

Adaptive Multi-Level Region Merging for Salient Object Detection

Keren Fu^{1,2}

fkrsuper@sjtu.edu.cn, keren@chalmers.se

Chen Gong¹

goodgongchen@sjtu.edu.cn

Yixiao Yun²

yixiao@chalmers.se

Yijun Li¹

leexiaoju@sjtu.edu.cn

Irene Yu-Hua Gu²

irenegu@chalmers.se

Jie Yang¹

jjeyang@sjtu.edu.cn

Jingyi Yu³

yu@eecis.udel.edu

¹ Institute of Image Processing and Pattern Recognition

Shanghai Jiao Tong University

Shanghai, P.R. China

² Department of Signals and Systems

Chalmers University of Technology

Gothenburg, Sweden

³ University of Delaware

Newark, USA

Salient object detection is a long-standing problem in computer vision and plays a critical role in understanding the mechanism of human visual attention. In applications that require object-level prior (e.g. image re-targeting), it is desirable that saliency detection highlights holistic objects. Lately over-segmentation techniques such as SLIC superpixel [6], Mean-shift [1], and graph-based [3] segmentations are popular among saliency detection due to their usefulness on eliminating background noise and reducing computation cost. However, individual small segments provide little information about global contents. Such schemes have limited capability on modeling global perceptual phenomena. Fig.1 shows a typical example. The entire flower tends to be perceived as a single entity by human visual system. It is easily imagined that saliency computation with the help of coarse segmentation is conducive to highlighting entire object while suppressing background. As it is important to control segmentation level to reflect proper image content, more recent approach benefits from multi-scale strategies to compute saliency on both coarse and fine scales with fusion [4]. [4] merges a region to its neighbor region if it is smaller than pre-defined sizes. The underlying problem may be that scale parameters in [4] are crucial to performance. A salient region may not appear in the proper level if it is smaller than the defined size. On the other hand, large background regions with close colors may not be merged together if they are larger than the defined size.

In this paper we propose an alternative solution, namely by quantifying *contour strength* to generate varied levels. Compared to [4], we use edge/contour strength and a globalization technique during merging. Our contributions include:

1. Develop an adaptive merging strategy for salient object detection rather than using several fixed “scales”. Our method generates intrinsic optimal “scales” when the merging continues.
2. Incorporate additional global information by graph-based spectral decomposition to enhance salient contours. It is useful in salient object rendering.
3. Performance obtained is similar to other state-of-the-art methods even though simple region saliency measurements are adopted for each region.

As shown in Fig.2, our framework first performs over-segmentation on an input image by using SLIC superpixels [6], from which merging begins. To acquire holistic contour of salient objects as the merging process proceeds, we propose a modified graph-based merging scheme inspired by [3] which sets out to merge regions by quantifying a pre-defined region comparison criterion. Specifically before merging starts, a globalization

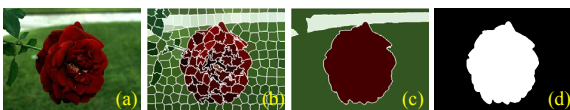


Figure 1: Multi-level segmentation for salient object detection. (a) shows a sample image from MSRA-1000 dataset [5]. (b) Over-segmentation using superpixels destroys the semantic content such as the flower. (c) A coarse segmentation derives from (b) maintains semantic holism. (d) object mask (ground truth).

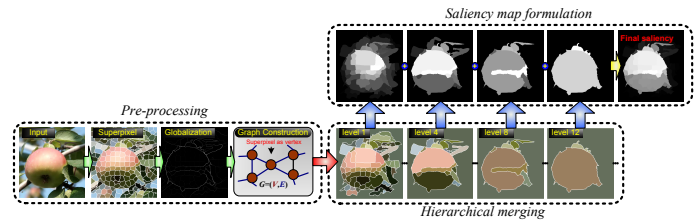


Figure 2: The processing pipeline of our approach.

procedure is proposed and conducted to pop out salient contours whereas suppress background clutter (Fig.2). At each level, we formulate an intermediate saliency map based on several simple region saliency measurements. Finally a salient object will be enhanced by summing across-level saliency maps (Fig.2).

Let initial SLIC superpixels be $R_i^0, i = 1, 2, \dots, N$. A graph $G = (V, E)$ is defined where vertices V are superpixels, and E are graph edges. Let $R^l = \{R_1^l, R_2^l, \dots\}$ be a partition of V in the l th level and $R_k^l \in R^l$ corresponds to its k th part (namely region). With the constructed edge E , a criterion D is defined to measure the pairwise difference of two regions R_i^l, R_j^l as:

$$D_{ij}^l = D(R_i^l, R_j^l) = \text{mean}_{v_k \in R_i^l, v_m \in R_j^l, e_{km} \in E} \{e_{km}\} \quad (1)$$

where “mean” is averaging operation over graph edges connecting R_i^l and R_j^l . In order to adapt merging to “large” differences (strong edges), we define a threshold Th to control the bandwidth of D_{ij}^l : at level l , we fuse two components R_i^l, R_j^l in R^l if their difference $D_{ij}^l \leq Th$. Suppose $R_i^l, R_j^l, R_k^l, \dots$ are regions that have been merged into one larger region R_{new}^l at this level, we then update $R^l \leftarrow (R^l / \{R_i^l, R_j^l, R_k^l, \dots\}) \cup R_{new}^l$ (“/” and “ \cup ” are set operation), where R_{new}^l is the newly generated region. At next level $l + 1$, Th is increased as $Th \leftarrow Th + T_s$ where T_s is a step length. In graph edge construction (i.e. E), a globalization procedure is proposed inspired by a contour detector *gPb* [2]. The technique attempts to achieve area completion by solving the eigen-problem on the local affinity matrix. This operation also meets the Gestalt psychological laws properties [7, 8] i.e. *closure* and *connectivity* based on which human perceive figures.

To show the effectiveness of the proposed region merging and integration scheme, each merged region is just evaluated using several simple region saliency measurements. Even though like this, we show the proposed method already can achieve competitive results against the best methods among the state-of-the-art.

- [1] D. Comaniciu et al. Mean shift: a robust approach toward feature space analysis. *TPAMI*, 24(5):603–619, 2002.
- [2] P. Arbelaez et al. Contour detection and hierarchical image segmentation. *TPAMI*, 33(5): 898–916, 2010.
- [3] P. Felzenszwalb et al. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, 2004.
- [4] Q. Yan et al. Hierarchical saliency detection. In *CVPR*, 2013.
- [5] R. Achanta et al. Frequency-tuned salient region detection. In *CVPR*, 2009.
- [6] R. Achanta et al. Slic superpixels compared to state-of-the-art superpixel methods. *TPAMI*, 34(11):2274–2282, 2012.
- [7] K. Koffka. Principles of gestalt psychology. 1935.
- [8] S. Palmer. Vision science: Photons to phenomenology. *The MIT press*, 1999.