Saliency Object Detection Based on Regions Merging and Its Application in Image Retrieval

Weiyi Wei, Hui Chen, Yahong Wen and Yu Wang

College of Computer Science and Engineering, Northwest Normal University,
Lanzhou, Gansu, 730070, China.
Email: 1611731310@qq.com

Abstract. In image retrieval, due to ignore the fact that different regions of the image have different levels of attraction to the human visual system in the traditional, so features are extracted of the entire image, result in the time is increased of the features matching. So we apply saliency object detection to image retrieval. However, extracting the saliency objects with complex background in the image is still a challenging problem. Thus we propose a region merging strategy to solve this problem. First, boundary super-pixels are clustered to generate the initial saliency map. Next, adjacent regions are merged by sorting the multiple feature values of each region. Finally, we get the final saliency map by merging regions by means of the distance from regions to the image center and the length of the boundary. After we apply the final saliency map to image retrieval. The experiments demonstrate that our method performs favorably on three datasets than state-of-art.

1. Introduction

Image retrieval aims at retrieving all images that contain the same object as the query within a potentially very large database of images. Content-based image retrieval [1] uses features such as color and texture for image retrieval, but these features ignore human visual characteristics.

Saliency object detection is an essential problem in computer vision which aims to highlight attractive regions in an image. It can be applied to such fields as image thumbnailing [2] and etc. In early, Itti et al [3] propose saliency detect model based on biological framework. It combines the color, brightness, direction and multi-scale features to compute saliency map. Later, researchers have found that contrast is the most important factor in human visual attention [4] Therefore, many saliency detection models use the contrast of all pixels in the image and the boundary information to calculate the saliency because boundary information is generally more likely to be background. Wei et al [5] utilize distance between object region and image boundaries to computer saliency. Yang et al [6] take the correlation of image boundaries and object region into consideration in image saliency detection. Meanwhile, there are many excellent regional merging algorithms. Uijlings et al [7] uses greedy algorithm to merge adjacent regions. Sun et al [8] proposed a phased region merging strategy based on the image features of different time periods which merge results of these methods are well. Therefore, we naturally exploit boundaries information to get initial result map and optimize the rough saliency map in two stages according to merging theory.

The saliency object is the regions of interest in the image, so we propose an optimized saliency object detection model and apply it to image retrieval.
2. Saliency Object Detection

In saliency object detection, the significant region is inaccurate in final saliency map when the image background is complex. In order to solve this problem, we propose a saliency detection method based on regions merging.

2.1. Construction of Initial Saliency Map.

We establish an initial saliency map in three steps and the specific process is as follows.

(1) Constructing color background map. In this paper, we use the method that means clustering and the number of clusters indicated use $M$. We set $M = 3$ (experiments show that the final effect is best when $M = 3$). And $C_m$ is defined as the color feature difference between super-pixel $i (i = 1, 2, \cdots, N)$ and the boundary class $m (m = 1, 2, \cdots, M)$. The class map $C_m$ can be formulated as

$$C_m = \frac{1}{p_m} \sum_{i} \left( \frac{1}{\exp\left(\frac{||mC_i||}{2\alpha_2}\right) + \eta} - ||C_i - C_m|| \right)$$

where $C_i$ and $C_m$ denotes the color features of each boundary class and super-pixel respectively, $||mC_i||$ denotes the Euclidean distance between each super-pixel and boundary class in Lab color space, $\alpha_1, \eta$ are balance factors, $p_m (m = 1, 2, \cdots, M)$ is the number of super-pixels in different boundary classes. We set $\alpha_1 = 0.2, \eta = 10$. The three background maps generated in this paper are shown in Figure 1.

![Figure 1. Color background map. From left to right: Input. 1st cluster Map. 2nd cluster Map. 3rd cluster Map. GT.](image)

(2) Constructing spatial correlation background map. The closer the distance between the two regions is, the greater the degree of influence and association in the space, and vice versa. We will construct the space correlation saliency map between each super-pixel and the boundary class by means of the spatial distance relationship between two regions. We defined $S_m$ as the spatial distance between each super-pixel and class $m (m = 1, 2, \cdots, M)$, so the $S_m$ can be formulated as

$$S_m = \frac{1}{p_m} \sum_{i} \exp\left(-\frac{||s_i - s_m||}{2\alpha_2^2}\right).$$

where $s_i, s_m$ denotes the spatial location of super-pixel $i$ and $m$. $\alpha_2 = 1$ is control parameter. The boundary clustering results are shown as Figure 2.

![Figure 2. Boundary clustering results. From left to right: Input. 1st Class Map. 2nd Class Map. 3rd Class Map](image)

(3) Constructing initial saliency map. As shown third map in Figure 3, because there are two black objects, it is very difficult to distinguish their saliency by means of color feature. But we can use spatial distance to enhance the saliency of the aircraft and weaken the saliency of the tree. Then the color background map and spatial correlation background map be merged. Let $S_m$ denotes the intensity factor to restrict the $C_m$. The initial saliency map can be written as $SC = \sum_{m=1}^{M} C_m \cdot S_m$. The merging result is shown as third map Figure 3.
2.2. Optimization of Regional Merger Strategy

It can be seen from the third map in Figure 3 that the results are not accurate. In order to optimize the rough initial saliency maps, a regional merging model is established in this section. In this paper, we merge the relevant regions with two steps.

The first phase of the merging:

The relationship tightness of each region with its adjacent regions must be calculated, and then merge them according to the degree of the relationship tightness. In this paper, the relationship tightness between regions is represented by color similarity and spatial tightness.

(1) Color similarity. Color similarity is defined as the difference in color feature between the super-pixel $i$ and $j$, it can be formulated as

$$C(i, j) = \left| c_i - c_j \right|.$$  

Where $\left| c_i - c_j \right|$ is the Euclidean distance between region $i$ and adjacent region $j$.

(2) Spatial tightness. The spatial tightness between two regions measures their degree of association in space. The correlation is measured by the length of the edge that their intersection. Let $ST(i, j)$, $S(i, j)$ and $T(i, j)$ as the spatial tightness, the spatial distance and spatial correlation between the super-pixel $i$ and adjacent super-pixel $j$, they can be written respectively as

$$ST(i, j) = (1 - S(i, j)) T(i, j)$$

and

$$S(i, j) = \left| L_i - s_j \right|$$

and

$$T(i, j) = \frac{B(i, j)}{\min(L(i), L(j))}.$$  

Where $s_i$, $s_j$ denote the spatial location for super-pixel $i$ and $j$, $B(i, j)$, $L(i)$ and $L(j)$ denote the length of their overlapping boundaries, the length of the boundary of super-pixels $i$ and $j$.

Finally, we define $P(i, j)$ as importance of the region $j$ to adjacent region $i$ by calculating their color similarity and spatial tightness.

$$P(i, j) = \omega_1 (1 - C(i, j)) + \omega_2 ST(i, j).$$  

Some small region $s_i$ must be merged into large region $s_j$ with greater relationship tightness. The merging results are shown in Figure 4. where $\omega_1 = 0.68$, $\omega_2 = 0.32$.

The second phase of the merging:

Based on above first merging, it is noted that majority of background regions are merged. However, there are some non-adjacent or adjacent backgrounds regions with greater intensity are not belong to saliency regions. Hence, in order to eliminate these background regions, we use $Sal$ denote these regional saliency value, according to the value $Sal$ decide whether merge it with background region or not. $Sal$ is influenced by the region-area($V$), the distance from regions to image center ($CS$) and the length of the boundaries they co-own for region boundaries and the image boundary ($BL$). The region with low value of $Sal$ is merged with background region.

(1) Region area $V$. Regions with larger area are more salient than smaller regions, so larger regions will be assigned greater saliency values. The formula can be written as

$$V = \frac{v}{\max(v)}.$$
Where $v$ is the area each associated object region in the image, and $\max(v)$ is the area of largest associated object region in the first merged map.

(2) Center distance $CS$. In general, the center regions of the image more attractive than others, and they are more likely to be saliency region [6,10], in this paper, therefore, calculates the regional $Sal$ according to the distance between the region and the center of the scene. Center distance can be expressed as $CS=\sum_{P}^{nP}[R,PC]/areaP$. Where $nP$ is the number of super-pixels in region $P$, $PC$ is the spatial location of the super-pixel included in the region $P$, $R$ is the center position of the image, $areaP$ is the area of the region $P$.

(3) Boundary length $BL$. The longer the overlapped boundary between a region with the image boundaries region is, the greater the possibility of being background regions [7], so $BL$ can be written as $BL=PB/BB$. Where $PB$ is the length of overlapped boundary that between each object region and image boundaries, and $BB$ is the total length value of the image boundaries. Hence, the regional saliency $Sal$ is defined as $Sal=V+CS-BL$. We merge the small objects into the large area by means of the $Sal$ sorted. When some regions’ $Sal$ value of several regions are very close, these regions can’t be merged. The result is shown in Figure 5.

![Figure 5. Second phase merge map. First phase merge map. From left to right: Input image. First phase of the merging. Second phase of the merging. GT.](image)

2.3. Experimental Results and Analysis

To evaluate the performance of the proposed model, we test on ASD, ECSSD and DUT-OMRON datasets. Moreover, we compare our method with five classic saliency detection algorithms. They are SER [11], SS [12], SR [13], SIM [9], FES [14].

(1) From subjective visual effects. This section selects parts representative images from three datasets to compare visual effects. The results are shown in Figure 6.

![Figure 6. Visual effect map. From left to right: input image. GT. Ours. SS. SR. SIM. SER. FES.](image)

In the first row in Figure 6, it is difficult to detect the object region because the image with the complex background. But the saliency object region can be shown complete and with clear boundaries by our algorithm. In the second row in Figure 6 is a multi-objective map. We can see from the generated saliency map, the results of our algorithm are closer to the GT map. The third row in Figure 6, a sharp contrast between black object regions and white background regions. Our algorithm can detect more clear regions. For other comparison algorithms, when reveal saliency object region the background regions also be highlighted.

(2) Objective evaluation. We evaluate the performance using Precision-Recall (PR) curves and F-measure. Given a saliency map, the PR values are computed by segmenting the saliency map with a
threshold within \([0, 255]\) to construct a binary map which is then compared with the corresponding ground truth. The result diagram is shown as in Figure 7.

![Contrast data diagram](image)

**Figure 7.** Contrast data diagram. The frist row denotes recall and precision curves of different methods in ASD dataset, ECSSD dataset and DUT-OMRON dataset. The second row denotes recall, precision and F-measure value in ASD dataset, ECSSD dataset, DUT-OMRON dataset.

In Figure 7, From the first two map in the first line, we can see that the numerical line of our algorithm is above the numerical line of the all comparison algorithm. In the first two map in the second line the F-measure of our algorithm is higher than other algorithms. In the ECSSD datasets, although the background in the images are more complex, the accuracy and recall of o algorithm is also close to 90%. Third column is the Recall values, Precision values, and F-measure in the DUT-OMRON datasets. The images in this datasets are complex, so all algorithms are less effective than 70%. But the highest F-measure 0.55 of our algorithm.

3. **Image Retrieval Result**

In this section, we use the hybrid features (including basic features and saliency features) to perform the image retrieval. In Figure 8, we can see that the saliency regions can be extracted by saliency object detection, and then the color and texture features are picked up in final saliency map to perform image retrieval with features matching in image datasets.

![Image Retrieval Result](image)

**Figure 8.** The first and second rows represent respectively: the input image and the saliency map that use our algorithm.

We use Sim denotes the similarity that retrieved image and the image of a datasets, it can be measured as: the color feature \(S\) and texture feature \(L\) are extracted by using the color moments and LPB operators in the saliency maps. Then, the comprehensive result \(M\) measured by the weighted sum of the two features. Finally, the Sim be measured use Euclidean distance. It can be wrote as follows

\[ M = \lambda_1 S + \lambda_2 \]

and

\[ Sim = \| M - M_i \| \] where \(M_i\) is the comprehensive value of each image in image of a datasets, where parameters \(\lambda_1, \lambda_2\) are 0.4 and 0.6 respectively.

In order to verify the retrieval effect of our algorithm, firstly, we establish an image dataset with 1,000 images. Then, use our saliency object detection algorithm to retrieve image. Finally, the retrieval results are compared between our retrieval algorithm and traditional retrieval algorithm. The comparison results are as follows
Figure 9. Comparison results map. First two rows are direct retrieval result maps, and the last two rows are retrieval result maps that use saliency object detection.

From the above results, we can see that using the salient object is better than that of the direct retrieval from Figure 9. The results such as follow

| Table 1. Precision of retrieval result. |
|----------------------------------------|
| Method       | elephant | flower |
| Basic Feature| 58.9     | 75.4   |
| Hybrid Feature| 69.8     | 83.6   |

As can be seen from the above table, the retrieval result based on the saliency object detection is more accurate. So a robust saliency object detection algorithm is very vital for image retrieval.

4. Conclusion
In this paper, a saliency object detection model for regional merging is proposed and applied it to image retrieval. Experimental results show that our saliency object detection algorithm is effective and retrieval result is more accurate.

5. Acknowledge
This work was supported by National Natural Science Foundation of China (Grant 61762080, 61762078), Science and Technology Plan of Gansu Province (17YF1FA119). Corresponding author: Hui Chen, E-mail: 1611731310@qq.com

6. References
[1] T Kato. Database architecture for content-based image retrieval [M]. SPIE Proceedings, 1992, 1662 pp 112-123.
[2] Marchesotti L, Cifarelli C, Csurka G. A framework for visual saliency detection with applications to image thumbnailing [C]. Int. Conf. on Computer Vision IEEE 2009 pp 2232-2239.
[3] L Itti, CKoch, and ENiebur. A model of saliency-Based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligente 1998 20(11) pp 1254-1259.
[4] Yu-fei Ma, Hong-jian Zhang. Contrast-based image attention analysis by using fuzzy growing[C]. Eleventh ACM Int. Conf. on Multimedia 2003 (New York) p 374.
[5] Y Wei, F Wen, W Zhu, J Sun. Geodesic saliency using background priors[C]. In Computer Vision-ECCV 2012 pp 29-42.
[6] C Yang, L Zhang, H Lu, X Ruan, M H Yang. Saliency detection via graph-based manifold ranking[C]. In Computer Vision and Pattern Recognition 2013 pp 3166-3173.
[7] J Uijlings, K E ASande, etal. Selective search for object recognition [J]. Int.Journal of Computer Vision 2013 104(2) pp 154-171.
[8] F Sun, K HQin, etal. Image saliency detection based on region merging [J]. Computer aided design and graphics Journal 2016 28(10) pp 1679-1687.
[9] Murray N, Vanrell M, Otazu X, CA Parraga. Saliency estimation using a non-parametric low-level vision model [C]. IEEE Conf. on Computer Vision and Pattern Recognition, Colorado Springs 2011 pp 433-440.
[10] N Richard, F Nielsen. Statistical region merging [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 2004 26 (11) pp 1-7.

[11] H J Seo, P Milanfar. Static and space-time visual saliency detection by self-resemblance [J]. Journal of Vision 2009 9(12) pp 1-27.

[12] X Hou, J Harel, C Koch. Image Signature: Highlighting sparse salient regions [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence 2012 34(1) pp 194-201.

[13] X. Hou and L. Zhang. Saliency detection: A spectral residual approach[C]. IEEE Conf. on Computer Vision and Pattern Recognition 2007 (Minneapolis USA) pp 1-8.

[14] HR Tavakoli, E Rahtu, J Heikkilä. Fast and efficient saliency detection using sparse sampling and kernel density estimation[C]. Scandinavian Conf. on Image Analysis 2011 6688(2) pp 666-675.