

Adaptive Location for Multiple Salient Objects Detection

Shaoyong Jia¹, Yuding Liang¹, Xianyang Chen¹, Yun Gu¹, Jie Yang¹,
Nikola Kasabov², and Yu Qiao¹✉

¹ Institute of Image Processing and Pattern Recognition,
Shanghai Jiao Tong University, Shanghai, China
{jiashaoyong, liangyuding, st_tommy, geron762, jieyang, qiaoyu}@sjtu.edu.cn

² Knowledge Engineering and Discovery Research Institute,
Auckland University of Technology, Auckland, New Zealand
nkasabov@aut.ac.nz

Abstract. Salient objects detection aims to locate objects that capture human attention within images. Recent progresses in saliency detection have exploited the center prior, to combine with other cues such as background information, object size or region contrast, achieving competitive results. However, previous approaches of center prior supposing salient object locates nearly at image center is very simple, fragile, especially not suitable for multiple salient objects detection, but the assumption is mostly heuristic. In this paper, we present an adaptive location method based on geodesic filtering framework to address these issues. First, we detect salient points by the adjustive color Harris algorithm. Second, we involve the Affinity Propagation (AP) method to automatically cluster the salient points for a coarse objects location. Then, we utilize geodesic filtering framework for a final saliency map by multiplying objects location and size. Experimental results on two more challenging databases of off-center and multiple salient objects demonstrate our approach is more robust to the location variations of salient objects, against state-of-the-art methods for saliency detection.

Keywords: Saliency detection · Adaptive location · AP cluster method

1 Introduction

Recent years have seen greatly increasing interest in salient object detection [3]. It is motivated by the importance of saliency detection in applications such as object detection and recognition [10], image segmentation [6], image and video compression [5] and visual tracking [11]. Because of the loss of high level knowledge, all bottom up methods depend on assumptions about the properties of objects and backgrounds. Among them, some researches usually add a Gaussian map to models for center prior to enhance saliency computation in [7, 17], which suppose that salient objects locate closely at the image center.

We observed two issues about previous assumptions of center prior. The first is to simply treat the image center as potential salient object location which ignores the fact that multiple salient objects are off-center at different levels. This is fragile and may fail on more challenging databases, such as the SED2 [2] of which the images contain multiple off-centered salient objects. In this case, image center assumption may become the bottleneck when it is integrated with other cues for saliency detection. Secondly, while some methods re-estimate the mean and radius of the gaussian map from an initial saliency map, this strategy is still not suitable for multiple off-center salient objects.

In this paper, we present an adaptive location method to address the above two problems. Our main contribution is a novel and reliable salient objects location measure, called *adaptive location*. Instead of simply assuming salient object locating at the image center, the proposed method aims to automatically detect the salient objects location. Our method is more robust as it characterizes the spatial layout of salient objects. In detail, we firstly detect salient points and cluster them by AP algorithm [4]. Then we utilize the *geodesic filtering* framework and “soft” region size computing method proposed in [17] for a final saliency map.

We enhance the baseline proposed in [17]. Since images in SED2 [2] have only two salient objects and simple background, we select the images of multiple objects against more complex background from DUT-OMRON [16] and



Fig. 1. Saliency detection results on challenging examples. (a) Input images; (b) GS [13]; (c) SF [9]; (d) Base [17]; (e) Ours; (f) ground truth.

PASCAL-S [8], called dataset MDUT-SAL, to further validate our algorithm. Experimental comparisons show that our approach outperforms Base [17], especially on MDUT-SAL. The examples in Fig. 1 show that comparisons against other methods of different difficulties: background interference, small salient object, background touching, three and four saliency objects. More importantly, the performance of all previous methods are further improved with our results combined than Base [17], and new state-of-the-art results are achieved.

2 Geodesic Filtering Framework

The papers [17, 18] both proposed geodesic filtering framework based on a regular superpixel image representation, which encodes the information of image segmentation in an implicit and soft manner.

Firstly, an image is converted into CIELab color space and decomposed into N superpixels representation by SLIC algorithm [1]. Then an undirected weighted graph is constructed by connecting spatially adjacent superpixels. The Euclidean distance between superpixels i and j is denoted as the edge weight $w_{i,j}$ according to average colors of superpixels. The *geodesic distance* between any two superpixels G_d is computed as:

$$G_d(i, j) = \min_{i=v_1, v_2, \dots, v_n=j} \sum_{k=1}^{n-1} w_{v_k, v_{k+1}} \quad (1)$$

where v_1, v_2, \dots, v_n is a shortest path in the graph linking nodes i and j , and $G_d(i, i)$ is set to 0. Then the *geodesic connectivity* is defined as:

$$G_c(i, j) = \exp\left(-\frac{G_d^2(i, j)}{2\sigma^2}\right) \quad (2)$$

Secondly, the *geodesic filtering* framework is defined to measure the properties of image regions from superpixels representation. Suppose $I(j)$ is the property value of superpixel j to be filtered, the geodesic filtering computes the property of the region that superpixels j belongs to as:

$$GF(I, j) = \frac{\sum_{j=1}^N G_c(i, j) \times I(j)}{\sum_{j=1}^N G_c i, j} \quad (3)$$

It aggregates and smoothes the property values within the same homogeneous region. After filtering, all superpixels in the same region have similar property values of that region. As proposed in [17], Eq. (3) is used to estimate salient object centerness by replacing $I(j)$ with a gaussian map $M(j)$, which is too simple and weak for multiple salient objects detection. We propose our approach in Sect. 3 to alleviate this problem. And an un-normalized version of GF by removing the denominator is used to estimate the object size.

3 Our Approach

Many saliency methods are biased to assign image center regions with higher saliency. However, previous methods simply use a gaussian fall-off map with mean at the image center and a fixed radius, or re-estimate the mean and radius of the gaussian map from an initial saliency map which highly depends on the quality of the initial saliency map. These strategies are problematic for multiple salient objects.

We propose a method which can detect the salient objects location automatically, which characterizes the spatial layout of salient objects. We follow below five steps to implement our algorithm with enough motivation in detail.

Image smoothing: some image background or noise may be so complex that they affect subsequent salient points detection. We smooth images firstly via L0 gradient minimization [15] which can remove low-amplitude structures and globally preserve and enhance salient edges. The salient points detection and clustering results before and after smoothing are shown in Fig. 2(d) and (e). We can see that the salient points coming from background are eliminated and the cluster center locates at the object center basically after smoothing the original input images.

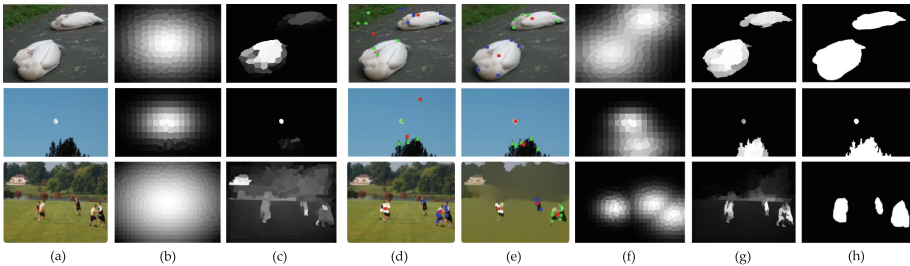


Fig. 2. Illustration of our approach. (a) Input images; (b) gaussian maps [17]; (c) final saliency maps [17]; (d) salient points without image smoothing and corresponding cluster center (red); (e) salient points with image smoothing and corresponding center (red); (f) our gaussian maps; (g) our final results; (h) ground truth (Color figure online).

Salient points detection: traditional luminance-based saliency detection methods incline to completely ignore the color information and thus are very sensitive to the background noises. In [12], they applied the boosting color saliency theory to Harris detector and show that the resulting saliency points are much more informative than the luminance-based Harris points.

In this paper, we adopt the color boosting Harris points [12] as salient points to catch the corners or marginal points of visual salient region in color image and eliminate the salient points near image boundary. Then the saliency points provide us a coarse location of the salient areas even if there are multiple salient

objects. As the color boosting Harris points usually gather around the saliency region, the salient points center usually locates at the object center. We denote the salient points as SP_k , $k = 1, 2, \dots, K$, where K is the number of salient points. Besides, it is good for subsequent clustering to locate salient objects adaptively. Note that even though few salient points from background noises do not make an obvious negative effect on cluster center even the final saliency map.

Adaptive location: In [14], they proposed the concept of convex hull derived from salient points and adopted k-means method to group superpixels inside and outside the convex hull for eliminating the effect of the noisy region included in the convex hull based on Bayesian model. However, they are simply used for single salient object, which is quite different with ours.

We adopt the AP method to cluster K salient points into l clusters, which is basically consistent with the number of salient objects, with m salient points respectively, represented as $SP_j^i = \{X_j^i, Y_j^i\}$, where $j = 1, 2, \dots, m$, $i = 1, 2, \dots, l$, namely:

$$K = \sum_{i=1}^{i=l} SP_m^i. \quad (4)$$

Then we calculate the center of each cluster, namely *adaptive location*, the average of spacial positions following below formula:

$$C_i = \frac{1}{m} \sum_{j=1}^{j=m} SP_j^i = \frac{1}{m} \sum_{j=1}^{j=m} \{X_j^i, Y_j^i\}. \quad (5)$$

And we define the cluster radius R_i as the average Euclidean distance between each salient point and corresponding cluster center:

$$R_i = \frac{1}{m} \sum_{j=1}^{j=m} \|SP_j^i - C_i\|^2. \quad (6)$$

Then we get a gaussian fall-off map G by combining R_i , as shown in Fig. 2(f), with mean at cluster center and standard deviation equals to its cluster radius for each cluster. Note that we add a small constant value to cluster radius to avoid the degenerate case when they are equal to 0.

Final saliency map: we replace I with G in Eq. 3 to acquire a saliency map based on our *adaptive location*. Then we completely follow the background prior and approximate computation of region size in [17] for final saliency map, shown in Fig. 2(g), which are much better than the Base [17] results in Fig. 2(c). This fully shows that we further optimize the proposed method in [17].

4 Experiments

For experimental comparison, we use a standard benchmark dataset SED2 [2] which contains 100 images of two salient objects with largely different sizes and

locations while background is relatively simple, and our more challenging MDUT-SAL, consisting of 220 images with multiple salient objects and complex background by combining most examples in DUT-OMRON [16] and PASCAL-S [8]. We follow [17] to compute the standard precision-recall curves (PR curves) and F-measures evaluation metrics. As complementary, we also introduce the mean absolute error (MAE) into the evaluation which measures how close a saliency map is to the ground truth.

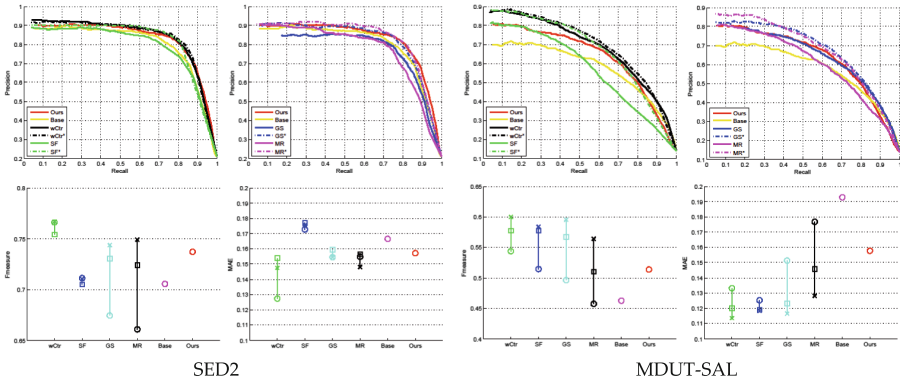


Fig. 3. (Better viewed in color) Precision-recall curves, F-measure and MAE of various methods on SED2 [2] (left) and MDUT-SAL. In the PR curves, results of dotted lines and (*) are obtained by combining our results. In the F-measure and MAE, the circle marker is the value of some state-of-the-art methods, square and cross markers are the results after combing with Base and ours, respectively (Color figure online).

We compare against the most recent state-of-the-art saliency methods, including saliency filter (SF) [9], manifold ranking (MR) [16], geodesic saliency (GS) [13], and saliency optimization (wCtr) [18]. All of them implemented algorithms based on SLIC [1] superpixels and achieved competitive results in recent years. Example results of recent state-of-the-art original results, after combining Base [17] and our approach are shown in Fig. 4.

4.1 Comparison with State-of-the-Art

Figure 3 reports the PR curves, F-measures and MAE of all methods on two databases, before and after combining with our approach. We can make several obviously observations. Firstly, our approach outperforms Base [17] in terms of three evaluation metrics especially on dataset MDUT-SAL, which demonstrates that our method is more robust and general for multiple salient objects detection. Secondly, all previous methods are higher improved after combination with our method on dataset MDUT-SAL. We consider that this is because SED2 [2] is relatively simple and other complex algorithms are possibly overfitted to SED2 dataset and do not generalize well to MDUT-SAL. Specifically, it is more

obvious that wCtr [18] which acquires the best result on both two databases, and improved results are best on multiple salient objects detection up to now. The motivation for combination has been fully proven in [17]. Finally, the performance gaps between previous methods are much smaller after combination as shown in Fig. 3 in sight of three metrics. Thus, the approach we proposed is an enhanced baselining for state-of-the-art methods.

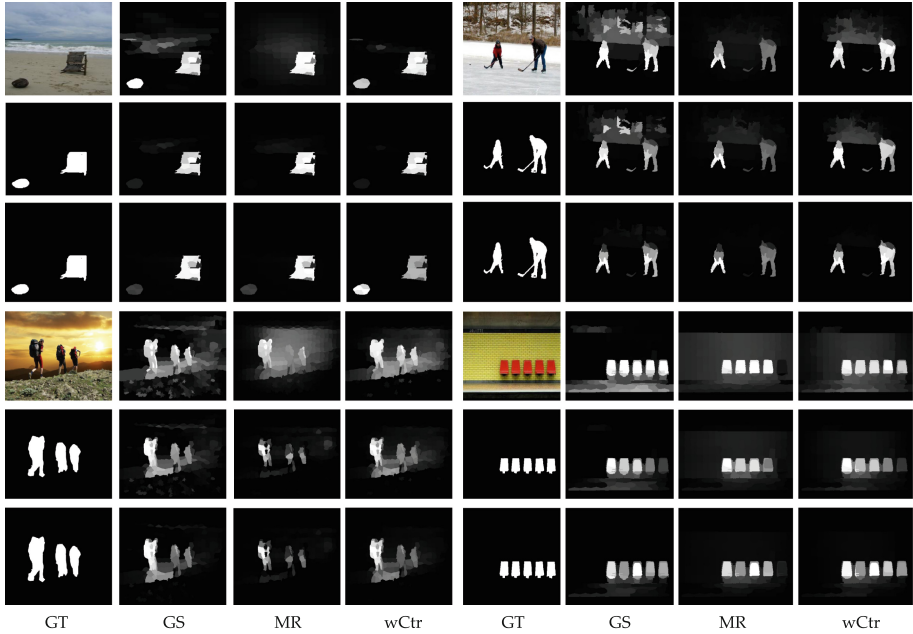


Fig. 4. Example results of three recent state-of-the-art methods. For each image, the first row shows the input image and related original results. The second row shows the ground truth and related improved results by combining Base [17]. The last row shows the ground truth and related enhanced results after combining our method.

5 Conclusion

We present an adaptive location for multiple salient objects detection based on *geodesic filtering* framework. It mainly introduces the salient points detection algorithm and Affinity Propagation (AP) clustering method to acquire a coarse salient objects location, called *adaptive location*. Then we use the *geodesic filtering* framework for a final fine saliency map. By comparing against the state-of-the-art methods, we find that our approach outperforms Base and improves other state-of-the-art methods after combination. For further validating our method, we propose a more challenging database MDUT-SAL than SED2. We hope our work and dataset can enhance the understanding of multiple salient objects detection in future.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2274–2282 (2012)
2. Alpert, S., Galun, M., Basri, R., Brandt, A.: Image segmentation by probabilistic bottom-up aggregation and cue integration. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8. IEEE (2007)
3. Borji, A., Cheng, M.M., Jiang, H., Li, J.: Salient object detection: a benchmark. *ArXiv e-prints* (2015)
4. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. *Science* **315**(5814), 972–976 (2007)
5. Guo, C., Zhang, L.: A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Trans. Image Process.* **19**(1), 185–198 (2010)
6. Li, Q., Zhou, Y., Yang, J.: Saliency based image segmentation. In: 2011 International Conference on Multimedia Technology (ICMT), pp. 5068–5071. IEEE (2011)
7. Li, Y., Fu, K., Zhou, L., Qiao, Y., Yang, J.: Saliency detection via foreground rendering and background exclusion. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 3263–3267. IEEE (2014)
8. Li, Y., Hou, X., Koch, C., Rehg, J.M., Yuille, A.L.: The secrets of salient object segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 280–287. IEEE (2014)
9. Perazzi, F., Krahenbuhl, P., Pritch, Y., Hornung, A.: Saliency filters: contrast based filtering for salient region detection. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 733–740. IEEE (2012)
10. Ren, Z., Gao, S., Chia, L.T., Tsang, I.H.: Region-based saliency detection and its application in object recognition. *IEEE Trans. Circ. Syst. Video Technol.* **24**(5), 769–779 (2014)
11. Stalder, S., Grabner, H., Van Gool, L.: Dynamic objectness for adaptive tracking. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012, Part III. LNCS, vol. 7726, pp. 43–56. Springer, Heidelberg (2013)
12. Van De Weijer, J., Gevers, T., Bagdanov, A.D.: Boosting color saliency in image feature detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(1), 150–156 (2006)
13. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 29–42. Springer, Heidelberg (2012)
14. Xie, Y., Lu, H.: Visual saliency detection based on bayesian model. In: 2011 18th IEEE International Conference on Image Processing (ICIP), pp. 645–648. IEEE (2011)
15. Xu, L., Lu, C., Xu, Y., Jia, J.: Image smoothing via l 0 gradient minimization. *ACM Trans. Graph. (TOG)* **30**, 174 (2011). ACM
16. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H.: Saliency detection via graph-based manifold ranking. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3166–3173. IEEE (2013)
17. Zhao, L., Liang, S., Wei, Y., Jia, J.: Size and location matter: a new baseline for salient object detection. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9005, pp. 578–592. Springer, Heidelberg (2015)
18. Zhu, W., Liang, S., Wei, Y., Sun, J.: Saliency optimization from robust background detection. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2814–2821. IEEE (2014)