SALIENCY DETECTION USING MAXIMUM SYMMETRIC SURROUND

Radhakrishna Achanta and Sabine Süsstrunk

School of Computer and Communication Sciences (IC)
Ecole Polytechnique Fédérale de Lausanne (EPFL)

ABSTRACT

Detection of visually salient image regions is useful for applications like object segmentation, adaptive compression, and object recognition. Recently, full-resolution salient maps that retain well-defined boundaries have attracted attention. In these maps, boundaries are preserved by retaining substantially more frequency content from the original image than older techniques. However, if the salient regions comprise more than half the pixels of the image, or if the background is complex, the background gets highlighted instead of the salient object. In this paper, we introduce a method for salient region detection that retains the advantages of such saliency maps while overcoming their shortcomings. Our method exploits features of color and luminance, is simple to implement and is computationally efficient. We compare our algorithm to six state-of-the-art salient region detection methods using publicly available ground truth. Our method outperforms the six algorithms by achieving both higher precision and better recall. We also show application of our saliency maps in an automatic salient object segmentation scheme using graph-cuts.

***Index Terms*—** Image saliency, segmentation, content-aware image re-targeting, seam carving.

1. INTRODUCTION

Visual saliency is the perceptual quality that makes an object, person, or pixel stand out relative to its neighbors and thus capture our attention. The focus of this paper is the automatic detection of visually salient regions in images. This has applications such as adaptive content delivery [1], adaptive region-of-interest based image compression, image segmentation [2], object recognition, and content aware image resizing [3]. Our algorithm finds low-level, pre-attentive, bottom-up saliency. It is inspired by the biological concept of center-surround contrast, but is not based on any biological model.

Current methods of saliency detection can be computationally expensive and often generate saliency maps that have low resolution or poorly defined borders. In addition, some methods produce higher saliency values in the vicinity of object edges instead of generating maps that uniformly cover the whole object. These drawbacks often arise from failing to exploit appropriate spatial frequency content of the original image, as analyzed by Achanta et al [4]. They introduce a frequency-tuned approach to estimate center-surround contrast using color and luminance features that offers three advantages over existing methods: uniformly highlighted salient regions with well-defined boundaries, full resolution, and computational efficiency. This leads to a global saliency estimation approach that relies on the premise that there is no information available about the scale of the object. While this method outperforms several existing methods in terms of precision, recall and speed, in the presence of large salient objects or complex backgrounds, it may fail to correctly highlight the salient regions.

This work is supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

Fig. 1. *Top row images are original images. Bottom row images are the corresponding saliency maps using our algorithm.*

In this paper we rely on the hypothesis that with respect to the image borders we can make assumptions about the scale of an object. We thus vary the bandwidth of the center surround-filtering near image borders using symmetric surrounds. Our algorithm retains the advantages of accuracy, speed, and simplicity, while at the same time overcoming the drawbacks of existing methods. We prove its effectiveness by performing a precision-recall comparison with six other methods on a publicly available ground truth database of 1000 images and in a graph-based segmentation scheme.

2. SALIENCY COMPUTATION METHODS

Saliency has been referred to as *visual attention* [5, 1], *unpredictability*, *rarity*, or *surprise* [6]. Saliency estimation methods can broadly be classified as biologically based, purely computational, or those that combine the two ideas. In general, most methods employ a low-level approach of determining contrast of image regions relative to their surroundings using one or more features of intensity, color, and orientation.

Itti et al. [7] base their method on the biologically plausible architecture proposed by Koch and Ullman [8]. They determine center-surround contrast using a Difference of Gaussians (DoG) approach. Frintrop et al. [9] present a method inspired by Itti's method, but they compute center-surround differences with square filters and use integral images to speed up the calculations.

Some methods are purely computational [1, 10, 11, 12] and are not explicitly based on biological vision principles. Ma and Zhang [1] and Achanta et al. [12, 4] estimate saliency using center-surround feature distances. Hu et al. [10] estimate saliency by applying heuristic measures on initial saliency measures obtained by histogram thresholding of feature maps. Gao and Vasconcelos [13] maximize the mutual information between the feature distributions

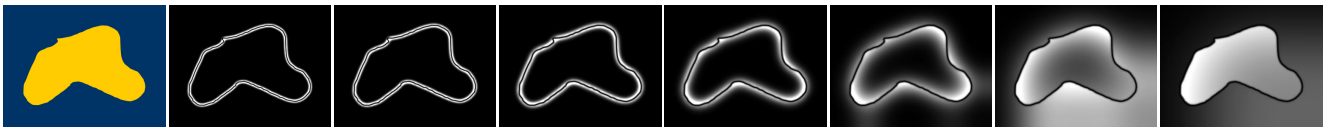

Fig. 2. *Band-pass filtering output with progressively increasing bandwidth from left to right (values in brackets show spatial frequency range). The high frequency cut-off is kept the same while the low-frequency cut-off is reduced. Pixels that are far removed from the salient object's boundaries need small cut-off frequencies to be successfully detected as salient.*

of center and surround regions in an image, while Hou and Zhang [11] rely on frequency domain processing.

The third category of methods are those that incorporate ideas that are partly based on biological models and partly on computational ones. For instance, Harel et al. [14] create feature maps with Itti's method but perform their normalization using a graph based approach. Other methods use a computational approach like maximization of information [15] that represents a biologically plausible model of saliency detection.

Some algorithms take a multi-scale approach [7, 12], while others operate on a single scale [1, 10]. Depending on the features used, either feature maps are created separately and combined to obtain the final saliency map [1, 10, 9, 16], or a combined saliency map is obtained directly [1, 12].

3. LIMITATIONS OF EXISTING METHODS

The saliency maps generated by several methods suffer from low resolution [7, 1, 14, 9, 11]. Itti's method produces saliency maps that are $1/256^{th}$ the original image size in pixels, while Hou and Zhang [11] output maps of size 64×64 pixels for any input image size. Some exceptions are the algorithms presented by Achanta et al. [12, 4] that output saliency maps of the same size as the input image.

Some methods may generate maps that have ill-defined object boundaries [7, 14, 9], limiting their usefulness in certain applications. Some others highlight the salient object boundaries but fail to highlight the entire salient region [1, 11], or, highlight smaller salient regions better than larger ones [12].

These limitations are explained from a frequency domain perspective by Achanta et al. [4] to be the consequence of limiting the range of spatial frequency content retained from the original image. The authors then propose a frequency-tuned algorithm for computing saliency maps that exploits almost all of the low frequency content and most of the high frequency content to obtain high quality saliency maps using color and intensity features. Their saliency map is obtained by computing the Euclidean distance of the average $CIELAB$ vector of all pixels of an input image with each pixel (also a $CIELAB$ vector) of a Gaussian blurred version (using a 3×3 or 5×5 binomial kernel) of the same input image:

$$S(x, y) = \|\mathbf{I}_\mu - \mathbf{I}_f(x, y)\| \quad (1)$$

where $S(x, y)$ is the pixel saliency value at position (x, y), $\mathbf{I}_\mu$ is the average of all $CIELAB$ pixel vectors of the image, $\mathbf{I}_f(x, y)$ is the corresponding $CIELAB$ image pixel vector in the Gaussian filtered version of the original image, and $\|\|$ is the L_2 norm (i.e. Euclidean distance in $CIELAB$ color space). The $CIELAB$ color space is used since Euclidean distances in this color space are approximately perceptually uniform.

The resulting saliency maps have uniformly highlighted salient regions with well-defined boundaries, which are proven to be an improvement over several state-of-the-art methods [4] for the given ground truth based database. However, in images where the salient region is very large, or the background is complex, the saliency maps highlight the background instead. This happens because in computing the average $CIELAB$ vector for the image in Eq. 1, the salient region contributes more to the image average than the rest of the image, thereby generating lower $S(x, y)$ values than the pixels of the background.

4. OUR SALIENCY DETECTION ALGORITHM

Achanta et al. [4] treat the entire image as the common surround (abstracted as the average image $CIELAB$ color vector) for any given pixel. The implicit premise is that in the absence of any knowledge of the scale of the salient object, it is best to pass all the low-frequency content. We base our new saliency detection algorithm on the premise that we can make assumptions about the scale of the object of detection based on its position in the image.

In Fig. 2 we note that the more central a pixel is within the salient object, the smaller has to be the low-frequency cut-off for detecting it. However, how central a pixel can be inside an object is limited by how far the pixel is from the boundary. That is, a pixel belonging to a salient object near the boundary will be less central inside the object. Therefore, assuming the salient object is fully within the image, and not cut-off by the image borders, we can afford to vary the bandwidth of the center-surround filter by increasing the low-frequency cut-off as we approach the image borders.

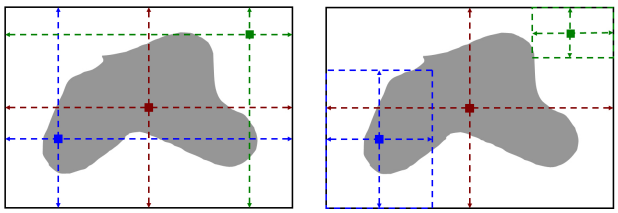

Fig. 3. *(A) In the method of [4], for a pixel at the center (red) or elsewhere (blue), the surround regions used for computing saliency remains the same, namely the whole image area. (B) Our new algorithm uses surround regions (sub-images) that are symmetric w.r.t the pixel whose saliency needs to be computed. This leads to varying center-surround bandwidth depending on the distance of the pixel from the image borders.*

In effect, as we approach the image borders we should use a more *local* surround region. We choose to do this by making the surround symmetric around the center with respect to the image borders as illustrated in Fig. 3 (B). This increases the low-frequency

cut-off of the center-surround filter. By choosing a symmetric surround for each pixel (as the center), we implicitly treat each pixel to be at the center of its own sub-image (see Fig. 3 (B)). This is different from the method of Achanta et al. [4], where the entire image is used as the common *global* surround (abstracted as the average image $CIELAB$ color vector) for any given pixel, resulting in an asymmetric surround for pixels that are not at the center of the image. This is explained graphically in Fig. 3 (A). Thus, for an input image of width w and height h, the symmetric surround saliency value at the given pixel $S_{ss}(x, y)$ is obtained as:

$$S_{ss}(x, y) = \|\mathbf{I}_{\mu}(x, y) - \mathbf{I}_f(x, y)\| \tag{2}$$

where $\mathbf{I}_{\mu}(x, y)$ is the average $CIELAB$ vector of the sub-image whose center pixel is at position (x, y) as given by:

$$\mathbf{I}_{\mu}(x, y) = \frac{1}{A} \sum_{i=x-x_o}^{x+x_o} \sum_{j=y-y_o}^{y+y_o} \mathbf{I}(i, j) \tag{3}$$

with offsets x_o, y_o, and area A of the sub-image computed as:

$$\begin{aligned} x_o &= min(x, w - x) \\ y_o &= min(y, h - y) \\ A &= (2x_o + 1)(2y_o + 1) \end{aligned} \tag{4}$$

The sub-images obtained in Eq. 3 using Eq. 4 are the maximum possible symmetric surround regions for a given pixel at the center. Consequently, the closer a pixel is to the edges, the narrower is its surround. To compute the $CIELAB$ averages of these sub-images, we take the computationally efficient approach of using integral images as done by [12, 9]. Examples of our saliency maps using our algorithm are shown in Figures 1 and 4. The advantage of narrowing the bandwidth near the borders is that the background is usually less highlighted. The disadvantage though is that if the salient object is cut by the image borders, i.e it is not completely inside the image, it is treated as background and is less likely to be detected.

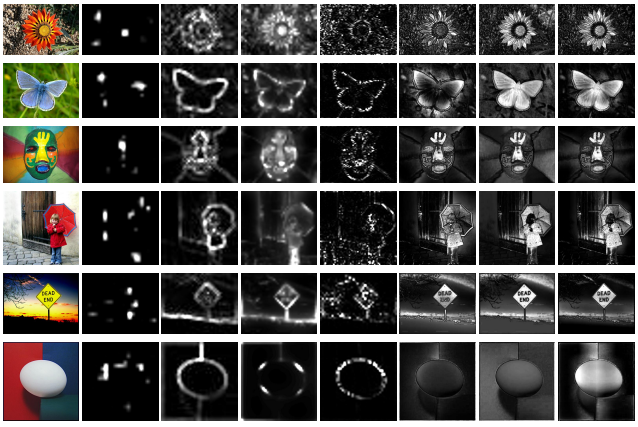

Fig. 4. *Visual comparison of saliency maps. Our method MSSS produces saliency maps that have well-defined borders, highlight whole object regions, and suppress the background better than most methods even in the presence of complex backgrounds or when the salient object is very large.*

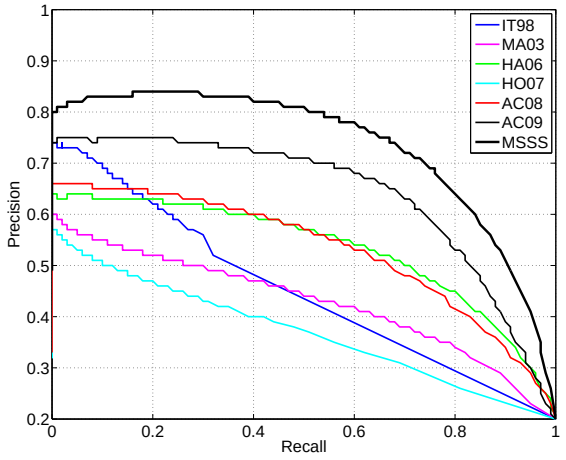

Fig. 5. *Precision-recall curve using groundtruth. Our new method MSSS shows the best precision-recall performance.*

5. COMPARISON WITH STATE-OF-THE ART

We compare the our saliency maps with six state-of-the-art methods. The six saliency detectors are Itti et al. [7], Ma and Zhang [1], Harel et al. [14], Hou and Zhang [11], Achanta et al. [12], and Achanta et al. [4], hereby referred to as IT98, MA03, HA06, HO07, AC08, and AC09, respectively. We refer to our proposed method as MSSS (for maximum symmetric surround saliency)[1].

In order to perform an objective comparison of the quality of the saliency maps with other methods, we use the precision-recall based method used by Achanta et al. [4]. For a given saliency map, with saliency values in the range $[0, 255]$, we perform simple binarization at each threshold value from 0 to 255, and compute the precision and recall values with respect to the ground truth data[2] from Achanta et al. [4]. The resulting precision versus recall curve is shown in Fig. 5. The algorithmic complexity of MSSS is linear in the number of pixels, i.e. $O(N)$. It is only marginally slower than AC09, which is the fastest full-resolution saliency detection algorithm to our knowledge.

6. GRAPH BASED SEGMENTATION

Graph cuts based methods are popular for image segmentation applications. Boykov and Jolly [17] perform interactive segmentation using graph cuts. They require a user to provide scribble based input to indicate foreground and background regions. A graph cuts based algorithm then segments foreground from background. We use a similar approach, however, instead of the user indicating the background and foreground pixels using scribbles, we use the saliency map to assign these pixels automatically.

As in the graph cuts formulation proposed by Boykov and Jolly [17], we assign binary values of salient or non-salient to a vector $V = [V_1, V_2....V_{|P|}]$ of size $|P|$, the number of pixels in an image. We seek an optimal cut between pixels belonging to salient and non-

[1]Source code for our method MSSS can be downloaded at `http://ivrg.epfl.ch/supplementary_material/RK_ICIP2010`

[2]`http://ivrg.epfl.ch/supplementary_material/RK_CVPR09`

salient regions. We use graph cuts to minimize the energy $E(V)$:

$$E(V) = \lambda E_1(V) + E_2(V) \tag{5}$$

where $E_1(V)$ accounts for the saliency value as obtained using Eq. 2, and $E_2(V)$ (Eq. 6) promotes coherence among similar pixel neighbors. $\lambda \geq 0$ specifies the relative importance of saliency value versus pixel similarity. $E_2(V)$ penalizes the assignment of different labels to neighboring pixels with similar $CIELAB$ vectors.

$$E_2(V) = \sum_{\{p,q\} \in N} exp^{\frac{-(\|\mathbf{I}(p)-\mathbf{I}(q)\|)}{2\sigma^2}} \times \frac{1}{dist(p,q)} \tag{6}$$

where N is the set of 8-connected neighboring pixels q around each pixel p of the image and $dist$ is the spatial distance between the pixels. We use $\lambda = 1.0$ and $\sigma = 10.0$ in our work. A few example results of segmentation are shown in Fig. 6. The segmentation scheme strongly depends on the quality of the saliency map. The output is better if the boundaries are well defined, the salient region is well highlighted, and the background is well suppressed. Thus, our method MSSS has an advantage over other saliency detection techniques for such an application.

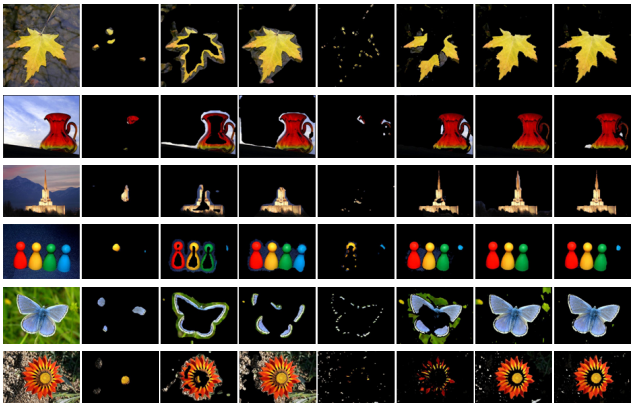

Fig. 6. *Salient object segmentation using graph cuts. From left to right, original image followed by output obtained using IT98, MA03, HA06, HO07, AC08, AC09, and MSSS. Our method MSSS is well suited for object segmentation using graph cuts.*

7. CONCLUSIONS

We present a novel saliency detection algorithm based on the idea of maximum symmetric surround. This method improves upon six existing state-of-the-art algorithms in precision and recall with respect to a ground truth database. Our algorithm uses low-level features of color and luminance. It is computationally efficient, easy to implement, and provides full resolution saliency maps that successfully suppress the background. We demonstrate the use of our saliency maps in salient object segmentation using graph-cuts.

8. REFERENCES

[1] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," *ACM International Conference on Multimedia*, pp. 374–381, November 2003.

[2] B. C. Ko and J.-Y. Nam, "Object-of-interest image segmentation based on human attention and semantic region clustering," *Journal of Optical Society of America A*, vol. 23, no. 10, pp. 2462–2470, October 2006.

[3] R. Achanta and S. Süsstrunk, "Saliency Detection for Content-aware Image Resizing," in *IEEE International Conference on Image Processing*, 2009.

[4] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, June 2009.

[5] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," *Artificial Intelligence*, vol. 78, no. 1-2, pp. 507–545, 1995.

[6] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," in *European Conference on Computer Vision*, 2004.

[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, November 1998.

[8] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.

[9] S. Frintrop, M. Klodt, and E. Rome, "A real-time visual attention system using integral images," in *International Conference on Computer Vision Systems*, March 2007.

[10] Y. Hu, X. Xie, W.-Y. Ma, L.-T. Chia, and D. Rajan, "Salient region detection using weighted feature maps based on the human visual attention model," *Springer Lecture Notes in Computer Science*, vol. 3332, no. 2, pp. 993–1000, October 2004.

[11] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 2007.

[12] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," *International Conference on Computer Vision Systems*, vol. 5008, pp. 66–75, 2008.

[13] D. Gao and N. Vasconcelos, "Bottom-up saliency is a discriminant process," *IEEE International Conference on Computer Vision*, pp. 1–6, Oct. 2007.

[14] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in Neural Information Processing Systems*, pp. 545–552, 2007.

[15] N. Bruce and J. Tsotsos, "Attention based on information maximization," *Journal of Vision*, vol. 7, no. 9, pp. 950–950, 6 2007.

[16] L. Itti and C. Koch, "Comparison of feature combination strategies for saliency-based visual attention systems," *SPIE Human Vision and Electronic Imaging IV*, pp. 473–482, May 1999.

[17] Y.Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," *IEEE International Conference on Computer Vision*, vol. 1, pp. 105–112, July 2001.