Adaptive Multi-Level Region Merging for Salient Object Detection

Keren Fu[1,2]
fkrsuper@sjtu.edu.cn,keren@chalmers.se
Chen Gong[1]
goodgongchen@sjtu.edu.cn
Yixiao Yun[2]
yixiao@chalmers.se
Yijun Li[1]
leexiaoju@sjtu.edu.cn
Irene Yu-Hua Gu[2]
irenegu@chalmers.se
Jie Yang[1]
jieyang@sjtu.edu.cn
Jingyi Yu[3]
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

Abstract

Most existing salient object detection algorithms face the problem of either under- or over-segmenting an image. More recent methods address the problem via multi-level segmentation. However, the number of segmentation levels is manually predetermined and only works well on specific class of images. In this paper, a new salient object detection scheme is presented based on adaptive multi-level region merging. A graph-based merging scheme is developed to reassemble regions based on their shared contour strength. This merging process is adaptive to complete contours of salient objects that can then be used for global perceptual analysis, e.g., foreground/ground separation. Such contour completion is enhanced by graph-based spectral decomposition. We show that even though simple region saliency measurements are adopted for each region, encouraging performance can be obtained after across-level integration. Experiments by comparing with 13 existing methods on three benchmark datasets including MSRA-1000, SOD and SED show the proposed method results in uniform object enhancement and achieves state-of-the-art performance.

1 Introduction

Salient object detection is a long-standing problem in computer vision and plays a critical role in understanding the mechanism of human visual attention. Applications in vision and graphics are numerous, especially in solving problems that require object-level prior such as "proto objects" detection [31] and segmentation [7, 16], content based image cropping [39],

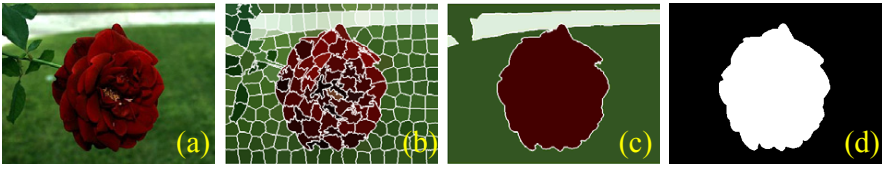

Figure 1: Multi-level segmentation for salient object detection. (a) shows a sample image from MSRA-1000 dataset [23]. (b) Over-segmentation using superpixels destroys the semantic content such as the flower. (c) A coarse segmentation derives from (b) maintains semantic holism. (d) object mask (ground truth).

thumbnailing [15], resizing and re-targeting [27, 33]. In such applications, it is desirable that a detection scheme highlights holistic objects.

Many existing methods [8, 9, 11, 17, 22, 25, 27, 29, 35] exploit contrast and rarity properties of local superpixels or regions. They employ over-segmentation techniques such as SLIC superpixel [24], Mean-shift [4] or graph-based [21] segmentations to decompose an image into small edge-preserving segments. These abstraction techniques are known to be useful for eliminating background noise and reducing computation by treating each segment as a processing unit. However, individual small segments provide little information about global contents. Such schemes have limited capability on modeling global perceptual phenomena [37, 38]. Fig.1 shows a typical example. The entire flower tends to be perceived as a single entity by human visual system. However, local-segment based schemes (e.g. [8]) partition the flower into parts (Fig.1 (b)), each of which alone does not reflect the meaning of "flower". In contrast, a coarse segmentation (derived from Fig.1 (b)) that attempts to keep semantic holism (Fig.1 (c)) better models such gist. It is easily imagined that saliency computation with the help of such coarse segmentation is conducive to highlighting entire objects while suppressing background.

As it is important to control segmentation level to reflect proper image content, some recent approaches benefit from multi-scale strategies to compute saliency on both coarse and fine scales with fusion. Yan *et al* [22] define three levels of sizes for regions and merge a region to its neighbor region if it is smaller than defined sizes. Despite good performance of [22], the underlying problem may be that scale parameters in [22] are crucial to performance. A salient region might not be in the proper level if it is smaller than the defined size. In addition, large background regions with close colors may not be merged together if they are larger than the defined size. Since appropriate merging may facilitate global perceptual phenomena analysis (Fig.1), to find coincidence of salient object in multi-scales, in this paper we propose an alternative solution, namely by quantifying *contour strength* to generate varied levels. Compared to [22], we use edge/contour strength and a globalization technique during merging, while [22] merges according to region size. Main advantages that lead to robust performance of the proposed method against [22] include: (*i*) use edges/contours and their strengths (rather than region size), reflecting object saliency that is often indicated by enclosed strong edges; (*ii*) use a globalization technique that better assist highlight objects and suppress background (Fig.5); (*iii*) the number of levels in the proposed method is much larger than [22] where only 3 scales are considered. It leads to robustness in more generic cases. In addition, our method is adaptive, i.e. no specification/manually determination of scale parameters is needed like [22]. It automatically merges regions sharing weak boundaries in each iteration. Main contributions of our work include:

1. Develop an adaptive merging strategy for salient object detection rather than using several fixed "scales". Our method generates intrinsic optimal "scales" during the merging.

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.

2 Related Work

The term "salient object" (also called "salient region") detection has emerged in light of its usefulness in image understanding and processing tasks [15, 16, 33, 39]. Existing methods attempt to compensate the drawbacks of previous eye-fixation models [2, 14, 19, 31] on two aspects: 1) to enhance the attenuated inner parts of large-scale objects while keeping the entire objects highlighted uniformly. 2) to output full resolution and edge-aware saliency maps. *The closer saliency maps are to binary ground truth masks, the better an algorithm is*. The literature of salient object detection is huge and we refer readers to comprehensive surveys [1, 2]. There are a number of ways to classify existing methods. We classify prior arts in terms of their processing units based on the starting point of this paper.

Pixel-based: The early work of Zhai *et al* [36] computes pixel-wise saliency via global luminance contrast. The technique is later extended to incorporate histogram-based color contrast [17]. Achanta *et al* [23] propose a frequency-tuned method that smooths an image using Gaussian filter first and then measures color distance to the image average. Shi *et al* [13] compute pixel-wise image saliency by aggregating complementary appearance contrast measures with spatial priors. Liu *et al* [28] segment salient objects by aggregating pixel saliency cues in a Conditional Random Field. Cheng *et al* [18] measure saliency by hierarchical soft abstraction. However, the drawback of using pixels as unit may be that simple computation of color contrast [23, 36] is less satisfactory for complex scenes whereas incorporating holistic pixel-wise information like [28] requires heavy computation.

Patch/Region/Superpixel-based: Gopalakrishnan *et al* [29] perform random walks on a graph with patches as nodes. Goferman *et al* [27] combine local feature and global feature to estimate patch saliency in multi-scale. Margolin *et al* [25] define patch distinctness as L1 norm in PCA coordinates and combine it together with color distinctness. However, local patch contrast [25, 27] can cause edges highlighted. Besides, patches are less well on edge-preserving rendering since they may contain edges or large color variation inside. To overcome this disadvantage of patches, tremendous efforts have focused on pre-segmentation techniques to obtain edge-aware superpixels/regions and shown success in eliminating unnecessary details and producing high quality saliency detection [8]. Examples in this category include saliency filters [8], color contrast and distribution based methods [11], the Bayesian framework [35], the geodesic approaches [12, 34], sparse and low rank matrix framework [32], manifold ranking [3], region contrast [17], region based saliency refinement [11, 23]. Despite these efforts, as aforementioned in Section 1, a small local segment alone hardly reflect global meanings.

Multi-scale based: Since over-partitioned segments have limited capabilities in modeling holism properties as shown in Fig.1, a number of latest approaches employ multi-scale segmentation schemes to extract non-local contrast information. Yan *et al* [22] merge regions according to user-defined scales (e.g., 3 size scales in their case) to eliminate small-size distracters. Jiang *et al* [9] learn several optimal scales from a series of manually defined scales which are quantized by segmentation parameter k in [21] that controls the extent of partition. By contrast our method embeds saliency estimation in an adaptive region fusion framework. It could effectively incorporate global cues and does not require sophisticated feature extraction and learning process [9].

3 Adaptive Multi-Level Region Merging

3.1 The Big Picture

As shown in Fig.2, our framework first performs over-segmentation on an input image by using SLIC superpixels [24], 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 [21] that sets out to merge regions by quantifying a pre-defined region comparison criterion. The rationale behind the method is that a salient object often presents high local or global contrast that contributes to *salient (i.e. strong) edges* (see 4.1). Specifically before merging starts, a globalization procedure is proposed and conducted to pop out salient contours and 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 integrating (summing) across-level saliency maps (Fig.2).

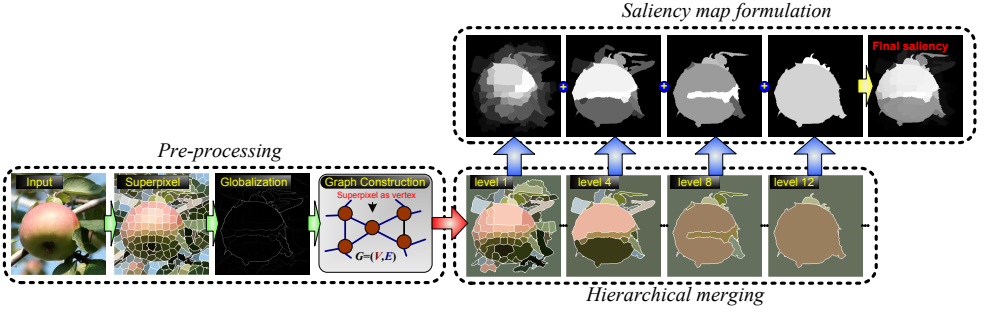

Figure 2: The processing pipeline of our approach.

3.2 Adaptive Region Merging Strategy

Let initial SLIC superpixels be R_i^0 and the corresponding average colors in CIE*Lab* color space be $\mathbf{c}_i^0$, $i = 1,2,...,N$. $N \approx 200$ superpixels are used for each input image. Let a graph $G = (V,E)$ be defined where vertices V are superpixels, and E are graph edges. Let $R^l = \{R_1^l, R_2^l, ...\}$ be a partition of V in the lth level and $R_k^l \in R^l$ corresponds to its kth part (namely region). With the constructed graph 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 $R_i^l, R_j^l, R_k^l, ...$ are regions that have been merged into one larger region R_{new}^l at this level, we then update $R^l \leftarrow (R^l / \{R_i^l, R_j^l, R_k^l, ...\}) \bigcup R_{new}^l$ ("/" and "$\bigcup$" 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 and the merging continues as above.

The proposed "merging and adapting" procedure continues until all regions in R^l are merged together, i.e., $|R^l| = 1$. The step size is fixed as $T_s = (e_{max} - e_{min})/n$ in all experiments, where e_{max}, e_{min} are the maximum and minimum graph edge, respectively. n is the "quantifying number" determined empirically. Practically $n = 30$ suffices (see 4.3).

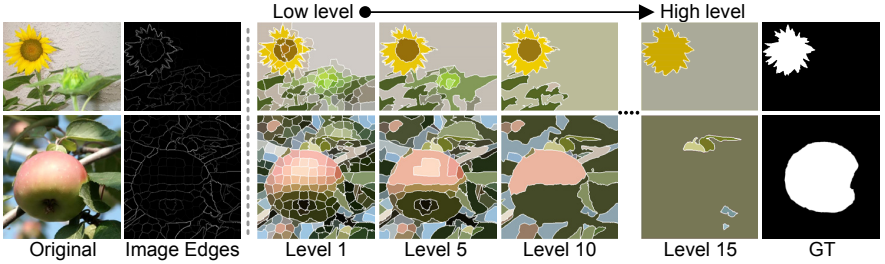

Figure 3: The intermediate results of our merging process on two sample images from the MSRA-1000 database [23]. In this illustration, the graph edges are equal to the adjacent color appearance differences between superpixels (i.e. image edges).

The function of threshold Th is described as: hierarchical segmentation results that abstract away trivial details on different levels are obtained by increasing Th. Since image edges between salient object and background are strong (Fig.5), no merging between them is performed. As a result, most strong edges could be retained till high levels.

It is also important to mention that our extensions to [21] are critical. First, the *mean function* in (1) is used rather than *min function* in [21]. This change keeps holistic contours stay longer during merging as it will better preserve the global edge contrast of regions. This is because "min" can be easily affected by some ambiguous boundaries, i.e. as long as a small portion of object boundaries are weak, the entire object will be merged into background immediately due to "min" effect. A downside of using the mean function is that it makes the merging problem NP-hard [21]. Luckily, the number of superpixel (N) is rather small (≈ 200 superpixels) and is not a bottleneck in computation. Second, our threshold Th is much simpler and easier to change than the parameter k in [21] which penalizes segmentation w.r.t. regions' area and needs to be modified nonlinearly through hierarchy. Above two changes are possible because our goal is different from [21]: we aim at adaptive region merging for salient object detection whereas [21] aims at a single level segmentation that is neither too coarse nor too fine.

Fig.3 shows the merging process on two sample images from MSRA-1000 [23]. Salient object contours are popped out gradually in the merging process. Note though the merging procedure is exploited to keep strong edges, object contours may still be destroyed at the final stage (see the second row in Fig.3). However before that, contours at several coarser levels are fully extracted and may be appropriate to analyze image's gist content. Our merging strategy differs from [22] that users are no longer required to set scale parameters directly since the number of scales is intrinsically determined according to image content.

3.3 Construction of Edges in the Graph

Given two adjacent superpixels R_i^0 and R_j^0, a straightforward way to construct E is to use image edges, approximated by $e_{ij} = ||\mathbf{c}_i^0 - \mathbf{c}_j^0||_2$, where $\mathbf{c}_i^0, \mathbf{c}_j^0$ are superpixels' color vectors (Fig.3). However, such pair-wise difference for graph construction is only local and can lead to inconsistent object contours (e.g., leaking edges) during merging. For example, due to some weak image edges, salient objects may be potentially merged with the background more easily (2nd row in Fig.3).

To address this issue, we further propose to use a globalization procedure inspired by a contour detector gPb [20]. The technique achieves area completion by solving the eigen-problem on the local affinity matrix. Note this operation also meets the Gestalt psychological

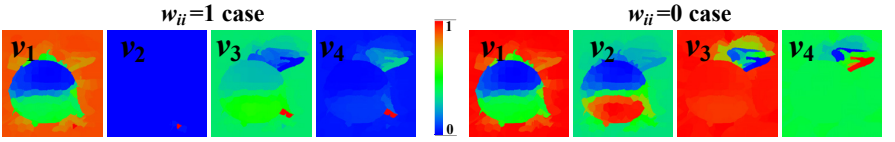

Figure 4: An illustration of the first four eigenvectors (translated to Pseudo-color images) from both cases (i.e. $w_{ii} = 1/0$). Setting the diagonal of the $\mathbf{W}$ to 0 captures relatively large homogenous regions with salient edges.

laws properties [37, 38] such as *closure* and *connectivity* based on which human perceive figures. We define entries of the graph affinity matrix $\mathbf{W}$ by using image edges:

$$w_{ij} = \begin{cases} \exp(-\alpha||\mathbf{c}_i^0 - \mathbf{c}_j^0||_2) & \text{if } R_i^0 \text{ and } R_j^0 \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where the α controls affinity. Since we can first normalized all image edges to interval $[0,1]$, $\alpha = 10$ is set empirically in (2). We then solve for the generalized eigenvectors $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{nvec}$ (correspond to $nvec+1$ smallest eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{nvec}$) of system:

$$(\mathbf{D} - \mathbf{W})\mathbf{v} = \lambda \mathbf{D}\mathbf{v} \tag{3}$$

where $\mathbf{D}$ is the degree matrix of $\mathbf{W}$. λ, $\mathbf{v}$ are eigen-value and vector to be solved.

Since eigenvectors of the smallest eigenvalues correspond to different clusters [5, 10], they carry contour information in our case where $\mathbf{v}$ is an image. Note self-reinforcement is avoided by setting $\mathbf{W}$'s diagonal entries $w_{ii} = 0$ rather than 1 [6]. This operation enables eliminating the case where isolated superpixels are detected as tight clusters in the solved eigenvectors, and hence can remove small sized distracters with isolated superpixels. Illustration for this is shown in Fig.4. By treating the decomposed eigenvectors as images (Fig.4), we compute the graph edge e_{ij} between two adjacent R_i^0 and R_j^0 by integrating the deviation from the $nvec$ smallest non-zero eigenvalues $\lambda_1, \lambda_2, ..., \lambda_{nvec}$ along with their eigenvectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{nvec}$ as:

$$e_{ij} = \sum_{k=1}^{nvec} \frac{1}{\sqrt{\lambda_k}} |\mathbf{v}_k(R_i^0) - \mathbf{v}_k(R_j^0)| \tag{4}$$

where $\mathbf{v}_k(R_i^0)$ indicates the value in eigenvector $\mathbf{v}_k$ corresponding to superpixel R_i^0. The weighting by $1/\sqrt{\lambda_k}$ is motivated by the physical interpretation of the generalized eigenvalue problem as a mass-spring system [20]. In practice, $nvec = 8$ suffices while further increasing it introduces extra noise. Compared with that in [20], our global procedure introduces superpixels to replace pixels, thus reducing the dimension of $\mathbf{W}$ from hundreds of thousands to only hundreds (more efficient in both time and memory). An example result after globalization is shown in Fig.2, which can be compared with that in Fig.3.

3.4 Simple Region Saliency Measurements

To show the effectiveness of the proposed region merging and integration scheme, each merged region is just evaluated using several simple region saliency measurements, though more complex features and measurements as in [9] can be adopted. Even though like this, we show the proposed method already can achieve competitive results against the best methods among the state-of-the-art.

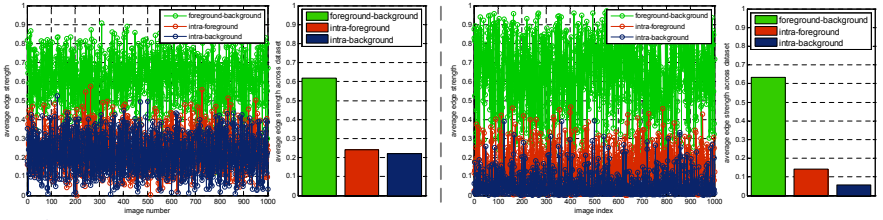

Figure 5: Left: average edge strength (computed via adjacent deviation) of foreground-background, intra-foreground and intra-background on 1000 images from the MSRA-1000 dataset [23] and the average edge strength over the complete dataset. Right: edge strength statistics after applying globalization.

Figure-ground contrast: As the survived regions are characterized by strong boundaries and likely to be "figures" [37, 38]. We simply compute the figure-ground contrast as a region's total color distance towards all superpixels in the four image boundaries (deemed as "ground"). A similar measurement is also used in [9] termed as "backgroundness".

Center bias: Statistical results in [1, 40] show that human attention occurs center bias, indicating that distinctive regions close to image center are likely to be salient. Therefore we weigh regions by using their location with respect to the image center. Instead of using the position of region centroid, we mask a Gaussian distribution which is located at the image center and average the probability values lying in each region. A similar measurement is used in [22] termed as "location heuristic".

We multiply scores of the above two measurements to obtain a combined saliency score for each region. Further, we notice that homogenous regions that touch image boundaries usually belong to background [34]. To effectively suppress such regions, we prune scores of regions touching more than one out of the four image borders to zero. After that, the saliency score of each region is assigned to the corresponding superpixels to formulate an intermediate saliency map (Fig.2).

4 Experiments and Results

We comprehensively compare our scheme with the state-of-the-art methods. The following metrics are used for evaluation: Precision-Recall, F-measure [1, 8, 18, 23], Mean Absolute Error (MAE) [8]. Benchmark datasets for evaluation include commonly used MSRA-1000 [23](1000 images), SOD [30] (300 images) and SED [26] which consists of two parts, i.e. SED1 (one object set) and SED2 (two objects set) each containing 100 images. We compare our technique with state-of-the-art salient region detection methods: CA (Context Aware) [27], FT (Frequency Tuned) [23], LC (Luminance Contrast) [36], HC (Histogram Contrast) [17], RC (Region Contrast) [17], SF (Saliency Filter) [8], LR (Low Rank) [32],GS (Geodesic Saliency) [34], HS (Hierarchical Saliency) [22], PCA [25], DRFI (Discriminative Regional Feature Integration) [9], GC (Global Cue) [18], MR (Manifold Ranking) [3]. Note we do not compare with eye fixation models such as Itti's [14] and Hou's [31] due to different aims.

4.1 Validation of Edge Hypothesis and Globalization Scheme

To validate the assumption that salient objects are enclosed by "*salient edges*", we compute the average foreground-background, inner-foreground and inner-background adjacent deviation on MSRA-1000 [23]. A superpixel is considered to belong to the foreground if more than half of its area covers the ground truth foreground object mask. Note all the deviation values in each image are first normalized to [0,1] and then averaged. Statistical results are

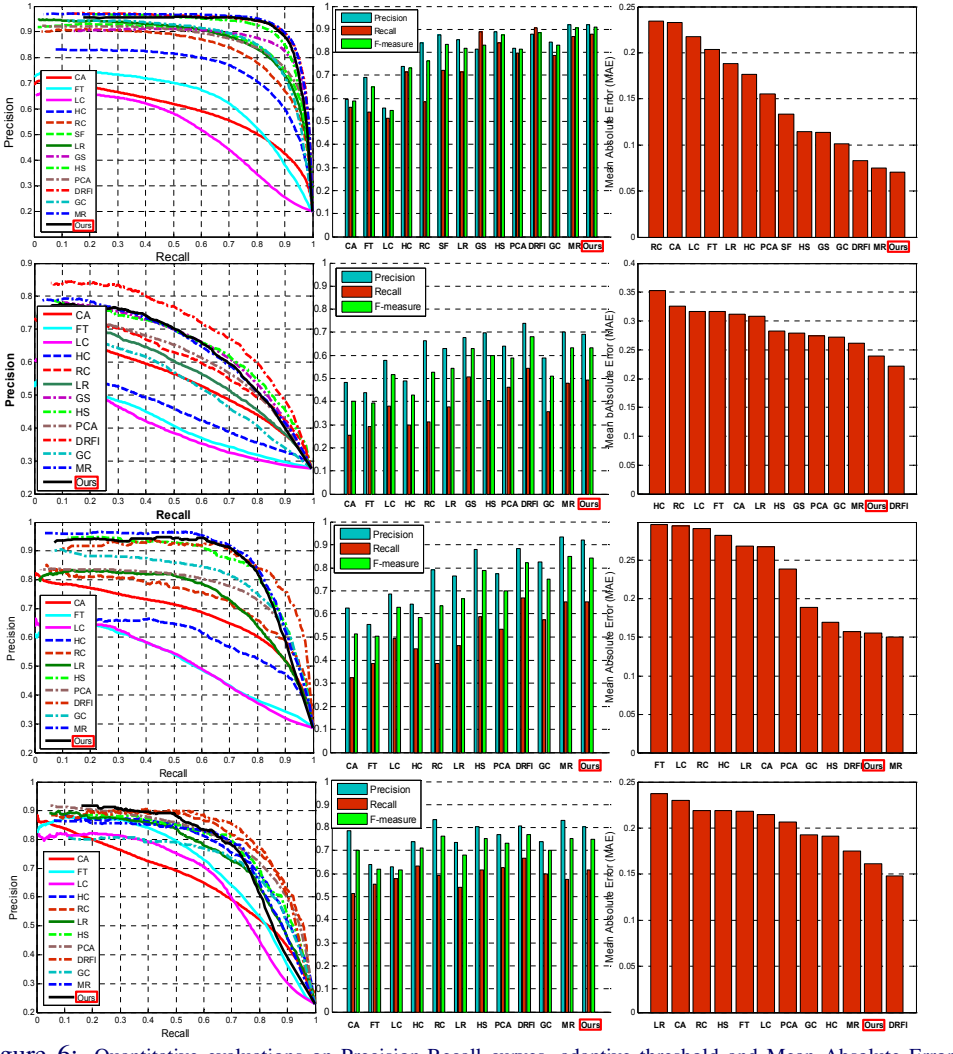

Figure 6: Quantitative evaluations on Precision-Recall curves, adaptive threshold and Mean Absolute Error (MAE) on three benchmark datasets: from top to bottom are MSRA-1000, SOD, SED1, and SED2. Note because SF only provides results on MSRA-1000 while GS only provides results on MSRA-1000 and SOD. We can not compare with them on the rest sets.

shown in the left sub-figure of Fig.5. The average foreground-background adjacent deviation is consistently much higher than inner-foreground and inner-background ones, which implies that *the merging process would bias towards regions within foreground or background rather than across the two.* Fig.5 right shows results after using globalization procedure, where intra-background edges are drastically suppressed, i.e. the background regions can be merged more easily.

4.2 Comparisons with the State-of-the-art Methods

Precision-Recall curves generated by using fixed threshold from 0 to 255 are shown in Fig.6. The performance of our method is comparable to the most recent state-of-the-art techniques,

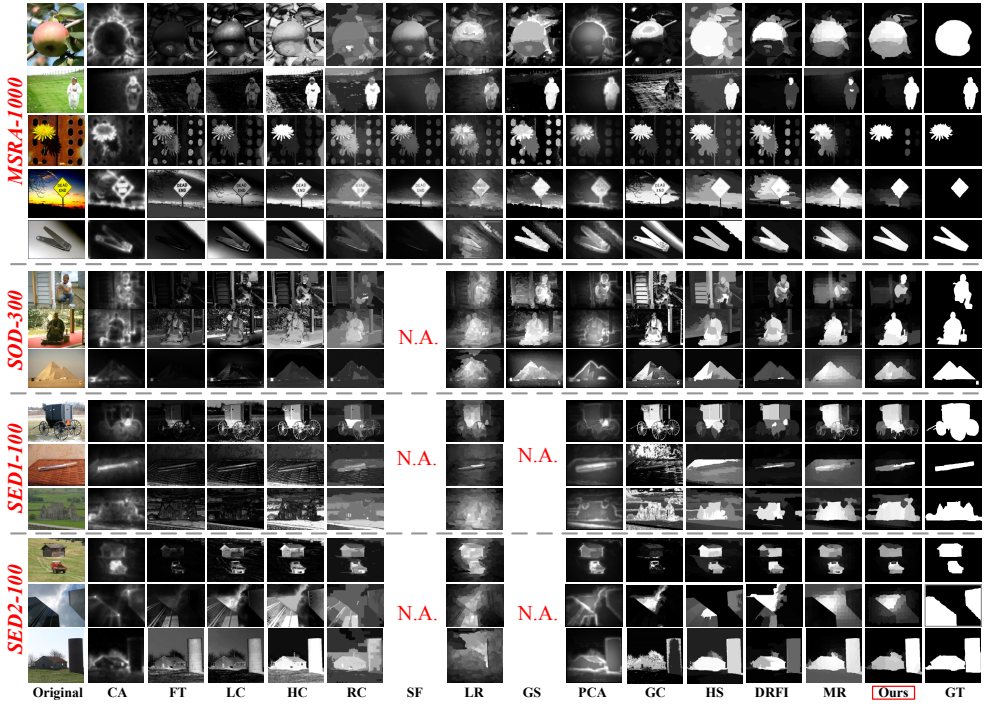

Figure 7: Visual comparisons on three benchmark datasets with 13 state-of-the-art methods. "N.A." means neither results nor code are publicly available for a certain method. Results generated by our method are closer to the ground truth as well as consistent with human perceptions.

e.g. outperforms HS [22] on MSRA-1000 and SED1 and achieves similar results on the rest. Besides, our method is comparable to DRFI [9] and MR [3], among which [9] adopts sophisticated learning and [3] uses two-stage refinement. In the adaptive threshold experiment [3, 8, 18, 23], our method achieves both the highest precision and F-measure on MSRA-1000, 3rd and 2nd F-measure on SOD and SED1. For SED2 whose images contain two objects, our method performs similar to MR [3]. Note in SED2 since many objects labeled violate the boundary prior (e.g. 13th row in Fig.7), both our method and MR perform less well. In such cases, it is better to keep a vague detection using only contrast. That is why RC and HC perform better than before.

To further evaluate the methods, we compute the MAE criterion [8]. As shown in Fig.6, our method produces the lowest error on MSRA-1000, and consistently 2nd on the rest, indicating our robustness against varied datasets. Note lower MAE means closer to the binary ground truth. In contrast, despite their good performance in PR curve and F-measure, RC [17], HC [17], and LR [32] achieve the highest error due to weak background suppression abilities (also can be observed from Fig.7).

Fig.7 shows visual comparisons on three datasets. Our method effectively suppresses the background clutter and uniformly emphasizes the foreground objects, attributed mainly to our hierarchical region merging strategy. Further globalization helps to pop out holistic salient contours. In addition, the proposed method is able to deal with images containing "color ramps". Such effects are usually caused by shadow or lighting conditions (4-5th rows in Fig.7). Our hierarchical merging scheme effectively combines them into background,

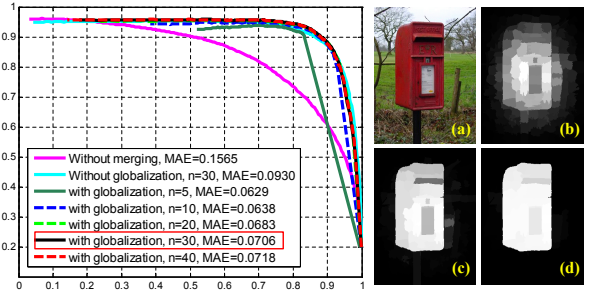

Figure 8: Left: Quantitative comparisons among other alternatives including simple saliency computation without merging, without globalization, and performance under varied quantifying number n. Right: (a) original, (b) without merging, (c) without globalization, (d) full version of our method.

preserving perceptual homogeneity. In contrast, the contrast-based GC [13], SF [8] and geodesic based GS [34] methods that use over-partitioned image segments are less better due to color heterogeneity.

Our method also handles challenging cases that cause the state-of-the-art methods to fail. For example, a key procedure in MR [3] is intermediate thresholding and re-propagation to refine the results (called "second stage" in [3]). The operation is critical in achieving high performance. Since this operation depends highly on the threshold, once isolated cluttered regions are segmented out, they can hardly be absorbed into the background even with the help of propagation, e.g., the shadow of the flower in 3rd row of Fig.7.

4.3 Merging and Globalization

We test our method without the merging procedure (only conduct saliency detection introduced in subsection 3.4 on initial superpixels) and without globalization (construct the graph edges using image edges as demonstrated in subsection 3.3) on MSRA-1000. Fig.8 shows significant performance degradation on both PR and MAE without merging or globalization, e.g., the MAE drops from 0.093 to 0.071 (over 20%). Recall since the PR curve is insensitive to maps' uniformness, it appears the same with and without globalization. We also find that our method is robust to the "quantifying number" n (Fig.8). When using an $n > 30$, our method produces similar results. Nevertheless, our method outperforms the state-of-the-art in MAE under varied n. Since larger n results in more levels to compute, we set $n = 30$ for a good PR curve although precision can be sacrificed for speed.

5 Conclusion

We have presented a new salient object detection scheme based on adaptive multi-level region merging. The core of our method is adaptive region merging and globalization. The former combines potential foreground and background regions and the latter improves contour completions. When combined together, they greatly improve the accuracy on detecting holistic objects and effectively suppress the background. Experiments have shown our method achieves state-of-the-art performance on three commonly used benchmark datasets.

Acknowledgments: This work 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.

References

[1] A. Borji et al. Salient object detection: A benchmark. In *ECCV*, 2012.
[2] A. Borji et al. State-of-the-art in visual attention modeling. *TPAMI*, 35(1):185–207, 2013.
[3] C. Yang et al. Saliency detection via graph-based manifold ranking. In *CVPR*, 2013.
[4] D. Comaniciu et al. Mean shift: a robust approach toward feature space analysis. *TPAMI*, 24(5):603–619, 2002.
[5] D. Tolliver et al. Graph partitioning by spectral rounding: Applications in image segmentation and clustering. In *CVPR*, 2006.
[6] D. Zhou et al. Learning with local and global consistency. In *NIPS*, 2003.
[7] E. Rahtu et al. Segmenting salient objects from images and videos. In *ECCV*, 2010.
[8] F. Perazzi et al. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, 2012.
[9] H. Jiang et al. Salient object detection: A discriminative regional feature integration approach. In *CVPR*, 2013.
[10] J. Shi et al. Normalized cuts and image segmentation. *TPAMI*, 22(8):888–905, 2000.
[11] K. Fu et al. Salient object detection via color contrast and color distribution. In *ACCV*, 2012.
[12] K. Fu et al. Geodesic saliency propagation for image salient region detection. In *ICIP*, 2013.
[13] K. Shi et al. Pisa: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors. In *CVPR*, 2013.
[14] L. Itti et al. A model of saliency-based visual attention for rapid scene analysis. *TPAMI*, 20(11):1254–1259, 1998.
[15] L. Marchesotti et al. A framework for visual saliency detection with applications to image thumbnailing. In *ICCV*, 2009.
[16] L. Wang et al. Automatic salient object extraction with contextual cue. In *ICCV*, 2011.
[17] M. Cheng et al. Global contrast based salient region detection. In *CVPR*, 2011.
[18] M. Cheng et al. Efficient salient region detection with soft image abstraction. In *ICCV*, 2013.
[19] N. Bruce et al. Saliency based on information maximization. In *NIPS*, 2005.
[20] P. Arbelaez et al. Contour detection and hierarchical image segmentation. *TPAMI*, 33(5):898–916, 2010.
[21] P. Felzenszwalb et al. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, 2004.
[22] Q. Yan et al. Hierarchical saliency detection. In *CVPR*, 2013.
[23] R. Achanta et al. Frequency-tuned salient region detection. In *CVPR*, 2009.
[24] R. Achanta et al. Slic superpixels compared to state-of-the-art superpixel methods. *TPAMI*, 34(11):2274–2282, 2012.
[25] R. Margolin et al. What makes a patch distinct. In *CVPR*, 2013.
[26] S. Alpert et al. Image segmentation by probabilistic bottom-up aggregation and cue integration. In *CVPR*, 2007.
[27] S. Goferman et al. Context-aware saliency detection. In *CVPR*, 2010.
[28] T. Liu et al. Learning to detect a salient object. *TPAMI*, 33(2):353–367, 2011.
[29] V. Gopalakrishnan et al. Random walks on graphs for salient object detection in images. *TIP*, 19(12):3232–3242, 2010.
[30] V. Movahedi et al. Design and perceptual validation of performance measures for salient object segmentation. In *IEEE Computer Society Workshop on Perceptual Organization in Computer Vision*, 2010.
[31] X. Hou et al. Saliency detection: A spectral residual approach. In *CVPR*, 2007.
[32] X. Shen et al. A unified approach to salient object detection via low rank matrix recovery. In *CVPR*, 2012.
[33] Y. Ding et al. Importance filtering for image retargeting. In *CVPR*, 2011.
[34] Y. Wei et al. Geodesic saliency using background priors. In *ECCV*, 2012.
[35] Y. Xie et al. Visual saliency detection based on bayesian model. In *ICIP*, 2011.
[36] Y. Zhai et al. Visual attention detection in video sequences using spatiotemporal cues. *ACM Multimedia*, pages 815–824, 2006.
[37] K. Koffka. Principles of gestalt psychology. 1935.
[38] S. Palmer. Vision science: Photons to phenomenology. *The MIT press*, 1999.
[39] F. Stentiford. Attention based auto image cropping. In *Workshop on Computational Attention and Applications, ICVS*, 2007.
[40] B. Tatler. The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor bases and image feature distributions. *JoV*, 14(7), 2007.