Geodesic Saliency Using Background Priors

Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun

Microsoft Research Asia
{yichenw,fangwen,v-wazhu,jiansun}@microsoft.com

Abstract. Generic object level saliency detection is important for many vision tasks. Previous approaches are mostly built on the prior that "appearance contrast between objects and backgrounds is high". Although various computational models have been developed, the problem remains challenging and huge behavioral discrepancies between previous approaches can be observed. This suggest that the problem may still be highly ill-posed by using this prior only.

In this work, we tackle the problem from a different viewpoint: we focus more on the background instead of the object. We exploit two common priors about backgrounds in natural images, namely *boundary and connectivity priors*, to provide more clues for the problem. Accordingly, we propose a novel saliency measure called *geodesic saliency*. It is intuitive, easy to interpret and allows fast implementation. Furthermore, it is complementary to previous approaches, because it benefits more from background priors while previous approaches do not.

Evaluation on two databases validates that geodesic saliency achieves superior results and outperforms previous approaches by a large margin, in both accuracy and speed (2 ms per image). This illustrates that appropriate prior exploitation is helpful for the ill-posed saliency detection problem.

1 Introduction

The human vision system can rapidly and accurately identify important regions in its visual field. In order to achieve such an ability in computer vision, extensive research efforts have been conducted on bottom up visual saliency analysis for years. Early works [1, 2] in this field are mostly based on biologically inspired models (where a human looks) and is evaluated on human eye fixation data [3, 4]. Many follow up works are along this direction [5–7].

Recent years have witnessed more interest in object level saliency detection (where the salient object is) [8–13] and evaluation is performed on human labeled objects (bounding boxes [8] or foreground masks [9]). This new trend is motivated by the increasing popularity of salient object based vision applications, such as object aware image retargeting [14, 15], image cropping for browsing [16], object segmentation for image editing [17] and recognition [18]. It has also been shown that the low level saliency cues are helpful to find generic objects irrespective to their categories [19, 20]. In this work we focus on the object level saliency detection problem in natural images.

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part III, LNCS 7574, pp. 29–42, 2012.

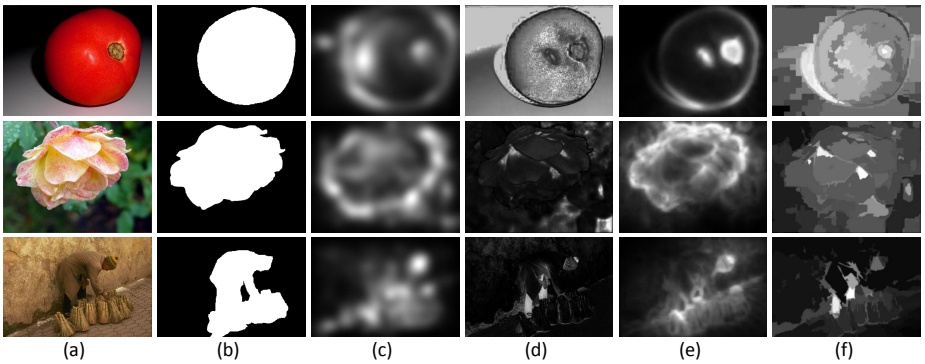

Fig. 1. Saliency maps of previous representative contrast prior based methods, on three example images of increasing complexity in objects and backgrounds. (a) Input images. (b) Ground truth salient object masks. (c)-(f) Results from one local method [1] and three global methods [9, 11, 12].

1.1 Contrast Prior in Previous Work

Due to the absence of high level knowledge, all bottom up saliency methods rely on assumptions or priors[1] on the properties of objects and backgrounds. Arguably, the most fundamental assumption is that "appearance contrast between object and background is high". This assumption states that a salient image pixel/patch presents high contrast within a certain context. It is intuitive and used in all saliency methods, explicitly or implicitly. In this paper we call it *contrast prior* for conciseness.

Depending on the extent of the context where the contrast is computed, previous methods can be categorized as local methods [1, 6, 8, 7, 10] or global methods [21, 9, 22, 11, 12]. *Local methods* compute various contrast measures in a local neighborhood of the pixel/patch, such as edge contrast [8], center-surround discriminative power [7], center-surround differences [1, 8, 13], curvature [10] and self information [6].

Global methods use the entire image to compute the saliency of individual pixels/patches. Some methods assume globally less frequent features are more salient and use frequency analysis in the spectral domain [21, 9]. Other methods compare each pixel/patch to all the others in the image and use the averaged appearance dissimilarity as the saliency measure [22, 11, 12], and the averaging is usually weighted by spatial distances between pixels/patches to take into account the fact that salient pixels/patches are usually grouped together to form a salient object.

Contrast prior based methods have achieved success in their own aspects, but still have certain limitations. Typically, the boundaries of the salient object can be found well, but the object interior is attenuated. This "object attenuation" problem is observed in all local methods and some global methods. It is alleviated

[1] We use "assumption" and "prior" interchangeably in this paper.

in global methods [11, 12], but these methods still have difficulties of highlighting the entire object uniformly.

In essence, the saliency object detection problem on general objects and background is highly ill-posed. There still lacks a common definition of "what saliency is" in the community, and simply using *contrast prior* alone is unlikely to succeed. While previous approaches are mostly based on their own understanding of "how the contrast prior should be implemented", huge behavioral discrepancies between previous methods can be observed, and such discrepancies are sometimes hard to understand.

This problem is exemplified in Figure 1, which shows results from previous representative methods, one local method [1] and three global methods [9, 11, 12]. The three examples present increasing complexities in objects and backgrounds: a simple object on a simple background; a more complex object on a more complex background; two salient regions on a complex background with low contrast. It is observed that the results from different methods vary significantly from each other, even for the first extremely simple example, which would not be ambiguous at all for most humans. This phenomenon is repeatedly observed in many images, either simple or complex. This consolidates our belief that using the contrast prior alone is insufficient for the ill-posed problem and more clues should be exploited.

1.2 Background Priors in Our Approach

We tackle this problem from a different direction. In addition to asking "what the salient object should look like", we ask the opposite question "what the background should look like". Intuitively, answering this question would help removing background clutters and in turn lead to better foreground detection. Let us look at the first image in Figure 1 as an example. The entire background is a large and smoothly connected region, and it should not be considered as foreground in any way, that is, its saliency should always be zero.

We propose to use two priors about common backgrounds in natural images, namely *boundary and connectivity priors*. They are not fully exploited in previous methods and are helpful to resolve previous problems that are otherwise difficult to address.

The first prior comes from the basic rule of photographic composition, that is, most photographers will not crop salient objects along the view frame. In other words, the image boundary is mostly background. We call it the *boundary prior*. It is more general than the previously used *center prior* [8, 3] that "the image center is more important", because salient objects can be placed off the center (one-third rule in professional photography), but they seldom touch the image boundary. This prior shares a similar spirit with the *bounding box prior* used in interactive image segmentation [17, 23], where the user is asked to provide a bounding box that encloses the object of interest. This prior is validated on several salient object databases, as discussed in Section 4.1.

The second prior is from the appearance characteristics of real world backgrounds in images, that is, background regions are usually large and homogeneous.

In other words, most image patches in the background can be easily connected to each other. We call it the *connectivity prior*. Note that the connectivity is in a piecewise manner. For example, sky and grass regions are homogeneous by themselves, but patches between them cannot be easily connected. Also the background appearance homogeneity should be interpreted in terms of human perception. For example, patches in the grass all look similar to humans, although their pixel-wise intensities might be quite different. This prior is different from the connectivity prior used in object segmentation [24, 25], which is assumed on the spatial continuity of the object instead of the background. Sometimes this prior is also supported by the fact that background regions are usually out of focus during photography, and therefore are more blurred and smoother than the object of interest.

2 Geodesic Saliency

Based on the contrast and the two background priors, we further observed that *most background regions can be easily connected to image boundaries while this is much harder for object regions*. This suggests that we can define the saliency of an image patch as the length of its shortest path to image boundaries. However, this assumes that all the boundary patches are background, which is not realistic enough as the salient object could be partially cropped on the boundary. This can be rectified by adding a virtual background node connected to all boundary patches in the image graph and computing the saliency of boundary patches, as described in Section 2.1. In this way, we propose a new and intuitive *geodesic saliency* measure, that is, *the saliency of an image patch is the length of its shortest path to the virtual background node*.

Figure 2 shows several shortest paths on image patches and the geodesic saliency results. These results are more plausible than those in Figure 1.

Geodesic saliency has several advantages. Firstly, it is intuitive and easy to interpret. It simultaneously exploits the three priors in an effective manner. Previous "object attenuation" problem is significantly alleviated because all patches inside a homogeneous object region usually share similar shortest paths to the image boundary and therefore have similar saliency. As a result, it achieves superior results than previous approaches.

Secondly, it mostly benefits from background priors and uses the contrast prior moderately. Therefore, it is complementary to methods that exploit the contrast prior in a more sophisticated manner [22, 11, 12]. A combination of geodesic saliency with such methods can usually improve both.

Last, it is easy to implement, fast and suitable for subsequent applications. For different practical needs in accuracy/speed trade off, we propose two variants of the geodesic saliency algorithm. For applications that require high speed, such as interactive image retargeting [14], image thumbnail generation/cropping for batch image browsing [16], and bounding box based object extraction [26, 19], we propose to compute geodesic saliency using rectangular patches on an image grid and an approximate shortest path algorithm [27]. We call this algorithm *GS*

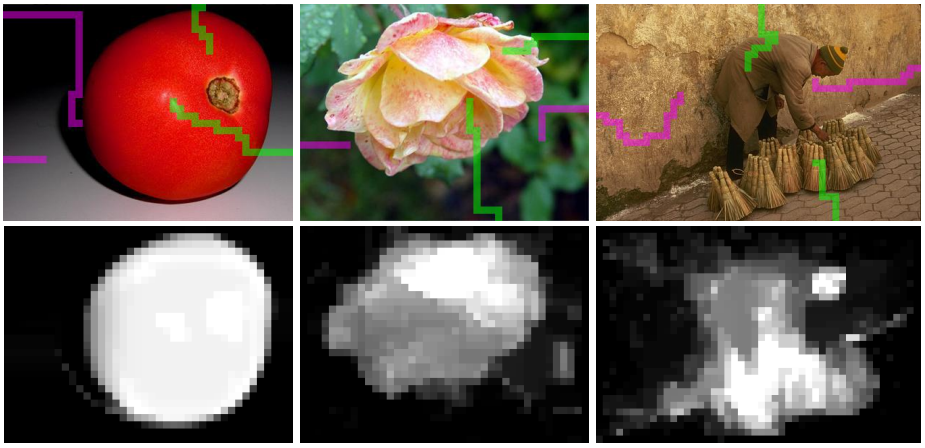

Fig. 2. (better viewed in color) Illustration of geodesic saliency (using regular patches). Top: three example images in Figure 1 and shortest paths of a few foreground (green) and background (magenta) image patches. Bottom: geodesic saliency results.

Grid. It runs in about 2 milliseconds for images of moderate size (400×400). For applications that require high accuracy such as object segmentation [17, 18], we propose to use superpixels [28] as image patches and an exact shortest path algorithm. We call this algorithm *GS Superpixel*. It is slower (a few seconds) but more accurate than *GS Grid*.

In Section 4, extensive evaluation on the widely used salient object databases in [8, 9] and a recent more challenging database in [29, 30] shows that geodesic saliency outperforms previous methods by a large margin, and it can be further improved by combination with previous methods. In addition, the *GS Grid* algorithm is significantly faster than all other methods and suitable for real time applications.

2.1 Algorithm

For an image, we build an undirected weighted graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. The vertices are all image patches $\{P_i\}$ plus a virtual background node B, $\mathcal{V} = \{P_i\} \cup \{B\}$. There are two types of edges: internal edges connect all adjacent patches and boundary edges connect image boundary patches to the background node, $\mathcal{E} = \{(P_i, P_j)|P_i \text{ is adjacent to } P_j\} \cup \{(P_i, B)|P_i \text{ is on image boundary}\}$. The geodesic saliency of a patch P is the accumulated edge weights along the shortest path from P to background node B on the graph $\mathcal{G}$,

$$saliency(P) = \min_{P_1=P, P_2, \ldots, P_n=B} \sum_{i=1}^{n-1} weight(P_i, P_{i+1}), s.t.(P_i, P_{i+1}) \in \mathcal{E}.$$

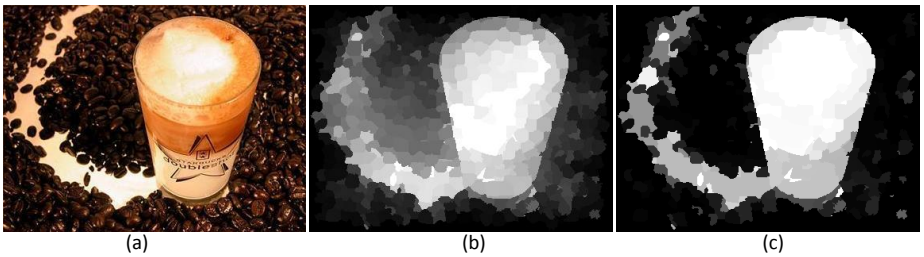

Fig. 3. Internal edge weight clipping alleviates the small weight accumulation problem. (a) Input image. (b) Geodesic saliency (using superpixels) without weight clipping. (c) Geodesic saliency (using superpixels) with weight clipping.

We firstly describe how to compute edge weights, and then introduce two variants of geodesic saliency algorithm: *GS Grid* is faster and *GS Superpixel* is more accurate.

Internal edge weight is the appearance distance between adjacent patches. This distance measure should be consistent with human perception of how similar two patches look alike. This problem is not easy in itself. For homogeneous textured background such as road or grass, while patches there look almost identical, simple appearance distances such as color histogram distance are usually small but non zero values. This causes the *small-weight-accumulation* problem, that is, many small weight edges can accumulate along a long path and form undesirable high saliency values in the center of the background. See Figure 3(b) for an example.

Instead of exploring complex features and sophisticated patch appearance distance measures, we take a simple and effective *weight clipping* approach to address this problem. The patch appearance distance is simply taken as the difference (normalized to $[0, 1]$) between the mean colors of two patches (in LAB color space). For each patch we pick its smallest appearance distance to all its neighbors, and then we select an "insignificance" distance threshold as the average value of all such smallest distances from all patches. If any distance is smaller than this threshold, it is considered insignificant and clipped to 0. Such computation of internal edge weights is very efficient and its effectiveness is illustrated in Figure 3.

Boundary edge weight characterizes how likely a boundary patch is not background. When the boundary prior is strictly valid, all boundary patches are background and such weights should be 0. However, this is too idealistic and not robust. The whole object could be missed even if it only slightly touches the image boundary, because all the patches on the object may be easily connected to the object patch on the image boundary. See Figure 4(b) for examples.

We observe that when the salient object is partially cropped by the image boundary, the boundary patches on the object are more salient than boundary patches in the background. Therefore, the boundary edge weights computation is treated as a one-dimensional saliency detection problem: given only image

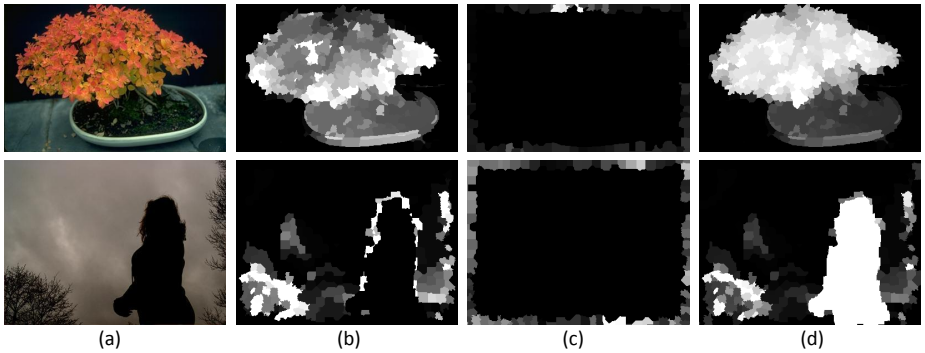

Fig. 4. Boundary edge weight computation makes boundary prior and geodesic saliency more robust. (a) Input image. (b) Geodesic saliency with boundary edge weights as 0. (c) The boundary edge weights computed using a one dimensional version of the saliency algorithm in [11]. (d) Geodesic saliency using boundary edge weights in (c).

boundary patches, compute the saliency of each boundary patch P_i as the weight of boundary edge (P_i, B). This step makes the boundary prior and geodesic saliency more robust, as illustrated in Figure 4(c)(d).

In principle, any previous image saliency method that can be reduced to a one-dimensional version can be used for this problem. We use the algorithm in [11] because it is also based on image patches and adapting it to using only boundary patches is straightforward. For patch appearance distance in this method, we also use the mean color difference.

The *GS Grid* algorithm uses rectangular image patches of 10×10 pixels on a regular image grid. Shortest paths for all patches are computed using the efficient geodesic distance transform [27]. Although this solution is only approximate, it is very close to the exact solution on a simple graph on an image grid. Because of its linear complexity in the number of graph nodes and sequential memory access (therefore cache friendly), it is extremely fast and also used in interactive image segmentation [31, 32]. Overall, the *GS Grid* algorithm runs in 2 ms for a 400×400 image and produces decent results.

The *GS Superpixel* algorithm uses irregular superpixels as image patches. It is more accurate because superpixels are better aligned with object and background region boundaries than the regular patches in *GS Grid*, and the patch appearance distance computation is more accurate. We use the superpixel segmentation algorithm in [28] to produce superpixels of roughly 10×10 pixels. It typically takes a few seconds for an image. Note that our approach is quite insensitive to the superpixel algorithm so any faster method can be used.

The shortest paths of all image patches are computed by Dijkstra's algorithm for better accuracy. In spite of the super linear complexity, the number of superpixels is usually small (around $1,500$) and this step takes no more than a few milliseconds.

3 Salient Object Databases

The MSRA database [8] is currently the largest salient object database and provides object bounding boxes. The database in [9] is a 1,000 image subset of [8] and provides human labeled object segmentation masks. Many recent object saliency detection methods [8–13, 16] are evaluated on these two databases.

Nevertheless, these databases have several limitations. All images contain only a single salient object. Most objects are large and near the image center. Most backgrounds are clean and present strong contrast with the objects. A few example images are shown in Figure 8.

Recently a more challenging salient object database was introduced [30]. It is based on the well known 300 images Berkeley segmentation dataset [29]. Those images usually contain multiple foreground objects of different sizes and positions in the image. The appearance of objects and backgrounds are also more complex. See Figure 10 for a few example images. In the work of [30], seven subjects are asked to label the foreground salient object masks and each subject can label multiple objects in one image. For each object mask of each subject, a consistency score is computed from the labeling of the other six subjects. However, [30] does not provide a single foreground salient object mask as ground truth.

For our evaluation, we obtain a foreground mask in each image by removing all labeled objects whose consistency scores are smaller than a threshold and combining the remaining objects' masks. This is reasonable because a small consistency score means there is divergence between the opinions of the seven subjects and the labeled object is probably less salient. We tried different thresholds from 0.7 to 1 (1 is maximum consistency and means all subjects label exactly the same mask for the object). We found that for most images the user labels are quite consistent and resulting foreground masks are insensitive to this threshold. Finally, we set the threshold as 0.7, trying to retain as many salient objects as possible. See Figure 10 for several example foreground masks.

4 Experiments

As explained above, besides the commonly used MSRA-1000 [9] and MSRA [8] databases, we also use the Berkeley-300 database [30] in our experiments. The latter is more challenging and has not been widely used in saliency detection yet.

Not surprisingly, we found that all our results and discoveries on the MSRA database are very similar to those found in MSRA-1000. Therefore we omit all the results on the MSRA database in this paper.

4.1 Validation of Boundary Prior

In this experiment, we evaluate how likely the image boundary pixels are background. For each image, we compute the percentage of all background boundary pixels in a band of 10 pixels in width (to be consistent with the patch size in

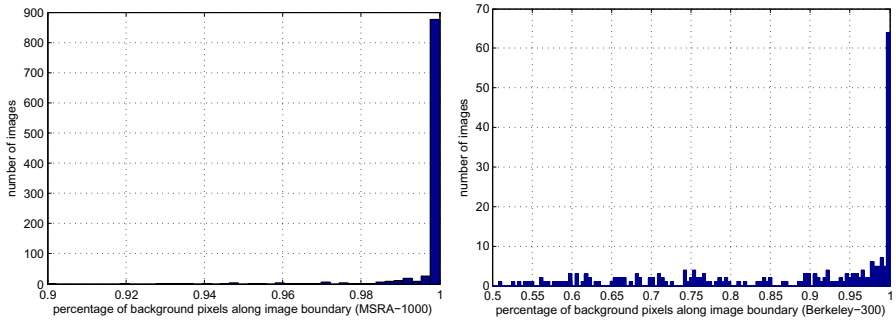

Fig. 5. Distribution of percentage of background pixels along the image boundary on the MSRA-1000 database [9] (left) and Berkeley-300 database [30] (right)

geodesic saliency) along the four image sides, using the ground truth foreground mask. The histograms of such percentages of all the images in MSRA-1000 and Berkeley-300 databases are shown in Figure 5.

As illustrated, the boundary prior is strongly valid in the MSRA-1000 database, which has 961 out of 1000 images with more than 98% boundary pixels as background. The prior is also observed in the Berkeley-300 database, but not as strongly. There are 186 out of 300 images with more than 70% boundary pixels as background. This database is especially challenging for our approach as some images indeed contain large salient objects cropped on the image boundary.

4.2 Evaluation of Geodesic Saliency

We compare the two geodesic saliency methods (GS_GD short for *GS Grid* and GS_SP short for *GS Superpixel*) with eight previous approaches: Itti's method(IT) [1], the spectral residual approach(SR) [21], graph based visual saliency(GB) [22], the frequency-tuned approach(FT) [9], context-aware saliency(CA) [11], Zhai's method(LC) [33], histogram based contrast(HC) and region based contrast(RC) [12]. Each method outputs a full resolution saliency map that is normalized to range $[0, 255]$. Such methods are selected due to their large diversity in computational models. For SR, FT, LC, HC and RC, we use the implementation from [12]. For IT, GB and CA, we use the public implementation from the original authors.

To test whether geodesic saliency is complementary to previous methods, we combine two methods by averaging their results. For example, the combined method "RC+GS_GD" produces a saliency map that is the average of the maps from the RC and GS_GD methods.

Similar as previous work, precision-recall curves are used to evaluate all methods, including the combined methods. Given a saliency map, a binary foreground object mask is created using a threshold. Its precision and recall values are computed with respect to the ground truth object mask. A precision-recall curve is then obtained by varying the threshold from 0 to 255.

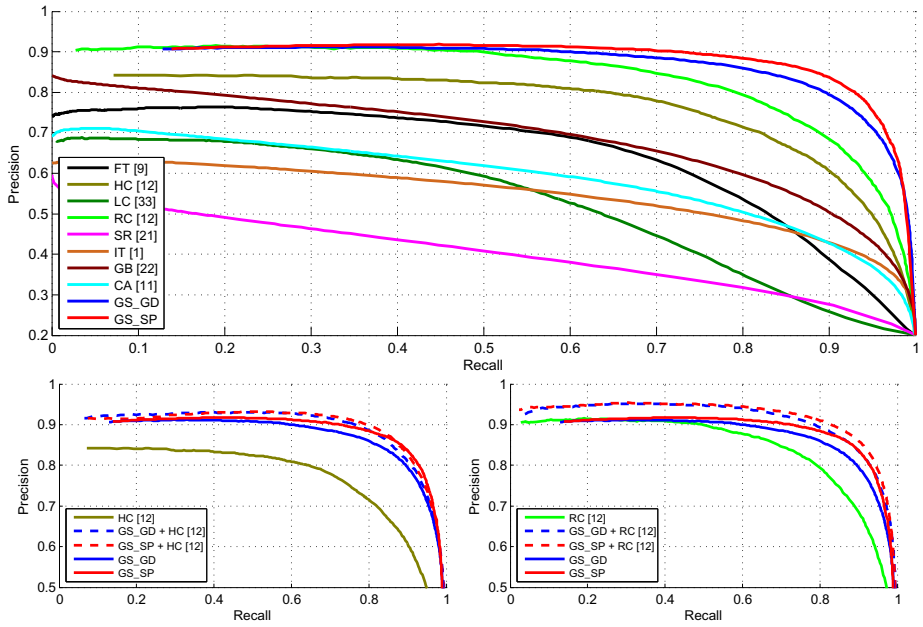

Fig. 6. (better viewed in color) Average precision-recall curves on the MSRA-1000 database [9]. Top: Geodesic saliency is better than all other methods. Bottom: Combination of geodesic saliency with HC(RC) is better than both.

Results on the MSRA-1000 database [9] are shown in Figure 6 and we have two conclusions: 1) geodesic saliency significantly outperforms previous methods. Especially, it can highlight the entire object uniformly and achieve high precision in high recall areas. Figure 8 shows example images and results. 2) After combination with geodesic saliency, all the eight previous methods are improved, and four of them (FT,GB,HC,RC) are better than both that are combined. In Figure 6 (bottom) we show results before and after combination for the two best combined methods, GS+HC and GS+RC.

To further understand how geodesic saliency differs from other methods, Figure 7 shows the saliency value distributions of all foreground and background pixels (using ground truth masks) for GS_GD and RC methods. Clearly, the foreground and background are better separated in GS_GD than in RC.

Results on the Berkeley-300 database [30] are shown in Figure 9 and example results are shown in Figure 10. The accuracy of all methods are much lower, indicating that this database is much more difficult. We still have similar conclusions about geodesic saliency: 1) it is better than previous methods; 2) after combination, all the eight previous methods are improved, and four of them (CA,IT,GB,RC) are better than both that are combined. In Figure 9 (bottom) we show results before and after combination for the two best combined methods, GS+GB and GS+RC.

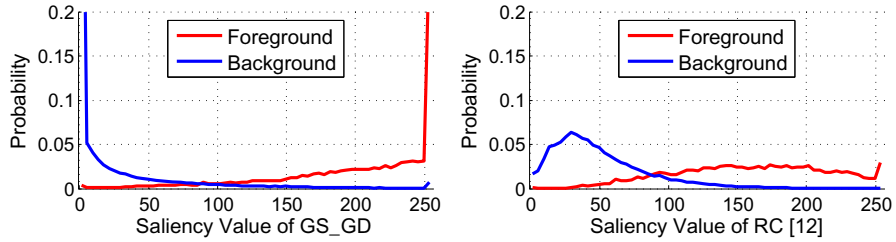

Fig. 7. Saliency value distributions of all foreground and background pixels in the MSRA-1000 database [9] for GS_GD and RC methods

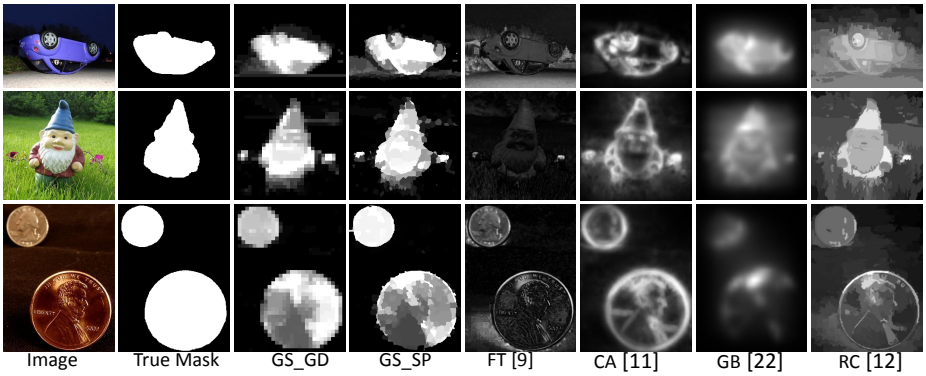

Fig. 8. Example results of different methods on the MSRA-1000 database [9]

Running times of all the methods are summarized in Table 1. GS_GD is significantly faster than all the methods and useful for real-time applications.

The only important free parameter is the patch size. In experiment, we found that best results are obtained when it is within $[\frac{1}{80}, \frac{1}{20}]$ of the image dimension. As most images are 400×400, we use 10×10 patches (roughly $\frac{1}{40}$).

We also evaluate the effect of the *weight clipping* and *boundary edge weight computation* by comparing the results with and without them. It turns out that geodesic saliency without these components is still better than previous methods, but by a very small margin. This shows that the background priors are indeed useful, and these algorithm components can further improve the performance.

5 Discussions

Given the superior results and the fact that geodesic saliency is so intuitive, this illustrates that appropriate prior exploitation is helpful for the ill-posed saliency detection problem. This is a similar conclusion that has been repeatedly observed in many other ill-posed vision problems.

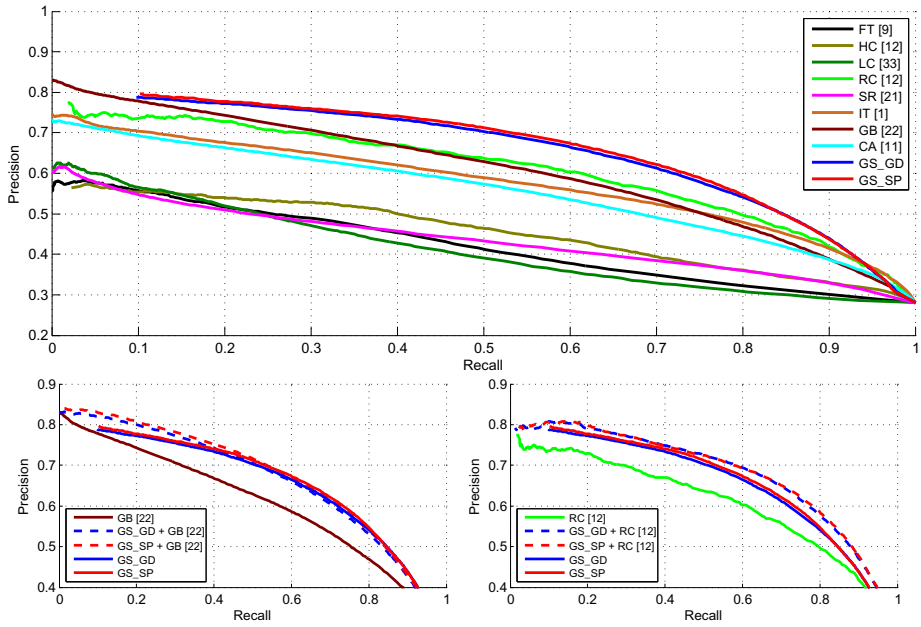

Fig. 9. (better viewed in color) Average precision-recall curves on the Berkeley-300 database [30]. Top: Geodesic saliency is better than all the other methods. Bottom: Combination of geodesic saliency with GB(RC) is better than both.

Table 1. Averaging running time (milliseconds per image) of different methods, measured on an Intel 2.33GHz CPU with 4GB RAM. IT, GB and CA use Matlab implementation, while other methods use C++.

GS_GD	GS_SP	IT [1]	SR [21]	GB [22]	FT [9]	CA [11]	LC [33]	HC [12]	RC [12]
2.0	7438	483	34	1557	8.5	59327	9.6	10.1	134.5

As our approach is built on statistical priors, inevitably it fails when such priors are invalid. Figures 11 shows typical failure cases: objects significantly touching the image boundary and complex backgrounds. Nevertheless, the priors hold for most images and geodesic saliency performs well in general.

There are several directions to improve geodesic saliency: 1) alleviate its dependency on the background priors; 2) compute saliency on the image boundary in a better way, instead of solving a one-dimensional saliency detection problem using only image boundary; 3) combine with other methods in a way better than our straightforward combination by averaging.

Finally, we would encourage future work to use more challenging databases (*e.g.*, Berkeley or even PASCAL) to advance the research in this field.

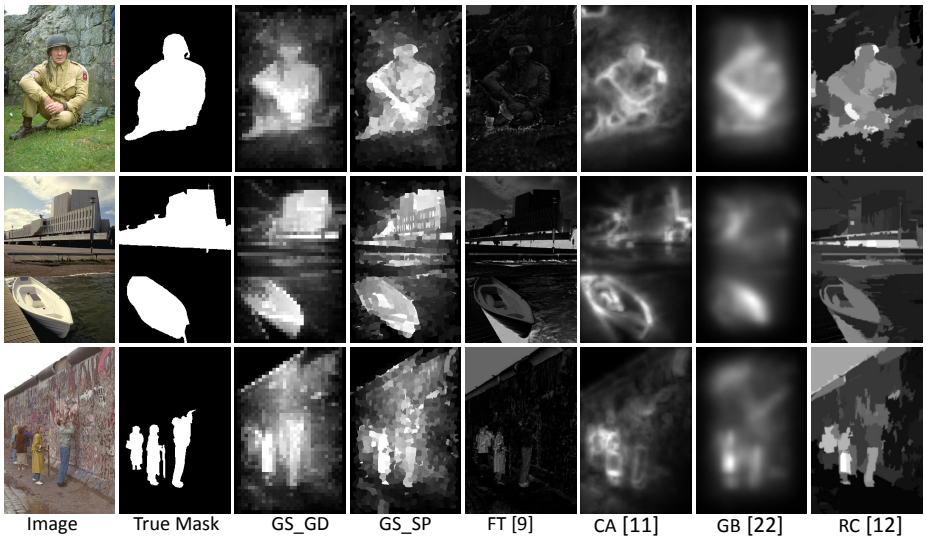

Fig. 10. Example results of different methods on the Berkeley-300 database [30]

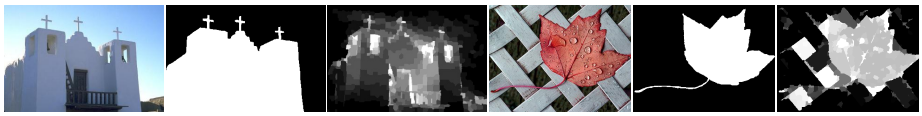

Fig. 11. Typical failure cases of geodesic saliency: (left) salient object is significantly cropped at the image boundary and identified as background; (right) small isolated background regions are identified as object

References

1. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 1254–1259 (1998)
2. Parkhurst, D., Law, K., Niebur, E.: Modeling the role of saliency in the allocation of overt visual attention. Vision Research (2002)
3. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: ICCV (2009)
4. Ramanathan, S., Katti, H., Sebe, N., Kankanhalli, M., Chua, T.-S.: An Eye Fixation Database for Saliency Detection in Images. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 30–43. Springer, Heidelberg (2010)
5. Itti, L., Baldi, P.: Bayesian surprise attracts human attentions. In: NIPS (2005)
6. Bruce, N.D.B., Tsotsos, K.: Saliency based on information maximization. In: NIPS (2005)
7. Gao, D., Mahadevan, V., Vasconcelos, N.: The discriminant center-surround hypothesis for bottom-up saliency. In: NIPS (2007)
8. Liu, T., Sun, J., Zheng, N., Tang, X., Shum, H.: Learning to detect a salient object. In: CVPR (2007)

9. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: CVPR (2009)
10. Valenti, R., Sebe, N., Gevers, T.: Image saliency by isocentric curvedness and color. In: ICCV (2009)
11. Goferman, S., Manor, L., Tal, A.: Context-aware saliency detection. In: CVPR (2010)
12. Cheng, M., Zhang, G., Mitra, N., Huang, X., Hu, S.: Global contrast based salient region detection. In: CVPR (2011)
13. Klein, D., Frintrop, S.: Center-surround divergence of feature statistics for salient object detection. In: ICCV (2011)
14. Rubinstein, M., Guterrez, D., Sorkine, O., Shamir, A.: A comparative study of image retargeting. In: SIGGRAPH Asia (2010)
15. Sun, J., Ling, H.: Scale and object aware image retargeting for thumbnail browsing. In: ICCV (2011)
16. Marchesotti, L., Cifarelli, C., Csurka, G.: A framework for visual saliency detection with applications to image thumbnailing. In: ICCV (2009)
17. Lempitsky, V., Kohli, P., Rother, C., Sharp, T.: Image segmentation with a bounding box prior. In: ICCV (2009)
18. Wang, L., Xue, J., Zheng, N., Hua, G.: Automatic salient object extraction with contextual cue. In: ICCV (2011)
19. Alexe, B., Deselaers, T., Ferrari, V.: What is an object. In: CVPR (2010)
20. Feng, J., Wei, Y., Tao, L., Zhang, C., Sun, J.: Salient object detection by composition. In: ICCV (2011)
21. Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. In: CVPR (2007)
22. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: NIPS (2006)
23. Grady, L., Jolly, M., Seitz, A.: Segmentation from a box. In: ICCV (2011)
24. Vicente, S., Kolmogorov, V., Rother, C.: Graph cut based image segmentation with connectivity priors. In: CVPR (2008)
25. Nowozin, S., Lampert, C.H.: Global connectivity potentials for random field models. In: CVPR (2009)
26. Butko, N.J., Movellan, J.R.: Optimal scanning for faster object detection. In: CVPR (2009)
27. Toivanen, P.J.: New geodesic distance transforms for gray-scale images. Pattern Recognition Letters 17, 437–450 (1996)
28. Veksler, O., Boykov, Y., Mehrani, P.: Superpixels and Supervoxels in an Energy Optimization Framework. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 211–224. Springer, Heidelberg (2010)
29. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV (2001), `http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/`
30. Movahedi, V., Elder, J.H.: Design and perceptual validation of performance measures for salient object segmentation. In: IEEE Computer Society Workshop on Perceptual Organization in Computer Vision (2010), `http://elderlab.yorku.ca/~vida/SOD/index.html`
31. Bai, X., Sapiro, G.: A geodesic framework for fast interactive image and video segmentation and matting. In: ICCV (2007)
32. Criminisi, A., Sharp, T., Blake, A.: GeoS: Geodesic Image Segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 99–112. Springer, Heidelberg (2008)
33. Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. ACM Multimedia (2006)