REGION ENHANCED SCALE-INVARIANT SALIENCY DETECTION

Feng Liu and Michael Gleicher

Department of Computer Sciences
University of Wisconsin-Madison

ABSTRACT

Saliency measures the low-level stimuli to human vision, and serves as an alternative to semantic image understanding. This paper presents a region enhanced scale-invariant saliency detection method. Our method constructs a scale-invariant saliency map from an image, segments the image into regions, and enhances the saliency map with the region information. Compared with previous methods, our method has advantages in providing robust scale-invariant saliency, giving meaningful region information for applications, and eliminating misleading high-contrast edges.

1. INTRODUCTION

Finding the objects that are important in an image requires semantic image understanding. Such understanding, however, is beyond the state of the art of research in vision and psychology. Therefore applications, such as image/video retrieval [1], video abstraction/ summarization [2], adaptive content delivery and image/ video retargeting [3, 4, 5, 6], rely on two heuristics to localize important objects. First, they use identifiable high level information, such as faces and texts, to determine the important areas. Second, they use image saliency, which simulates the low-level stimuli to human vision, as an indication of the importance.

The use of identifiable high level information is limited because it does not exist in some images and is hard to extract automatically when it does exist. Image saliency (typically forms of contrast) is always available. However, it fails to provide enough information for applications to localize the salient objects. The low-level spatial features do not necessarily map well to the salient objects. For instance, high-contrast edges between regions usually stand out, which will mislead applications into identifying the wrong salient object as illustrated in Fig. 2 (b). Moreover, existing image saliency detection methods fail to identify salient image features that may occur at various scales. Failing to address the scale-invariant saliency, salient features at some scales will be lost, which will mislead applications.

In this paper, we present a region enhanced scale-invariant saliency detection method, which combines both the scale-invariant saliency and region information. Our method obtains the scale-invariant saliency through a multi-resolution feature contrast calculation. The idea is to calculate the image feature contrast at an image scale matching the feature scale. That is, the contrast of the large-scale features is calculated at a coarse image scale and that of the small-scale features is calculated at a fine scale. To achieve salient region localization, our method enhances the saliency with region information from image segmentation by averaging the saliency value in each region. Our method provides the following advantages over previous methods: First, it provides robust scale-invariant saliency. Second, it provides salient regions, and is a close approximation to important object extraction. Third, it eliminates the misleading high-contrast edges.

After a brief overview of previous works, we will describe scale-invariant saliency detection and region enhanced saliency calculation in Section 3. We will discuss the results in Section 4.

2. RELATED WORK

Results from vision science, such as [7, 8], suggest that saliency can be measured by low-level feature contrast. This serves as the theoretical basis of all the existing methods.

Itti et al. [9, 10] detect local spatial feature discontinuities in a static image pyramid with a fixed number of scales, as feature contrast maps, and combine them into a single saliency map. Ma et al. [11] use a simple contrast based image saliency map, which has proved to be effective in practice. These saliency detection methods have two main limitations. First, they either empirically select a fixed image scale [11] or a fixed number of image scales for the feature contrast computation [9, 10]. This scale dependence causes parameter sensitivity. Different images have different optimal image scales to calculate the saliency, and it is difficult (sometimes impossible) to automatically determine a suitable scale for each image. Second, the above saliency detection methods are pixels/blocks based, and the resulting low-level spatial salient feature not only can not provide information for important region/object localization, but also will mislead applications by high contrast edges in the saliency map, as illustrated in Fig.1 and 2.

Recently, several researchers have addressed these issues

This work is supported in part by NSF IIS-0097456 and IIS-0416284.

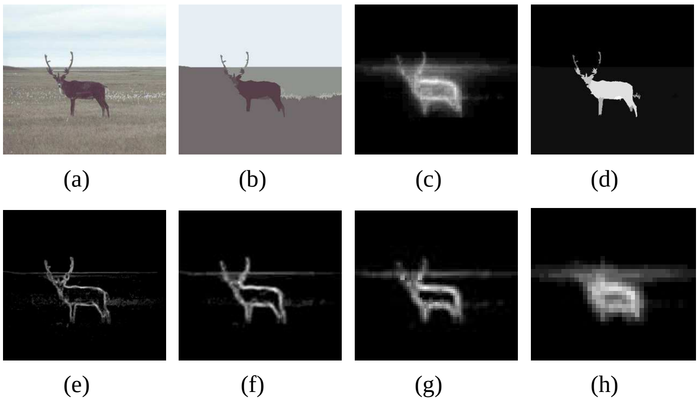


Fig. 1. Region enhanced saliency detection. (a): original image. (b): segmentation result. (c): scale-invariant saliency. (d): region enhanced saliency (e), (f), (g) and (h): saliency at image scale 0, 1, 2 and 3 respectively.

through region based saliency analysis. Li et al. [12] presented a salient region extraction method. They use k-means clustering to segment the image into homogenous regions, and use k-means clustering again to classify the regions into salient and non-salient groups based on the observation that salient regions usually are in the image center. Hu et al. [13] extract a salient region by constructing a convex hull from the salient points. In their recent work [14], they transform image features into a 2D space through a polar transformation, and identify regions by estimating the subspaces. They consider both the region feature contrast and its geometric properties to determine the saliency. These pure region based methods localize regions well, however, heavily depend on the region extraction in practice, requiring reasonable regions by tuning the parameters. Errors from region extraction can cause the saliency detection to fail catastrophically.

Other research combines low-level saliency with high-level information to achieve attentive region/object localization. Saliency is combined with high-level information such as faces and text to find the region of interest in [3, 5, 6]. Ma et al. [2] combines saliency with motion, camera, and faces to build a video attention model. Setlur et al. [15] combines saliency with region information in a similar way to ours, however, their method is scale-dependent.

3. REGION ENHANCED SALIENCY

We propose a hybrid approach, combining low-level saliency and region information. Our method first calculates a scale-invariant saliency map based on pixels/blocks, and then enhances it with region information from image segmentation by averaging the saliency value in each region. Region information provides two advantages. First, it is always available in images while other high information such as faces and texts are not. Second, regions provide direct information for applications to localize objects. However, region extraction is not robust. On the other hand, the pixels/blocks based scale-invariant saliency excels at robustness, and is weak at region localization. Our scheme combines the advantages of the above two information.

3.1. Scale-invariant saliency

Unlike previous scale-dependent methods, that obtain the saliency by calculating feature contrast at a fixed scale or a fixed number of scales, we construct a scale-invariant saliency map through a multi-scale analysis. We use the method of Ma et al. [11] to calculate the contrast at each image scale. The underlying idea of this multi-scale method is to calculate the image feature contrast at an image scale matching the feature scale. In another word, features will stand out at an image scale matching to their feature scales. For example, the large scale features will be highlighted at a coarse scale and the small-scale features will be highlighted at a fine scale. The algorithm is outlined as follows:

Step 1: Transform the image into a perceptually uniform color space (Lu*v*).

Step 2: Build Gaussian image pyramid from the image. The number of levels, n_l, is calculated from the original image size (w, h) as $\log_2 (\min (w, h)/10)$.

Step 3: Build the contrast pyramid by calculating the contrast map at each scale as illustrated in Fig. 1(e), (f), (g) and (h). The contrast value $C_{i,j,l}$ at image scale l is defined as the weighted sum of the differences between the pixel (i, j) at scale l and each other pixel in its neighborhood. That is,

$$C_{i,j,l} = \sum_{q \in \Theta} w_{i,j,l} d(p_{i,j,l}, p_q) \quad (1)$$
$$w_{i,j,l} = 1 - r_{i,j,l}/r_{l,max}$$

where Θ is the neighborhood of pixel (i, j) at scale l, $p_{i,j,l}$ is the color of the pixel at i, j at scale l, p_q is the color of the pixel in $p_{i,j,l}$'s neighborhood, and d is the magnitude of the color difference using the L^2 norm. The weighting factor $w_{i,j,l}$ is used to account for the heuristics that the center of an image is usually more visually salient. $r_{i,j,l}$ is the distance from (i, j) to the image center and $r_{l,max}$ is the maximal distance to the image center. This is similar to [11] with scale added. Our final result is not sensitive to the size of the neighborhood Θ due to the multi-scale scheme.

Step 4: Reconstruct the saliency map from the contrast pyramid by summing up the contrast map at all the scales.

From Fig. 1(e), (f), (g) and (h), we can see that the contrast map at different scale differs. For example, the interior region of the deer's body is not salient at scale 0 and scale 1, but it is salient at scale 3 as illustrated in Fig. 1(e), (f) and (h) respectively because the body is a large scale feature, it can stand out only at a coarse scale. On the other hand, the deer's antler stands out at scale 0 and scale 1, but not at scale 3, because the antler is of small scale, it can only stand out at a finer scale. The saliency map from the multi-scale analysis, illustrated in Fig. 1(c), presents saliency with different scales.

3.2. Region enhanced saliency

A problem of pixels/blocks based saliency detection methods is that they can not provide accurate information to localize salient objects. Moreover, misleading high-contrast edges often stands out instead of salient regions/objects, for example, the line separating the sky and the grassland poping out as shown in Fig. 1(e), (f), (g), (h) as well as (c).

We use region information extracted from the image to enhance the scale-invariant saliency map. We calculate the saliency value of each region as the average saliency value of pixels within it in a similar way to [15]. The final saliency map is illustrated in Fig. 1(d). From this result, we can see that the misleading high-contrast edge is eliminated and the salient regions with accurate boundaries stand out.

Extracting regions from an image is a well-studied field. We adopt the mean shift image segmentation algorithm [16] however other image segmentation algorithms can be used instead. Mean shift is a non-linear method based on non-parametric density estimation, and models image data as clusters in both spatial and range domain. It can adapt to the local image structure automatically. An example is shown in Fig. 1(b).

4. RESULTS

We compare our region enhanced saliency detection method with previous methods, namely a previous pixels/blocks based method and a pure region based method.

Fig. 2 shows some typical examples that compare region enhanced saliency to a pixels/blocks based saliency, which is implemented as a contrast map at a fixed scale [11]. From these results, we can see: First, our method eliminates misleading high contrast edges. For example, if the saliency map in Fig. 2(b) is used to find a region of interest using typical algorithms [5, 6], instead of the hill as indicated by a white rectangle in Fig. 2(c), the edge is selected as the interesting region. Second, region enhanced saliency map provides valuable information for applications. For example, saliency maps in Fig. 2(f), (i), (l) and (o) provide sharp and accurate potential object boundaries, and potential object components. In case of simple background, salient object can be obtained directly from the saliency map as illustrated in Fig. 2(o).

Fig. 3 shows some typical examples that compare the region enhanced saliency to a pure region based saliency. We use the same mean shift image segmentation algorithm in both our method and the pure region based method. We implemented the pure region based method as follows: segment an image into regions, and calculate the region color contrast among its neighborhood as the region saliency. Also the regions closer to the image center are given higher weight. We examine both algorithms on different image segmentation parameters, which will result in different regions from the same image. From Fig. 3, we can see that the pure region based method heavily depends on region scales via segmentation parameters while our algorithm gives almost consistent results.

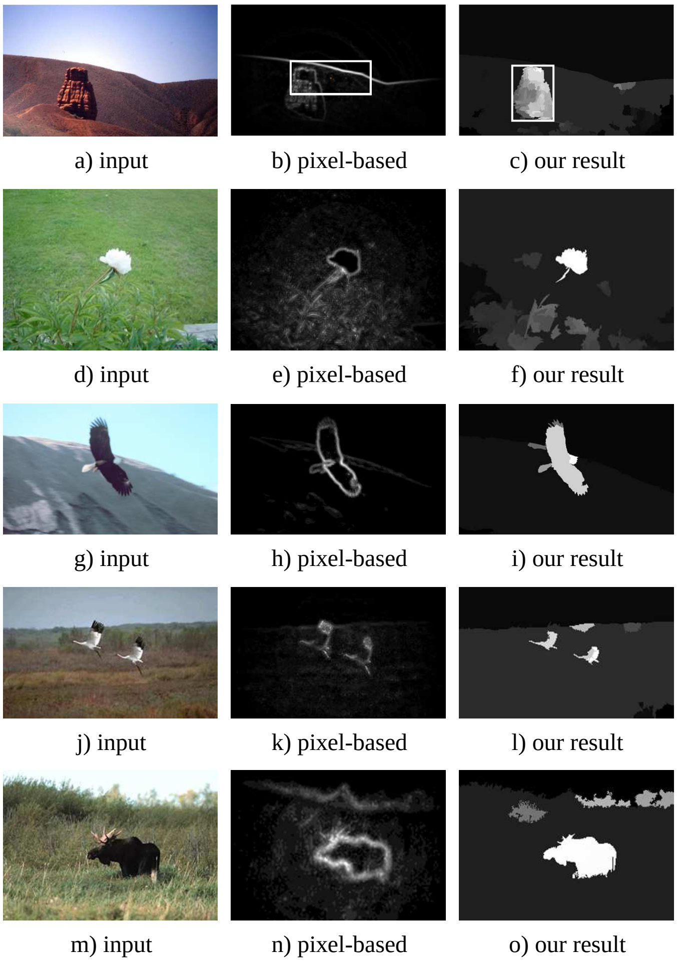
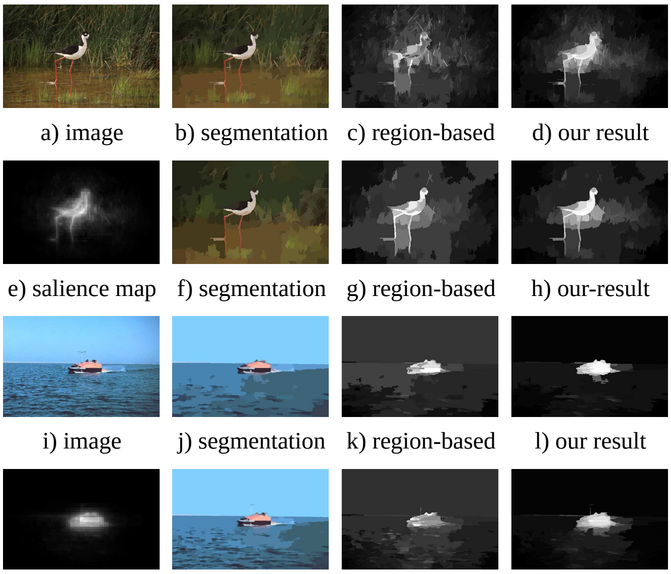

a) input
b) pixel-based
c) our result
d) input
e) pixel-based
f) our result
g) input
h) pixel-based
i) our result
j) input
k) pixel-based
l) our result
m) input
n) pixel-based
o) our result

Fig. 2. Region enhanced saliency vs. pixels/blocks based saliency. Original images are in the left column, the pixels/blocks based saliency maps are in the middle column, and the region enhanced saliency maps are in the right column.

5. CONCLUSION

In this paper, we present a region enhanced saliency detection method. We construct a scale-invariant saliency map through a multi-scale analysis and enhance it with region information



a) image b) segmentation c) region-based d) our result

e) salience map f) segmentation g) region-based h) our-result

i) image j) segmentation k) region-based l) our result

m)salience map n) segmentation o) region-based p) our result

Fig. 3. Region enhanced saliency vs. pure region based saliency. For the bird example, (a) is the original image, (e) is the scale-invariant saliency, (b) and (f) are segmentation result with different segmentation parameters, (c) and (g) are the corresponding pure region based saliency, and (d) and (h) are the corresponding region enhanced saliency. Similar captions apply to the ship example.

from image segmentation. Our method provides useful region information without suffering from the unreliability of image segmentation. Since our saliency is scale-invariant and free from misleading high contrast edges, it can be a reliable approximation or basis for important region/object localization.

6. REFERENCES

[1] Hao Liu, Xing Xie, Xiaoou Tang, Zhi-Wei Li, and Wei-Ying Ma, "Effective browsing of web image search results," in *MIR '04: Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*. 2004, pp. 84–90, ACM Press.

[2] Y.F. Ma, L. Lu, H.J. Zhang, and M.J. Li, "A user attention model for video summarization," in *Proceedings ACM Multimedia 2002*, 2002, pp. 533–542.

[3] Li-Qun Chen, Xing Xie, Xin Fan, Wei-Ying Ma, Hong-Jiang Zhang, and He-Qin Zhou, "A visual attention model for adapting images on small displays," *ACM Multimedia Systems Journal*, pp. 353–364, 2003.

[4] Jun Wang, Marcel Reinders, Reginald Lagendijk, Jasper Lindenberg, and Mohan Kankanhalli, "Video content presentation on tiny devices," in *ICME 2004: Proceedings of IEEE International Conference on Multimedia and Expo*, 2004.

[5] Feng Liu and Michael Gleicher, "Automatic image retargeting with fisheye-view warping," in *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, 2005, pp. 153–162.

[6] Bongwon Suh, Haibin Ling, Benjamin B. Bederson, and David W. Jacobs, "Automatic thumbnail cropping and its effectiveness," in *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*. 2003, pp. 95–104, ACM Press.

[7] Hans-Christoph Nothdurft, "Salience from feature contrast: additivity across dimensions," *Vision Research*, vol. 40, no. 11-12, pp. 1183–1201, 2000.

[8] Hans-Christoph Nothdurft, "Salience from feature contrast: variations with texture density," *Vision Research*, vol. 40, no. 23, pp. 3181–3200, Jan. 2000.

[9] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[10] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, Mar 2001.

[11] Yu-Fei Ma and Hong-Jiang Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proceedings ACM Multimedia 2003*, 2003, pp. 374–381.

[12] Ying Li, Yu-Fei Ma, and Hong-Jiang Zhang, "Salient region detection and tracking in video," in *ICME 2003: Proceedings of IEEE International Conference on Multimedia and Expo*, 2003.

[13] Yiqun Hu, Xing Xie, Wei-Ying Ma, Liang-Tien Chia, and Deepu Rajan, "Salient region detection using weighted feature maps based on the human visual attention model.," in *IEEE PCM*, 2004, pp. 993–1000.

[14] Yiqun Hu, Deepu Rajan, and Liang-Tien Chia, "Robust subspace analysis for detecting visual attention regions in images," in *Proceedings ACM Multimedia 2005*, 2005.

[15] Vidya Setlur, Saeko Takagi, Ramesh Raskar, Michael Gleicher, and Bruce Gooch, "Automatic image retargeting," in *International Conference on Mobile and Ubiquitous Multimedia*, Dec. 2005.

[16] Dorin Comaniciu and Peter Meer, "Mean shift analysis and applications," in *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, 1999, pp. 1197 – 1203.