GEODESIC SALIENCY PROPAGATION FOR IMAGE SALIENT REGION DETECTION

Keren Fu[A] *Chen Gong*[A] *Irene Y.H. Gu*[B] *Jie Yang*[A*]

[A] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China
[B] Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden

ABSTRACT

This paper proposes a novel geodesic saliency propagation method where detected salient objects may be isolated from both the background and other clutter by adding global considerations in the detection process. The method transmits saliency energy from a coarse saliency map to all image parts rather than from image boundaries in conventional cases. The coarse saliency map is computed using the combination of global contrast and Harris convex hull. Superpixels from pre-segmented image are used as pre-processing to further enhance the efficiency. The proposed propagation is geodesic distance assisted and retains the local connectivity of objects. It is capable of rendering a uniform saliency map while suppressing the background, leading to salient objects being popped out. Experiments were conducted on a benchmark dataset, visual comparisons and performance evaluations with 9 existing methods have shown that the proposed method is robust and achieves the state-of-the-art performance.

Index Terms— Saliency detection, Geodesic distance, Saliency propagation, Saliency map

1. INTRODUCTION

Saliency detection aims at modeling visual attentions of human eyes and produces a saliency map as the output. Saliency detection methods can be roughly categorized into two types: *bottom up* approaches that are stimulus and data driven, and *top down* approaches that are tasks and knowledge driven. Based on different expectations, bottom up approaches can be furthered subdivided in two groups: approaches towards human attention modeling [1, 2, 3] and towards salient region detection [4, 5, 6, 7, 8, 9, 10, 11, 12]. The proposed method belongs to the latter subcategory. The goal is to highlight an entire salient object in a saliency map, leading to more elegant performance for applications like object detection and segmentation, content based image retrieval and resizing.

Recently, Wei *et al* [11] introduce a geodesic saliency detection method that uses background priors. In their method, the boundary parts of image are deemed as background components having "non-saliency" property. Such "non-saliency" is then propagated to inner image parts along a path with the shortest geodesic distance. As a salient object is usually isolated from its background, the geodesic distance between image boundaries and object parts is relatively large. Since "non-saliency" parts hardly propagate towards a salient object, enhanced detection of salient objects may be obtained. In practice, they define the image saliency of a specific superpixel as its shortest geodesic distance to the boundary component set. The drawback in [11] is that only the background information is used, where objects appeared to be isolated from the background belong to salient objects. This could fail in some cases, e.g. images contain multiple meaningless but isolated clutter parts. Motivated by this, we propose a method that salient objects may be isolated from both the background and other meaningless clutter/objects. The key to the solution is to include global considerations in the detection process. We propose a novel method that separates a salient object from background by exploiting information from both the background, clutter and other objects. Compared with [11] where energy is propagated from image boundaries, our method transmits saliency energy from a weighted coarse saliency map that is globally computed to all image parts. Further, our method differs from [11] in propagation means.

According to the analysis above, our method is based on the following two properties of salient objects:

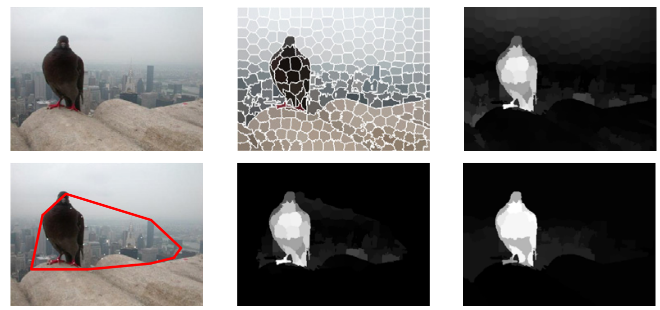

Fig. 1. The results from individual steps in the proposed approach. From left to right, top to bottom: original image, superpixel pre-segmentation, image of global contrast, Harris convex hull, coarse saliency map, and final saliency map from saliency propagation.

*This research is partly supported by National Science Foundation, China (No: 61273258, 61105001), Ph.D. Programs Foundation of Ministry of Education of China (No. 20120073110018). Jie Yang is the corresponding author (email: jieyang@sjtu.edu.cn).

(i) Salient objects are popped out from both the background and clutter;
(ii) Both salient objects and image background consist of homogenous components.

The property (ii) holds since objects usually contain some homogenous surfaces with inner connectivity [16]. This inner connectivity is somewhat used as color "compactness" in [8, 5, 9] by measuring the variances of color distributions. We propose to directly exploit the "connectivity" instead of the "compactness" property. Since "connectivity" also holds for most backgrounds as they could contain large homogenous part, such property helps to suppress the background clutter, as we will show later. Under these assumptions, we propose a novel method to propagate the saliency from relatively important image parts to their surroundings. A coarse saliency map is first computed by combining a global contrast assumption [7] and Harris convex hull [12], saliency propagation is then proposed to transmit the saliency energy of coarse saliency map to all image parts using geodesic distances.

The rest paper is organized as follows. Section 2 describes the proposed method. Section 3 shows the comparisons and evaluations. The conclusion is given in Section 4.

2. THE PROPOSED METHOD

In the proposed method, a coarse saliency map is first computed by exploiting the first property of salient objects (global popping out). Saliency propagation method is then designed to propagate the saliency energy from the coarse map to all parts of the image. Fig.1 shows the resulting images in each step. Before computing the coarse map, we first pre-segment the image into superpixels [13] of relatively consistent size in order to facilitate the computation.

2.1. Coarse saliency map

Let superpixels of image be denoted by R_i, $i = 1, 2, \cdots, N$, and the average colors in LAB color space and position vectors be c_i^{Lab} and p_i, respectively. LAB color space is chosen for better characterizing the human visual character. Based on the global popping out property of a salient object, we use the following two simple features to compute the coarse saliency map, respectively the global contrast and information conveyance.

Global Contrast: Observing that if a superpixel has a high global contrast to the remaining parts [7], it is more likely that it is a part of a salient object. This is exploited for computing the global contrast features of superpixels R_i by

$$F_i^{global} = \sum_j ||c_i^{Lab} - c_j^{Lab}||_2^2 \tag{1}$$

where the quadratic term is used for better suppressing the background. In most cases, (1) may highlight a salient object though partially and unevenly. Then all superpixels' global contrast features are normalized to the interval [0,1].

Information Conveyance: Much attention has been paid on extracting salient objects since these objects convey essential information in images [3]. Many salient point detection methods, e.g. "keypoints" detection by Harris corner detection or SIFT have been used in object recognition and image retrieval. Such keypoints contain more structural information meanwhile they are scale and orientation invariant. To extract salient objects, we employ the Harris convex hulls [12], since such convex hulls provide coarse regions that are more likely to contain salient objects, and superpixels in convex hulls are likely to convey more information than the rest image parts. Based on this, we define the information feature as

$$F_i^{information} = \begin{cases} 1 & \text{if } p_i \text{ is in the convex hull} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

It is worth noting that before using a convex hull to bound the boosted Harris corners [15], corners that are near image boundaries are eliminated since salient objects seldom touch image boundaries [11].

Combination: The above two features are then combined to obtain the coarse saliency map by

$$S_i^{coarse} = F_i^{global} \cdot F_i^{information} \tag{3}$$

One advantage of combining these two terms is that the background clutter in F_i^{global} is reduced while different parts in the convex hull are weighted. Fig.1 shows an example of the resulting coarse saliency map. Notice that the convex hull may also include some undesired background and exclude parts of the desired object as well (Fig.5). Thus we introduce a novel saliency propagation method as below to tackle this issue.

2.2. Saliency propagation

A key contribution of this paper is to use a global propagation procedure based on geodesic distance. This allows one to propagate energy from a binary or weighted coarse saliency map. It is worth noting that the proposed propagation approach is different from those in [11, 14], where these methods define the final propagation value of individual patch as its shortest geodesic distance from a given subset. Our propagation procedure is formulated as

$$S_i^{propagation} = \sum_j f_{j \to i} S_j^{coarse} \tag{4}$$

where $S_i^{propagation}$ is the transmitted energy to the superpixel R_i from R_j, $j = 1, \cdots, N$. S_j^{coarse} is the corresponding saliency energy in the coarse saliency map defined in (3). The saliency energy transmitted from R_j to R_i is represented by $f_{j \to i} S_i^{coarse}$, where $f_{j \to i}$ is the propagation intensity between R_j and R_i. Under the local connectivity consideration, $f_{j \to i}$ is defined as

$$f_{j \to i} = \frac{1}{N} e^{-\beta d(R_i, R_j)} \tag{5}$$

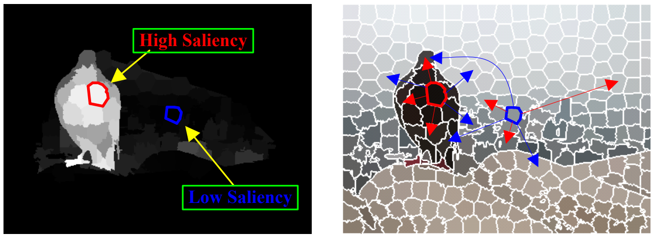

Fig. 2. Illustration for saliency propagation. From left to right: the coarse saliency map and the description for propagation intensities. In the left subimage, superpixel with the red boundary is in relatively high saliency while superpixel in the blue boundary is in low saliency. In the right subimage, the saliency energy propagates from these two superpixels to the other parts. The red arrows represent high propagated intensity area while the blue arrows stand for low propagated intensity area.

where β is a parameter controlling the transmitting intensity ($\beta = 0.05$ is set for unnormalized LAB color space in our tests), $N = \sum_j e^{-\beta d(R_i, R_j)}$ is used for normalizing the saliency energy. $d(R_i, R_j)$ is the geodesic distance between R_i and R_j:

$$
\begin{aligned}
d(R_i, R_j) &= \min_{P_1=i, P_2, \ldots, P_n=j} \sum_{k=1}^{n-1} \text{ColorDifference}(R_{P_k}, R_{P_{k+1}}) \\
&= \min_{P_1=i, P_2, \cdots, P_n=j} \sum_{k=1}^{n-1} u\left(\Delta_{P_k} - d_0\right) \Delta_{P_k} \\
&\text{s.t.: } R_{P_k} \text{ and } R_{P_{k+1}} \text{ are adjacent superpixels}
\end{aligned}
\tag{6}
$$

where $\Delta_{P_k} = ||c_{P_k}^{Lab} - c_{P_{k+1}}^{Lab}||_2$, $u(\cdot)$ is the step function, and d_0 is a small threshold to avoid small color difference value being accumulated [11]. (6) implies that one needs to find a path from R_i to R_j (denoted as $\{P_1, \cdots, P_n\}$) to ensure the global minimization in (6). The symmetry $d(R_i, R_j) = d(R_j, R_i)$ implies $f_{i \to j} = f_{j \to i}$. Our final saliency map is given by normalizing all $S_i^{propagation}$ to the interval [0, 1].

It is worth mentioning that, due to using superpixels to replace pixels, (6) is an approximation of the standard geodesic distance where the gradient integration is performed along the best path consisting of pixels [14]. Based on using homogenous superpixel regions, the gradient integration is approximated by using the sum of adjacent color appearance differences in (6). In our method, the point-to-point shortest path between R_i and R_j is computed via the *Dijkstra's* algorithm. We treat each superpixel as a vertex in the undirected graph and the corresponding edge weight connected to the adjacent vertices (corresponds to neighbor superpixels) is modeled using color appearance differences. Then the geodesic distance is obtained according to (6).

Fig.2 shows an illustration of propagation intensity from using a coarse map. Since an object often contains some homogenous parts, the initial saliency value of a superpixel could be spread to the other connected homogenous parts, indicating the propagation may be achieved through connectivity. This can also be observed in the background where image background can be divided into large homogenous regions. Noting that the coarse saliency could render more saliency to the target object while less to the background (Fig.2). Hence, low saliency in background part could be spread over the entire background region after propagation. Eventually, the background could be suppressed to very low saliency values.

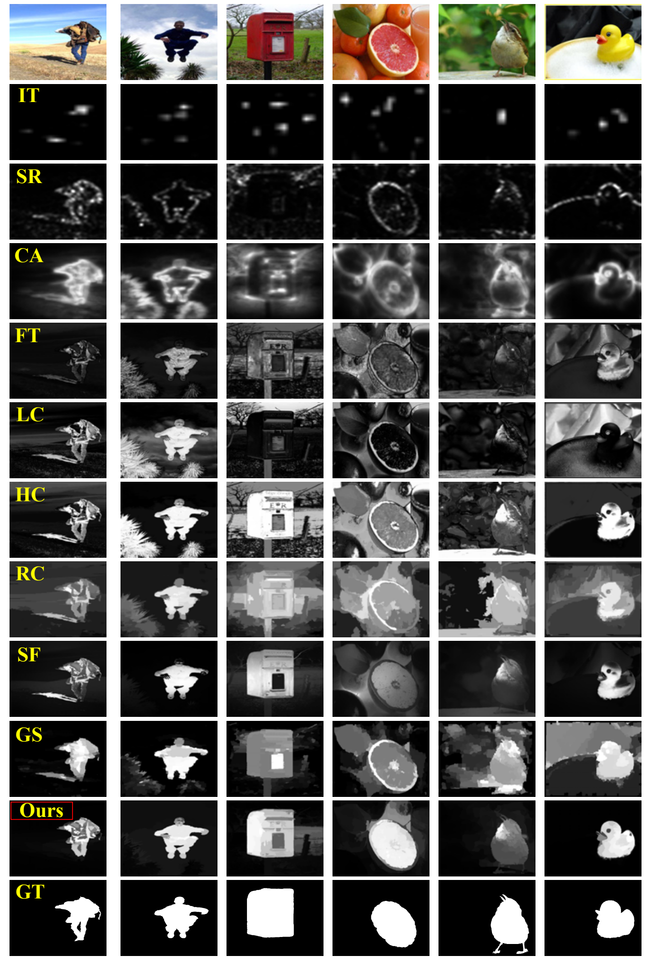

Fig. 3. Visual comparisons of resulted saliency maps from the proposed method and 9 existing methods. From top to bottom: original images, IT, SR, CA, FT, LC, HC, RC, SF, GS, the proposed method, and GT (ground truth).

Such energy propagation has two advantages: one is the uniform saliency rendering that is benefitted from the connectivity of both object and background; the other is the saliency energy of background in a convex hull may be distributed to the entire image background meanwhile the object saliency is preferably repaired (Fig.3 and Fig.5), hence the background is suppressed whereas complete salient object is popped out.

3. EXPERIMENTS AND RESULTS

Tests and Comparisons: The proposed method was tested on a benchmark dataset [6] which contains 1000 images, and compared with 9 existing state-of-the-art methods including

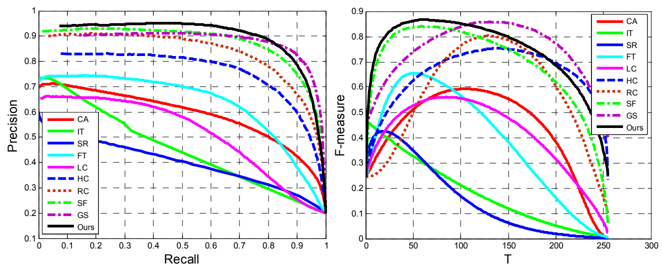

Fig. 4. Performance comparisons for the proposed method and the 9 existing methods. Left: precision-and-recall curves; Right: F-measure curves.

IT [1], SR [2], CA [10], FT [6], LC [4], HC [7], RC [7], SF [9], and Geodesic Saliency (GS) using background priors [11]. Note for GS method, we use their superpixel based implementation [11] since it achieves better performance than grid based. Fig.3 shows some visual results. Observing our test results and comparisons with the 9 existing methods, the proposed method is shown to have provided marked improvement on popping out salient objects meanwhile keeping much cleaner background, probably due to the benefit of jointly exploiting global contrast assumption and Harris convex hull in the saliency propagation procedure.

Evaluations: Evaluations were conducted on a range of test results. In these tests, binary masks from the saliency maps were generated by fixing the threshold (similar to [6, 7, 8, 9]) over a range of values $T = [0, 255]$. Evaluations were performed using two criteria: one is by computing the average precision and recall values over the entire range of T, the other is to apply the F-measure by integrating both precision and recall. Fig.4 shows the precision-recall curves and the F-measure curves respectively, obtained from the proposed method and 9 state-of-the-art methods.

Observing the evaluation results in Fig.4 (left), one can see from the precision-recall curve that the proposed method outperforms all these 9 methods by a noticeable margin (highest precision is 0.95 for our method and 0.93 for SF and 0.91 for GS). Further, observing F-measure in Fig.4 (right) shows that the proposed algorithm achieves the best performance for $T \leq 110$, and is secondary best for $T > 110$ and ranks the third for $T > 190$ (for $T > 110$, GS method is somewhat better under the F-measure, which may be attributed to the higher recall of GS under a relative large threshold T). Further, T=61 has led to the highest F-measure score 0.87 from the proposed method (0.86 for GS and 0.84 for SF).

Discussion on performance against the fluctuations in global contrast and Harris convex hulls: It is worth noting that the proposed method does not highly rely on the quality of coarse saliency maps. The soft conditions are that a convex hull should at least cover part of salient objects while object parts should have a relative high saliency energy as comparing with its surroundings in the hull. As can be seen from examples in Fig.5, these two conditions can be easily satisfied in most cases. In these cases, although a convex hull may covers only part of object, or more than half area of the hull covers the background, the proposed method can pop out the real salient objects while suppressing the background.

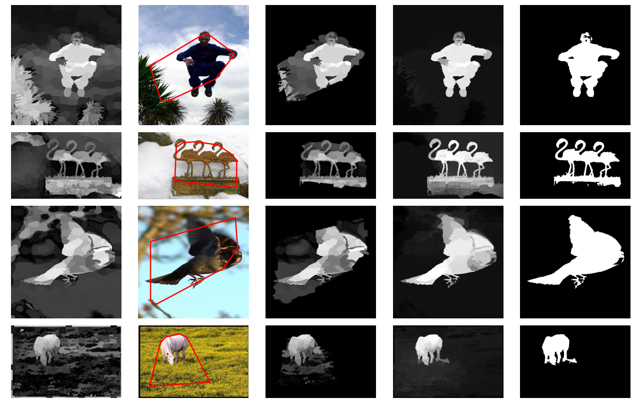

Fig. 5. Robustness of performance against the fluctuations in global contrast and Harris convex hulls: results from the proposed method. From left to right: results of global contrast; detected Harris convex hulls; results of coarse saliency maps; results after the proposed saliency propagation; results of object masks from saliency maps using the adaptive threshold, defined as two times the mean saliency of the map [6, 7, 8, 9].

Last but not least, we found out that in practical cases, the second condition can be relaxed. The tests also showed that although the background may have initially similar/higher saliency values than the salient object, it could still yield lower saliency values after the final propagation. This usually happens when the area of the homogenous parts of the background is much larger than the salient object. This indicated that the proposed method is robust against the variants in detected global contrast and Harris hulls.

4. CONCLUSION

This paper proposes a novel and efficient method for salient object detection, where a coarse map is employed through combining global contrast and Harris convex hulls, followed by propagating the saliency energy to whole image areas through using the geodesic distance between superpixels. Visual comparisons of test results from the proposed method and 9 state-of-the-art existing methods have shown visible improvement of the proposed method. This is probably due to the combination effects of using global contrast and convex hulls in the coarse map, and the geodesic distance in the saliency energy propagation, where more uniform saliency rendering and higher background suppression are observed during the tests, leading to the better popping out of objects of interest. Results from further tests using a range of fixed thresholds and evaluations under recision-and-recall and F-measure have provided further support to the effectiveness and robustness of the proposed method. Finally, we have observed from our tests that relaxing the assumptions on global contrast and Harris convex hulls to some extent would not significantly impact the results.

5. REFERENCES

[1] L.Itti, C.Koch, E.NieBur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", IEEE trans. PAMI, vol 20, no.11, pp.1254–1259, 1998.

[2] X.Hou, L.Zhang, "Saliency detection: A spectral residual approach", proc. CVPR, 2007.

[3] N.D.B. Bruce, J.K. Tsotsos, "Saliency Based on Information Maximization", proc. Int'l conf.NIPS, 2005.

[4] Y.Zhai, M. Shah, "Visual attention detection in video sequences using spatiotemporal cues", ACM Multimedia, pp.815–824, 2006.

[5] T.Liu, Z.Yuan et al, "Learning to detect a salient object", IEEE Trans. PAMI, vol 33, no. 2, pp. 353–367, 2011.

[6] R.Achanta, S.Hemami et al, "Frequency-tuned salient region detection", in proc. CVPR, 2009.

[7] M.Cheng, G.Zhang, et al, "Global contrast based salient region detection" proc. CVPR, 2011.

[8] K. Fu, C.Gong, et al, " Salient Object Detection via Color Contrast and Color Distribution", in proc. ACCV, 2012.

[9] F.Perazzi, P. Krahenbul et al, "Saliency Filters: Contrast Based Filtering for Salient Region Detection", proc. CVPR, 2012.

[10] S.Goferman, L.Zelnik-Manor, A. Tal, "Context-aware saliency detection", in proc. CVPR, 2010.

[11] Y.Wei, F.Wen, et al, "Geodesic Saliency Using Background Priors", in proc. ECCV, 2012.

[12] Y.Xie, H.Lu, "Visual saliency detection based on Bayesian model", Proc. Int'l conf ICIP, 2011

[13] R.Achanta, et al, "SLIC Superpixels", Technical report, 2010.

[14] A.Criminisi, T.Sharp et al, "Geodesic image and video editing", ACM Trans on Graph., vol.29, no.5, pp.134, 2010.

[15] J.Weijer, T.Gevers et al, "Boosting color saliency in image feature detection", IEEE Trans. PAMI, vol 28, pp. 150–156, 2006.

[16] S.Vicente, V.Kolmogorov et al, "Graph cut based image segmentation with connectivity priors", in proc. CVPR, 2008.