Salient Object Segmentation using a Switch Scheme

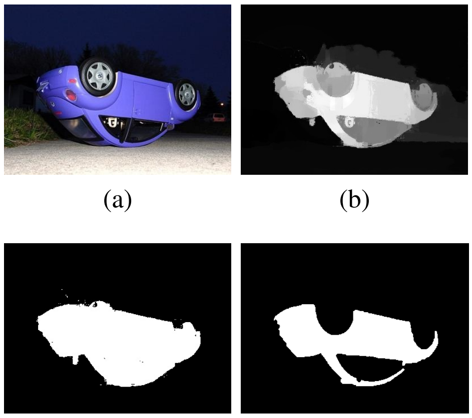
Ran SHI* , King Ngi NGAN* and Songnan LI*
* Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong

Abstract—**In this paper, we propose a novel switch scheme and a saliency map binarization method for salient object segmentation. With the proposed switch scheme, the saliency map can be segmented by different methods according to its quality, which is evaluated by a method proposed in this paper. We also develop a binarization method by integrating three properties of the salient object. This method exclusively derives information from the saliency map (i.e., without referring to the original image). Experimental results demonstrate that the proposed binarization method can generate better segmentation results and the switch scheme can further improve the segmentation results by fully exploiting the merit of both segmentation methods.**

I. Introduction

Salient object segmentation aims at extracting objects which capture most perceptual attention from images. It is a subsequent application of saliency maps and plays an important role for many content-based applications, such as content-based retargeting [1], content-based compression [2] and content-based retrieval [3], etc. In the past decades, researchers have made great efforts on developing saliency models and constantly improving the saliency map' s quality. The improvement of saliency models also gives impetus to the development of salient object segmentation.

According to the use of saliency maps, salient object segmentation methods can be classified into two categories. One is that segmentation methods are only based on the saliency map. They do not derive any information from original images. Two typical methods are Adaptive Threshold [4] and OTSU [5]. Adaptive Threshold selected twice the mean of the saliency map as the optimal threshold to separate salient objects and background. OTSU as a classic binarization method searched for an optimal threshold by maximizing the dissimilarity between salient objects and the background. Both methods just treat salient object segmentation as a general binarization problem of a gray image and adopt a global measure to select the optimal threshold. They do not fully explore properties of salient objects. The other category extracts some information from saliency map and then return to the original image to segment the salient object. These methods do not solely rely on saliency map but also exploit information from original images. The typical one is SaliencyCut [6]. It extracted pixels as object seeds and background seeds from saliency maps by empirical thresholds, which then were used to construct object and background color models in the original image. Finally, Grabcut [7] algorithm was adopted to generate the salient object. In this method, the ultimate criterion for judging whether a pixel belonging to the salient object is its color similarity with the seeds. However, the saliency map is generated by integrating many cues based on characteristics of the human visual system rather than the color information only. Therefore, this method eventually discards many useful cues within the saliency map. One example is shown in Fig.1. We can see that the car has been highlighted enough in the saliency map. It is suitable to be segmented using the threshold based binarization methods. However, since the color of wheels and window is similar to that of the background, they are labeled as background by SaliencyCut. So SaliencyCut may generate a poor segmentation result even with a saliency map of good quality. Therefore, the methods relying solely on the saliency map should be the first choice when the saliency map is good for segmentation. However, this idea introduces a new problem about how to evaluate the quality of a saliency map for the salient object segmentation. In [8], six features (saliency coverage, saliency map compactness, saliency histogram, color separation, segmentation quality and boundary quality) were combined to the evaluate quality of saliency maps without ground truth. It only trained a binary classifier to compare relative quality of pairs of saliency maps. Although it cannot judge whether a saliency map is suitable for the segmentation, the proposed features are worthy of our reference.



(a) (b) (c) (d)

Fig. 1: Salient object segmentation results by different methods. (a) Original image [9]; (b) Saliency map [10]; (c) Result by our binarization method (d) Result by SaliencyCut.

Based on the analysis above, we solve the salient object segmentation problem using a switch scheme in this paper. Given a saliency map, we propose an assessment to evaluate its quality according to saliency compactness and distribution of its histogram. If the saliency map is good, then we use a threshold based binarization method proposed in this paper to segment the salient object. It can select the best result from possible candidates by measuring saliency occupation ratio, saliency contrast and boundary smoothness. Otherwise, SaliencyCut is adopted as an auxiliary method to tentatively segment salient objects when the saliency map has poor quality. The rest of this paper is organized as follows. Section 2 describes our salient segmentation method in detail. Experimental results are presented in Section 3. Finally, Section 4 concludes our paper.

II. Proposed Method

A. The Evaluation of the Saliency Map

The first step of our method is to judge whether the binarization of the saliency map is suitable using a threshold. Examples are shown in Fig.2. For the saliency map (a), all high saliency values concentrate on the object. It is suitable to segment this object by searching for an optimal threshold. However, in saliency map (c) the distribution of saliency values is more scattered. The object is partially highlighted and some background regions share similar saliency values with the object. It is difficult for threshold based methods to well segment the object. Therefore, the compactness of high saliency values is an important attribute for our judgment. In [8], the compactness degree was measured by the density of the minimal image window covering a fixed proportion of the total saliency. However, the density is a relative index which may not correlate well with the saliency compactness. In our method the highest compact rectangle region W^* of large saliency values is searched by the method formulated as:

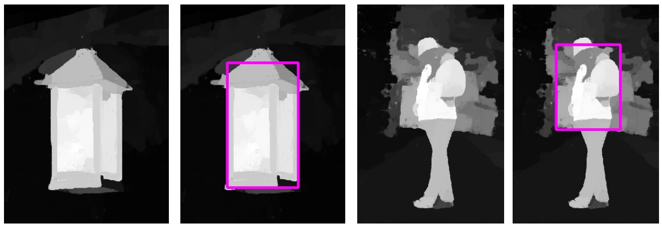
$$W^* = \underset{W \subseteq S_{map}}{\arg\max} f(W) \tag{1}$$

$$f(W) = \sum_{p \in W} s'_p \tag{2}$$

$$s'_p = s_p - \mu(1 + bias) \tag{3}$$

$$bias = \sum_{k=1}^{16} (H_d(k) \cdot \frac{|index(k) - 128|}{128}) \tag{4}$$

where W represents one rectangle region in the saliency map S_{map}, and s_p is the saliency value of p. s'_p is generated by Eqs.(3) and (4) where μ is the mean of the whole saliency map and H_d is the histogram of the saliency map by descending order. $index(k)$ indicates the original index of the kth bin before the sorting. We adopt ESS [11] to effectively search for W^*. Since s'_p may be positive or negative, a high compact region should include positive values as much as possible and exclude negative values. W^* actually locate the region with maximum of net saliency value. However, "high value" itself is a relative definition. So we adjust s'_p by the mean and the histogram. The mean is treated as a base. If saliency values mostly distribute at the two ends of the histogram, the $bias$ should be larger to make the threshold of "high value" much higher. Thus, we sort the histogram by descending order and use the summation of top 16 bins to measure the concentration degree of the distribution. Meanwhile, the bins locating closer to the two ends are assigned higher weights. In Fig.2 (b) and (d), the highest compact region of each saliency map is labeled by the pink bounding box. We can clearly observe the



(a) (b) (c) (d)

Fig. 2: Saliency maps [10] with different quality and the most compact regions. (a) An exemplar saliency map with good quality; (b) The highest compact region in (a) indicated by a pink bounding box; (c) An exemplar saliency map with poor quality; (d) The highest compact region in (c) indicated by a pink bounding box.

difference between these two bounding boxes. Due to high compactness of Fig.2 (a), the bounding box can cover most part of the object. However, the pink bounding box can only cover the upper body of the person in Fig.2 (d) due to the low compactness of Fig.2 (c). Furthermore, the compactness degree of W^* is measured by:

$$C = \frac{\sum_{p \in W^*} s_p}{\sum_{q \in S_{map}} s_q} \tag{5}$$

The larger C indicates higher compactness degree. Different from [8], we use the saliency ratio covered by W^* as an absolute index to measure the compactness. For Fig.2 (a) and (c), their compactness degrees are 0.83 and 0.47 respectively. Thus, we can set a threshold T_C to distinguish them.

B. Saliency Map Binarizaion

Since the saliency map is a gray image, we treat salient object segmentation as binarizaiton by an optimal threshold. Therefore, there are 255 possible thresholds corresponding to 255 possible candidate segmentation results (We exclude a case that the whole image is a salient object). Then, the best result can be selected from these candidates. Based on observation, properties of a salient object in the saliency map can be summarized as follows:

1) The salient object should occupy a high portion of all saliency values.
2) The salient object should have a high contrast with surrounding regions.
3) The salient object should have a smooth boundary.

For a certain threshold i, its corresponding result is R_i. We evaluate the quality of R_i in terms of these three properties. For the first property, we calculate the portion of saliency that is occupied by R_i:

$$Ratio(R_i) = \frac{\sum_{p \in R_i} s_p}{\sum_{q \in S_{map}} s_q} \tag{6}$$

For the second property, the contrast is measured as:

$$Contrast(R_i) = \frac{1}{L_{B_i}} \sum_{p \in B_i} (\frac{s_p}{M_{\mathcal{N}_p}} \sum_{q \in \mathcal{N}_p} (s_p - s_q)) \tag{7}$$

where p is a pixel on R_i' s boundary B_i. L_{B_i} is the length of B_i. $\mathcal{N}_p$ is a set of p' s four-neighboring background pixels. $M_{\mathcal{N}_p}$ indicates the number of elements in this set. We measure the contrast with surrounding regions by calculating average saliency difference between boundary pixels and their neighboring background pixels. We also adopt saliency value s_p as a weight. Because saliency value itself indicates the possibility of one pixel belonging to the salient object. It means the contrast should be enhanced for boundary pixels with high saliency values.

For a poor segmentation result, its rough boundary usually exhibits an irregular shape with a lot of ragged parts. In order to improve the quality, "Opening" and "Closing" are two useful morphologic operations to remove the ragged parts of the boundary and keep the smooth parts. According to the observation above, we can measure the smoothness in terms of similarity between the boundaries before and after the "Opening" and "Closing" operations. One side effect of this measurement is that the sizes of ragged parts are correlated to different sizes of objects. In order to relieve this side effect, the mask size of "Opening" and "Closing" operations should be adapted to the size of R_i. We generate one set about boundary that is $\Omega_i = \{x | x \in B_{R_i \circ m_i} \cap B_{R_i \bullet m_i} \cap B_i\}$, where $\circ$ and $\bullet$ denote the "Opening" and "Closing" operations, respectively. m_i is a square operation mask whose length of the side is βL_{R_i} where β is set to 0.004 and L_{R_i} is the length of the diagonal of R_i' s bounding box. The smoothness is measured:

$$Smooth(R_i) = \frac{Card(\Omega_i)}{Card(B_i)} \tag{8}$$

where $Card(\cdot)$ is an operation for counting the number of elements in one set. From Eq.(8), we can see that larger value indicates smoother boundary since the boundary is not much affected after two morphologic operations.

Finally, we combine these three properties together to select the best segmentation result R^*.

$$R^* = \underset{i \in [1,\ldots,255]}{\arg\max}\ Ratio(R_i) \cdot Contrast(R_i) \cdot Smooth(R_i) \tag{9}$$

In Fig.3 from the top to the bottom row, we successively discard Ratio term, Contrast term or Smooth term, but keep the other two terms in each case to select the best result. The comparison results reflect that each property has its own function for selecting the best result. All three properties should be taken into account in the binarization method.

C. The Switch Scheme

According to the quality of the saliency map evaluated by C, we design a switch scheme to designate one suitable method for segmenting it. Specifically, the saliency map with C above the threshold T_C is segmented by our binarization method. Otherwise, it is segmented by SaliencyCut.

III. Experimental Results

In order to evaluate the segmentation performance, we test our method on the ASD database [4] which contains 1,000 images selected from the MSRA database [9] for salient objects. The pixel-wise binary ground truths are provided by [10]. We use the ST model [10] to generate saliency maps. For each segmentation result, we adopt $F_1measure$ to evaluate its accuracy compared with its corresponding ground truth. $F_1measure$ is the harmonic mean of precision and recall where precision and recall share the equal weight.

$$F_1measure = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{10}$$

We tested the performance of our method (the switch scheme + the binarization method) with four different T_C, i.e. 0.7, 0.75, 0.8 and 0.85. We also compared them with the adaptive threshold method (AT) [4], OTSU [5], SaliencyCut [6] (SC) and our binarization method only (OUR-B). The average $F_1measure$ of 1000 results indicates the overall performance of these methods. The higher average indicates the better performance. The performances of these five methods are shown in Fig.4. Obviously, salient object segmentation result depends on the quality of the saliency map. Due to the good performance of ST model which guarantees high quality of the saliency map, we can see that simple segmentation methods such as AT and OTSU can also achieve good performance. The result also demonstrates that SC is still a high competitive segmentation method. The main reason is that on one hand, color dissimilarity itself is an useful feature to distinguish the salient object from the background. On the other hand, Grab-Cut can effectively utilize the partial information provided by the saliency map to extract the salient object. When the quality of the saliency map is poor, we can still attempt to segment the salient object by SC. Our binarization method achieves better performance compared with those of AT and OTSU. More comparison results are shown in Fig.5. In comparison to the results generated by AT and OTSU, our results are more precise and complete. It demonstrates that the three properties of the salient object considered in our method are reasonable and the proposed measurements can describe these properties

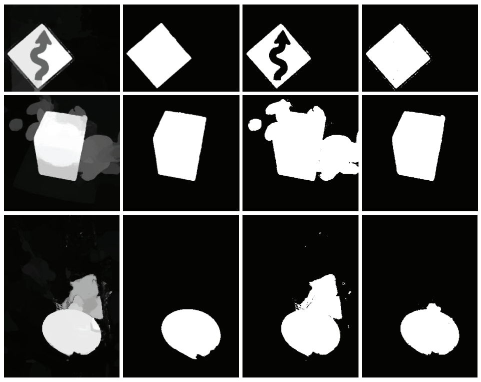


Fig. 3: Segmentation results by our method with or without evaluating certain property. The images from the left column to the right column are saliency maps [10], ground truths [10], results by our method without evaluating one selected property (from the top to the bottom discarding Ratio, Contrast or Smooth term) and results with evaluating all three properties.

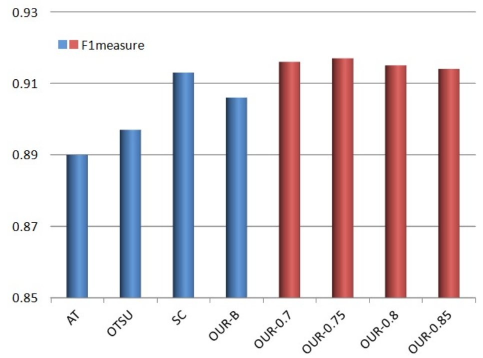


Fig. 4: Performance of different salient object segmentation methods. The performances of our method with different T_C are marked as red; others are marked as blue.

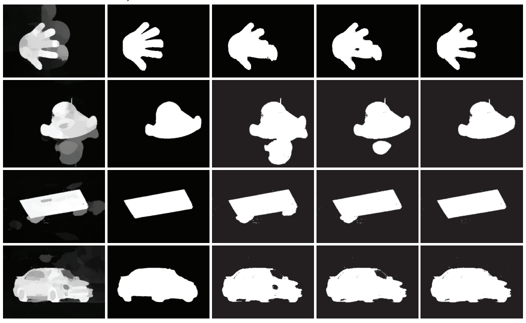


Fig. 5: Examples of segmentation results. The images from the left column to the right column are saliency maps, ground truths, results by AT, results by OTSU and results by OUR-B.

well for the good saliency maps. However, since OUR-B cannot well segment the saliency maps with poor quality where the properties cannot be held, its performance is lower than that of SC. When we combine our binarization method and SC together using the switch scheme with different T_C, better performances can be achieved. It demonstrates that the proposed measurement is not sensitive to the T_C, and it can effectively judge whether the saliency map is suitable for our binarization method. The best performance can be achieved when T_C is set to 0.75, in which case 563 saliency maps are segmented by our binarization method and the others are segmented by SC. More results of Our-0.75 are shown in Fig.6. The result marked by a red star is the one generated by the designated segmentation method. We can see that if our method mistakenly designates a segmentation method, the worse result will be generated. According to the evaluation of the proposed measurement, better segmentation results can be generated by designating a suitable segmentation method.

IV. Conclusion

The core idea of this paper is that salient object segmentation should solely rely on the saliency map when its quality is good. Therefore, a measurement of saliency compactness degree is proposed for evaluating the quality of the saliency map. The switch scheme based on this measurement is used to fully take the merits from our binarization method and SaliencyCut so as to generate better segmentation results.

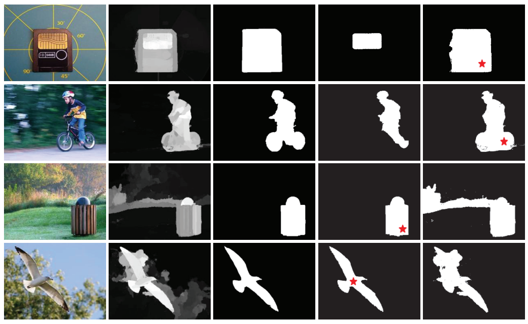


Fig. 6: Examples of segmentation results. The images from the left column to the right column are original images, saliency maps, ground truths, results by SC and results by OUR-B. The red star indicates the result generated by the designated method.

Our binarization method searches for an optimal threshold on the saliency map by quantifying three properties of the salient object including high saliency occupation ratio, high local contrast and smooth boundary. It can guarantee good segmentation results when it is designated to conduct the segmentation.

Acknowledgement

This work is partially supported by a grant from the Research Grants Council of the Hong Kong SAR, China (Project CUHK 14201115).

References

[1] Ariel Shamir and Shai Avidan, “Seam carving for media retargeting,” *Communications of the ACM*, vol. 52, no. 1, pp. 77–85, 2009.

[2] Chenlei Guo and Liming Zhang, “A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression,” *Image Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 185–198, 2010.

[3] Hong Fu, Zheru Chi, and Dagan Feng, “Attention-driven image interpretation with application to image retrieval,” *Pattern Recognition*, vol. 39, no. 9, pp. 1604–1621, 2006.

[4] Ravi Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk, “Frequency-tuned salient region detection,” in *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*, 2009, pp. 1597–1604.

[5] Nobuyuki Otsu, “A threshold selection method from gray-level histograms,” *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[6] Ming Cheng, Niloy J Mitra, Xumin Huang, Philip HS Torr, and Song Hu, “Global contrast based salient region detection,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, no. 3, pp. 569–582, 2015.

[7] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts,” *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 309–314, 2004.

[8] Long Mai and Feng Liu, “Comparing salient object detection results without ground truth,” in *Computer Vision–ECCV 2014*, pp. 76–91. 2014.

[9] Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, Xiaoou Tang, and Heung-Yeung Shum, “Learning to detect a salient object,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 2, pp. 353–367, 2011.

[10] Zhi Liu, Wenbin Zou, and Olivier Le Meur, “Saliency tree: A novel saliency detection framework,” *Image Processing, IEEE Transactions on*, vol. 23, no. 5, pp. 1937–1952, 2014.

[11] Christoph H Lampert, Matthew B Blaschko, and Thomas Hofmann, “Efficient subwindow search: A branch and bound framework for object localization,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 12, pp. 2129–2142, 2009.