W}$. In other words, we have $\hat{f}(\mathbb{W}) \geq f(W)$, $\forall W \in \mathbb{W}$, using W_{min} and W_{max} , the upper bound can be estimated as:

$$\hat{f}(\mathbb{W}) = \frac{\sum_{(x,y) \in W_{max}} S(x,y)}{\sum_{(x,y) \in I} S(x,y)} + \frac{\sum_{(x,y) \in W_{max}} S(x,y)}{C + \text{Area}(W_{min})}. \quad (4)$$

Based on this upper bound estimation, we propose our MSD salient object detection algorithm as shown in table 1, in which the branch-and-bound procedure is similar to that of [18].

Table 1. MSD Salient Object Detection Algorithm

Require: Image saliency map $S \subseteq R^{m \times n}$
Upperbound function $\hat{f}(\mathbb{W})$ as Eq. 4
Ensure: $W^* = \arg \max_{W \subseteq S} f(W)$

Initialize P as an empty priority queue
Set $\mathbb{W} = [0, n-1] \times [0, n-1] \times [0, m-1] \times [0, m-1]$
Repeat
 Split $\mathbb{W} = \mathbb{W}_1 \cup \mathbb{W}_2$ and $\mathbb{W}_1 \cap \mathbb{W}_2 = \emptyset$
 For $i = 1$ **to** 2
 Find W_i^{min} and W_i^{max} from $\mathbb{W}_i$
 Push $(\mathbb{W}_i, \hat{f}(\mathbb{W}_i))$ into P
 End For
 Retrieve top state $\mathbb{W}$ from P
Until $\mathbb{W}$ contains only one window, e.g. $W^{min} = W^{max}$
Return $W^* = W^{min}$

5 Experimental Results

5.1 Database

In order to evaluate the results, a public dataset [4] is used to test our algorithm. The dataset provides 5000 high quality images each of which contains a salient object. Each salient object is labeled by nine users by drawing a bounding box around it. Since different users have different understanding of saliency, the point voted more than four times is considered as the salient point. The averaged saliency map S_g is then obtained from user annotation.

Given the ground truth S_g which is the binary mask of the salient object, we evaluate the performance of our method based on precision, recall, and F-measure. Suppose S_d is the salient region found by our method, the precision, recall and F-measure can be defined as:

$$pre = \frac{\sum S_g \times S_d}{\sum S_d}, rec = \frac{\sum S_g \times S_d}{\sum S_g}, F - measure = \frac{(1 + \alpha) \times pre \times rec}{\alpha \times pre + rec}. \quad (5)$$

where α is a positive constant which weights the precision over recall while calculates F-measure. We take $\alpha = 0.5$ as suggested in [8, 4].

5.2 Comparison MSD with exhaustive search

First of all, we compare our method with the exhaustive search. λ is set to 95% as [4] suggested. Fig. 3 shows the results obtained by the exhaustive search and

Fig. 3. Detection results comparison among our MSD, exhaustive search and MSR. The first row are two examples of saliency maps for Hou’s [7], Achanta’s [12] and Bruce’s [9] methods respectively. The second row are localization results by exhaustive search on the three saliency maps. The third row are results by MSR. And the last row are our MSD results. Detected results are labeled with bold red line. Blue plain rectangles are ground truth from [4].

our method on the second and the last rows respectively. As 95% is an arbitrary value and not decided based on the content of the saliency map, the detected results include a large part of the nonsalient object area. While, in our method, small salient area away from main salient object region is dropped under the constraint of the saliency density. Therefore, our result is more accurate than the exhaustive search. Performances of precision, recall and F-measure on Fig. 5 further validate our claim. In order to balance the bias caused by arbitrarily choosing λ , we test four different λ values as Fig. 4 shows. Precision is improved while recall is reduced as λ becomes smaller and there is no direct way to choose an optimal λ .

5.3 Comparison MSD with MSR

To compare our MSD with MSR, we test both of the methods on three different saliency maps. The threshold τ is obtained by Otsu [22] for all saliency maps in MSR. C is set to be 60625, 2025 and 16200 for Hou’s, Bruce’s and Achanta’s saliency map respectively in MSD, through parameter evaluation. Fig. 5 shows the comparison results. The average of precision, recall and F-measure are reported on each group. On Hou’s saliency map, edges/corners are detected as salient parts. By using MSR, very small region is bounded while larger salient regions are detected by our MSD. The F-measure and recall are significantly improved by our MSD. For the other two saliency maps, our MSD also outperforms MSR. The results on three different types of saliency maps show that our

Fig. 4. Precision, recall and F-measure for exhaustive search with four λ {95%, 90%, 85%, 80%} on Hou's saliency map.

Fig. 5. Comparisons precision, recall and F-measure for our MSD to exhaustive search (ES) and MSR on (a) Hou's saliency map, (b) Bruce's saliency map and (c) Achanta's saliency map.

method improves the F-measure, and at the same time, keeps the high precision rate.

5.4 Evaluation on different saliency maps

Since the salient object detection result is based on the saliency map, the more accurate the saliency detection is the better performance of the object detection method obtains. It is worth noting that a single salient object region in [4] is obtained through supervised learning. Both [10] and [8] have prior knowledge about the region size provided by image segmentation. Thus, they are not directly comparable with our method. However, even without any prior knowledge of the salient object, our method on Bruce's saliency map outperforms Ma's method [16] which directly uses detected salient region (F-measure 61%) and search result on Itti's saliency map [21] which finds the minimum rectangle containing 95% salient points by the exhaustive search (F-measure 69%). For our method on Hou's saliency map, it obtains comparable result compared with

Fig. 6. Comparison our MSD to other salient object detection results by precision, recall and F-measure. 1: Ma’s saliency map and their salient region detection result [16]; 2: Exhaustive search smallest subwindow containing 95% salient points from Itti’s saliency map [21]; 3: MSD on Hou’s saliency map; 4: MSD on Bruce’s saliency map; 5: MSD on Achanta’s saliency map; 6: MSD on the combined saliency map.

the exhaustive search on Itti’s saliency map. Though our result on Achanta’s saliency map is not as good as the result on [21], the precision is still higher than searching result on Itti’s and Ma’s saliency maps.

Since different bottom-up saliency map generation method has different advantages and disadvantages, to minimize the influence to the saliency object detection result, three previous saliency maps are fused together. Each saliency map is normalized into $[0, 1]$ and the combination saliency map is obtained by adding them together then normalizing the summation into $[0, 1]$. As shown in Fig. 6 method 6, after combining three saliency maps together, F-measure is 74.67% which is 1.61% larger than the optimal result (73.06%) from Bruce’s saliency map. This performance is comparable to the learning based salient object detection results e.g. [4, 8] but our method is much simple and time efficient than them.

5.5 Parameter evaluation

To evaluate the influence of the only parameter C in our MSD method, different values C are tested as shown in Fig. 7. When C is small, the method is sensitive to the density change and prone to converge to a region with higher average density but relative smaller size. When a large value of C is selected, density term becomes trivial in objective function $f(W)$ and the whole algorithm converges to a larger region with lower average density. In Fig. 7(a), within the range of $[35000, 84000]$, it is thus not sensitive to the selection of C . Similarly, when C is in the range $[1500, 3300]$, the F-measure is above 71.3% in Fig. 7(b); when C in the range $[14200, 23500]$, the F-measure is above 63.5% in Fig. 7(c). From these results, we can see that the region based saliency map has a smaller optimal C

Fig. 7. Optimal C-value range for MSD on (a) Hou's saliency map, (b) Bruce's saliency map, (c) Achanta's saliency map and (d) Combined saliency map (x-coordinate is C-value measured in unit 10^4 and y-coordinate is the corresponding averaged F-measure.).

than edge/corner based methods (As Fig. 3 shows Bruce's saliency map is denser than Achanta's.). That further indicates that density term in Eq. 3 is important when the salient points are densely distributed on the saliency map.

5.6 Time complexity

The average computational time tested on 5000 images for exhaustive search, MSR and our method based on Hou's, Bruce's and Achanta's saliency map are shown in table 2. It is obvious from table 2 that our method is very time efficient compared to the exhaustive search and has comparable time efficiency to MSR. The algorithm is tested on a Duo Core desktop of 2.66GHz, implemented with C++.

6 Conclusion

We propose in this paper a novel method to efficiently detect salient objects from images. Salient object detection is first formulated with the saliency density. A branch-and-bound search algorithm is developed to optimize the newly formulated problem globally. Without a prior knowledge of the salient object,

Table 2. Time complexity comparison by seconds.

Method	saliency map by [7]	saliency map by [9]	saliency map by [12]
Exhaustive Search	22.2359	22.1646	23.3587
MSR	0.0039	0.0048	0.0056
Our MSD	0.0113	0.0351	3.4718

Fig. 8. More salient object localization results by our MSD method (The first row are our results on Hou’s saliency map [7]. The second row are our results on Achanta’s saliency map [12] and the last row are based on Bruce’s saliency map [9]. The red bold rectangle is the detected result while the blue plain one is the ground truth from [4].).

our method can adapt to different sizes and shapes of the object, and is less sensitive to the cluttered background. The experiments on a public dataset of 5000 images show that our method greatly improves the existing baseline methods on the measurements of precision, recall and F-measure. Our method gains comparable performance compared to learning based salient object detection results with a high time efficiency. Tests on different saliency maps indicate our method works well with different types of saliency maps. Our future work includes the localization of multiple salient objects in images and content based image and video retargeting.

References

1. Gao, D., Vasoncelos, N.: Discriminant saliency for visual recognition from cluttered scenes. In: Advances in Neural Information Processing Systems. (2004) 481–488
2. Liu, F., Gleicher, M.: Video retargeting: Automating pan and scan. In: Proc. ACM Multimedia. (2006) 241–250
3. Bradley, A.P., Stentiford, F.W.: Visual attention for region of interest coding in jpeg 2000. Journal of Visual Communication and Image Representation **14** (2003) 232–250

4. Liu, T., Sun, J., Zheng, N., Tang, X., Shum, H.: Learning to detect a salient object. In: IEEE Conf. on Computer Vision and Pattern Recognition. (2007) 1–8
5. Liu, F., Gleicher, M.: Automatic image retargeting with fisheye-view warping. In: Proc. of ACM symposium on User interface software and technology. (2005) 153–162
6. Suh, B., Ling, H., Bederson, B.B., Jacobs, D.W.: Automatic thumbnail cropping and its effectiveness. In: Proc. of ACM symposium on User interface software and technology. (2003) 95–104
7. Hou, X., Zhang, L.: Dynamic visual attention: Searching for coding length increments. In: Neural Information Processing Systems. (2008) 681–688
8. Valenti, R., Sebe, N., Gevers, T.: Image saliency by isocentric curvedness and color. In: IEEE Intl. Conf. on Computer Vision. (2009)
9. Bruce, N., Tsotsos, J.: Saliency based on information maximization. In: Neural Information Processing Systems. (2005) 155–162
10. Wang, Z., Li, B.: A two-stage approach to saliency detection in images. In: IEEE Intl. Conf. on Acoustics, Speech and Signal Processing. (2008) 965–968
11. Miyazato, K., Kimura, A., Takagi, S., Yamato, J.: Real-time estimation of human visual attention with dynamic bayesian network and mcmc-based particle filter. In: IEEE Intl. conf. on Multimedia and Expo. (2009) 250–257
12. Achanta, R., Hemami, S., Estraday, F., Susstrunk, S.: Frequency-tuned salient region detection. In: IEEE Conf. on Computer Vision and Pattern Recognition. (2009) 1597–1604
13. Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. In: IEEE Conf. on Computer Vision and Pattern Recognition. (2007) 1–8
14. Viswanath Gopalakrishnan, Y.H., Rajan, D.: Salient region detection by modeling distributions of color and orientation. IEEE Trans. on Multimedia **11** (2009) 892–905
15. Junwei Han, Ngan King N., M.L., Zhang, H.J.: Unsupervised extraction of visual attention objects in color images. IEEE Trans. on Circuits and Systems for Video Technology **16** (2006) 141–145
16. Ma, Y.F., Zhang, H.J.: Contrast-based image attention analysis by using fuzzy growing. In: Proc. ACM Multimedia. (2003) 374–381
17. Ko, B.C., Nam, J.Y.: Object-of-interest image segmentation based on human attention and semantic region clustering. Virtual Journal for Biomedical Optics **1** (2006) 2462–2470
18. Lampert, C.H., Blaschko, M.B., Hofmann, T.: Efficient subwindow search: A branch and bound framework for object localization. IEEE Trans. on Pattern Analysis and Machine Intelligence **31** (2009) 2129–2142
19. An, S., Peursum, P., Liu, W., Venkatesh, S.: Efficient algorithms for subwindow search in object detection and localization. Volume 0. (2009) 264–271
20. Lawler, E.L., Wood, D.E.: Branch-and-bound methods: A survey. Operations Research **14** (1966) 699–719
21. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence **20** (1998) 1254–1259
22. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. on System, Man and Cybernetics **9** (1979) 62–66