OBJECT SALIENCY USING A BACKGROUND PRIOR

Chintak Sheth and R. Venkatesh Babu

Video Analytics Lab,
Department of Computational and Data Sciences,
Indian Institute of Scince, Bangalore, India

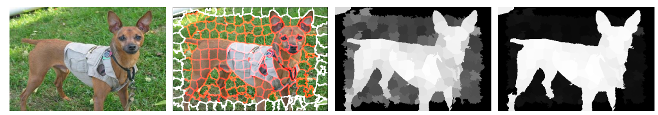
ABSTRACT

What makes an object salient? Almost all the works so far determine object saliency based on the amount of the contrast of a patch or super pixel with it's surrounding. Due to this approach, objects consisting of multiple colors, which is usually the case with a majority of natural objects, are allocated varying saliency values. Hence, post-processing for it's application is another problem. Taking note of this and keeping in mind the ease of extension to different applications, we provide a new perspective to this problem. We propose a simple yet powerful method for modelling the background for salient object detection. As a corollary of "Rule of Thirds" we model the background as the most occurring super pixels lying along the image border. Saliency is determined based on the distance of other super pixels from the background super pixels. Comparison of the proposed approach with the state of the art shows how our approach can provide more consistent saliency values throughout an object.

1. INTRODUCTION

Human vision has an innate ability to focus on objects of interest in an image, while disregarding unnecessary background information. Saliency detection, i.e. finding these regions of interest is currently an actively pursued problem in the computer vision community. Apart from helping understand the way in which humans perceive objects, finding salient regions and objects helps in speeding up various vision algorithms like object detection [1, 2], classification [3], retrieval, image editing [4], image and video compression [5] etc. Apart from these applications, a major motivation for salient region analysis is due to recent developments in egocentric acquisition led by Google glass and Go-Pro cameras. In analysing egocentric vision, to understand and predict human visual attention and behaviour, saliency maps are of vital importance [6].

Saliency map computation was initially formulated as a predictor of human fixations in images [7] and recent methods using deep convolutional neural networks [8] have shown significant progress on this task. With advances in saliency map computation, modelling the object saliency was actively used to propose bounding boxes in images as potential object locations [9, 10, 11, 12]. These boxes were used to speed up object detection [10] or to perform weakly supervised object annotation for training a detector [13]. Apart from fitting bounding boxes for object detection, increased pixel level accuracy of saliency maps has also enabled accurate segmentation of objects using saliency maps [14, 15].



(a) (b) (c) (d)

Fig. 1. (a) input image (b) SLIC super pixels - border super pixels in white. (c) Saliency computed as the Bhattacharyya distance from the background super pixels, Sec 3.2. (d) Final Saliency map after linear transformation, Sec 4.1.

This work focuses on combining super-pixel segmentations and background prior for saliency map generation, to solve the joint problem of finding the most salient region in an image as well as providing accurate segmentation. For developing a background prior, we average all the ground truths in the MSRA-1000 [16] and Complex Scene Saliency Detection [17] dataset (as shown in Fig 1(a) and 1(b)) we find that super pixels along the border are mostly considered as background. Using this as our background model we determine the saliency of a region based on it's dissimilarity to background regions. Furthermore, we show that by this modelling, the proposed approach is able to achieve state-of-the-art saliency maps, without the need for additional bottom-up saliency prior or additional hand-crafted features.

Before explaining the proposed approach in detail, a brief outline of the following paper is as follows. An overview of related work on visual saliency is presented in section 2. Section 3 provides an in-detail explanation of proposed approach. Section 4 shows experimentally the performance of proposed approach and we provide conclusions in section 5.

2. RELATED WORK

Over the years a lot of different frameworks have been proposed for object saliency detection. One major class of saliency computation methods focus on finding salient regions in an image to help solve the object detection problem. Jiang et al. [14] used object level shape prior to estimate the saliency, based on the observation that salient objects in an image have a well-defined closed boundary. Their proposed approach integrated multi-scale super pixels with object boundary information to formulate an iterative energy minimization framework such that salient object regions align with existing shape prior. In contrast to such shape modelling based saliency computation, Cheng et al. [11] used local contrast information within an image and spatial coherency to compute the saliency maps. Perazzi et al. [18] extended the use of contrast information in fusion with image abstraction to design Saliency filters. These Saliency filters are formulated as high dimensional Gaussian filters to estimate contrast based saliency. Further analysing such local region based prior, Goferman et al. [12] proposed a context-aware saliency algorithm based on four principles of human visual attention i.e., local low-level prior, global prior, visual organization rules and high-level factors.

Looking at improving the pixel level accuracy of saliency maps, especially to segment salient objects in an unsupervised framework, Achanta et al. [19] proposed low level prior to compute saliency maps. They determined saliency based on low-level features of luminance and color to generate high quality saliency maps which were then used to obtain semantically meaningful objects. Apart from such contrast based saliency maps Yan et al. [17] developed a scale based multi-layer approach to analyse saliency cues. This scale based region handling was performed by computing saliency values for regions in an efficient manner by using tree based modelling. Apart from computing the saliency of image regions recent work by Margolin et al. [20] analysed the characteristics of patches to understand "What makes a patch distinct?" This was done by analysing the statistics of patches in an image as well as a corpus of patches obtained from a training set. In addition by incorporation several high-level cues and prior to color based patch statistics Margolin et al. [20] were able to propose efficient object level saliency maps. Recently, Srivatsa et al. [21] has considered objectness measure computed using object proposals as an indicator of regional saliency in the image.

In contrast to existing methods which rely on analysing the contrast information in an image and rely on object shape prior, we propose a simple yet efficient algorithm by modelling the background prior in an image based on the boundary super pixel characteristics. By modelling saliency as region dissimilarity with background model and fitting piece wise linear transformation we are able to achieve semantically meaningful object level segmentation along with highly accurate saliency maps.

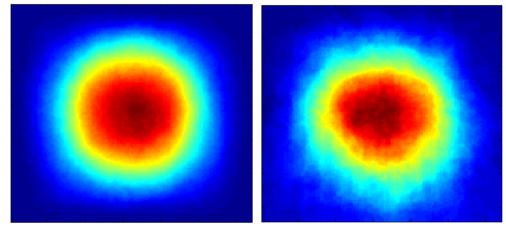


Fig. 2. Salient object prior using MSRA1000 [16] and CSSD [17] datasets.

3. SALIENCY VIA BACKGROUND MODELLING

In this section, we describe our method – Alg. 1, for object saliency detection. Contrary to most of the recent works [9, 12, 16, 22, 23] we formulate the problem as that of background modelling.

3.1. Region Representation

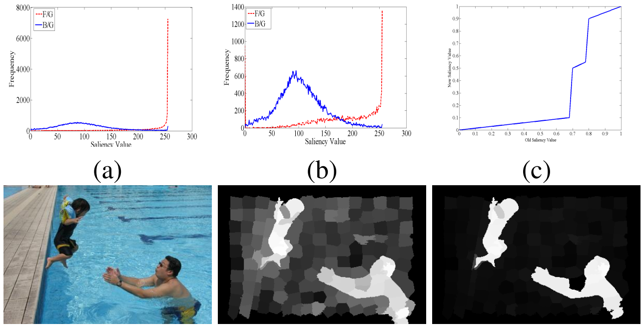
For being able to model image regions correctly, we must use regions having more or less similar color information, at the same time use the inherent variations within the region to compare with other regions. To obtain such image regions, we over-segment the image using SLIC super pixels [24] to obtain regions with similar color information. We represent each super pixel as a joint histogram in the CIE L*a*b* color space with 10 bins per channel.

3.2. Background Model

The *"Rule of Thirds"* or the "golden rule", a location prior which has been studied in photo quality assessment [25, 26, 27] states that to attract attention, the object of interest should lie at one of the four intersections in the center of the image to approximate the "golden ratio" (about 1.618). This rather simplistic corollary of the "golden rule" helps us compute consistent saliency values for an object very efficiently. This can also be observed in Fig. 2 which shows the probability of salient object in two publicly available datasets. Hence, the majority of super pixels along the border, (B_l) make up our *background* model ($BG_l \subset B_l$). In the proposed approach, Bhattacharya distance between the color histograms are used as the distance metric. The distance d_{ij} represent the Bhattacharya distance between the i-th and j-th superpixel.

$$d_{ij} = \sqrt{1 - \sum_{x=1}^{m} \sqrt{hist_i(x) \cdot hist_j(x)}} \tag{1}$$

Where $hist_i$ represents the joint CIE L*a*b* histogram of R_i – the super pixel with label i, d_{ij} is the Bhattacharyya



(a) (b) (c)

(d) (e) (f)

Fig. 3. (a), (b) Plot showing Saliency value distribution for Foreground and Background (MSRA1000 and CSSD resp.). (c) Linear transformation proposed for background suppression. (d)-(f) Example of this approach.

distance between $hist_i$, $hist_j$ and m denotes the number of bins in the histogram. The saliency of other/interior regions is determined as the minimum Bhattacharyya distance, eq (1) from any of the background region/super pixel, $R_i \in BG_l$.

In natural images the object of interest may very well extend across one half of the image and in the process may have some of the regions lying along the border. In order to prevent it from being wrongly classified as a background super pixel, we propose a single parameter, *number of similar border super pixels (N_b)*. We update our model as follows:

$$BG_l = \left\{ R_i \mid \frac{\sum_{j=1}^{N_b} sort_j(d_{ij})}{N_b} < \tau, \forall i, j \in B_l \right\} \qquad (2)$$

Where R_i, is a border super pixel with label i and R_j, is a border super pixel with label j. The intuition behind N_b is simple and straightforward, the distance will be larger for border super pixels which occur less frequently than N_b times along the border. In other words, N_b denotes the minimum number of border super pixel support for a super pixel to be assigned to the background model. Since salient object on the boundary will have less than N_b super pixel as support they will be excluded from the background model. Hence using eq (2) we are able to classify them as foreground super pixels. Therefore, by varying this parameter we can effectively handle cases when a part of the region lies along the border of the image. The qualitative effect of varying N_b has been illustrated in Fig. 4.

4. EXPERIMENTAL RESULTS

Our unoptimized MATLAB implementation takes about 0.60 seconds to process an image of resolution 400 × 300 on a PC with a 3.40 GHz CPU and 8GB memory. The time stated is inclusive of the time for SLIC super pixel generation. In our experiments we use $\tau = 0.75$, the threshold value in eq (2). For SLIC super pixel generation we use the executable provided by [24] with 200 super pixels.

Algorithm 1 ***Object Saliency via Background Modelling***

Require: Super pixel labels (**L**), Joint histogram $hist_i \ \forall\, i \ \in \mathbf{L}$, Parameter N_b
Ensure: Saliency map S

$\forall\, (i, j),\ dist(i, j) = \mathbf{BhattaDistance}(hist_i, hist_j)$;
B_l = Boundary super pixel labels from **L**
if $N_b == 0$ **then**
  $BG_l = B_l$;
else
  for all pairs of $i, j \in B_l$ **do**
    $spDists(i) = sort_{\forall j \neq i}(dist(i, j), ascending)$;
    $D_i = \frac{1}{N_b} \sum_{i=1}^{N_b}(spDists(i))$;
    if $D_i < \tau$ **then**
      $BG_l = BG_l \cup \{i\}$;
    end if
  end for
end if
for all $i \notin BG_l$ **do**
  $s = 1$;
  for all $j \in BG_l$ **do**
    $s = min(\mathbf{BhattaDistance}(hist_i, hist_j), s)$;
  end for
  $S(i) = s$ //$S(i)$ = Saliency of i-th superpixel
end for
return S

Function BhattaDistance($hist_x$, $hist_y$)
for $k = 1 : m$ **do**
  $bc = bc + \sqrt{(hist_x(k) \cdot hist_y(k))}$;
end for
return $\sqrt{(1 - bc)}$;
EndFunction

4.1. Piece-wise Linear Transform for Segmentation

Ideally, we would predict that all the regions of the background are given a near 0 saliency value. However, due to texture and intensity variations in natural images we note that the above approach provides a range of saliency values for background regions. For using saliency maps in applications we propose the use of a piece-wise linear transform as shown in Fig. 3(c). For estimating it's characteristics we compute two curves, Fig. 3(a,b). For each saliency map, we plot the histogram of pixel saliency values for region marked as foreground in ground truth (Red curves) and similarly for the background region (Blue curves). We compute this for both MSRA-1000 and CSSD dataset Fig. 3(a,b) respectively. This gives us insight into where the optimal threshold point lies. In Fig. 3(f) we show how the saliency map of Fig. 3(e) was transformed so as to give us uniform saliency values for the entire object.

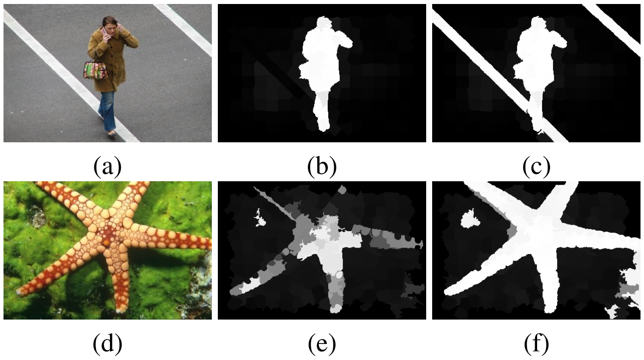


Fig. 4. (a), (d) Input Images. (b), (e) Saliency map with $N_b = 0$. (c),(f) Saliency map with $N_b = 5$

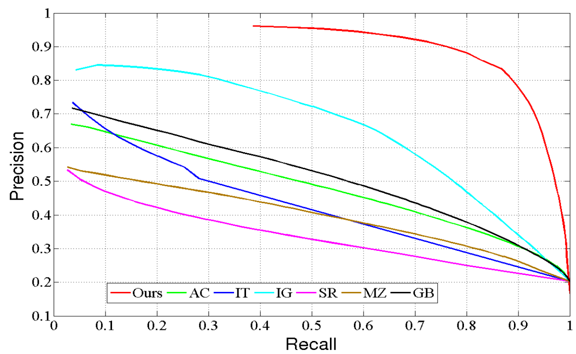


Fig. 5. PR-curve on MSRA-1000 dataset.

4.2. MSRA-1000 [16]

We tested our method on the saliency dataset MSRA-1000 [16] which contains 1000 natural images with their corresponding ground truths. In Fig 5, we compare our method with 6 state-of-the-art methods - IT [7], MZ [28], GB [22], SR [23], AC [19] and IG [16]. The saliency maps for all the above algorithms were obtained from [16]. In MSRA-1000 dataset most of the images are well restricted to the interior portions (as is further confirmed by Fig. 2(a)) and hence a value of $N_b = 1$ is able to provide us with very good results. As can be observed in the precision-recall curve, the proposed method performs extremely well in comparison to standard saliency computation algorithms. We observe a maximum precision of 0.96 in comparison to a precision of 0.85 by Achanta et al. (IG)[16] which is the next best, i.e. we observe more than 10% improvement with our proposed approach.

4.3. Qualitative Evaluation

In this experiment we provided a qualitative evaluation of proposed method on the CSSD saliency dataset provided by Yan et al. [17] consisting of much more complex image scenery and with regions of the object lying along the image border. In Fig. 6 we provide a qualitative comparison between our method and approaches by Yan et al. [17] and Yang et al. [29]. This example brilliantly shows how we are able to provide uniform saliency values to an entire object irrespective of the local contrast of different regions in the object. As can be observed, both these approaches provide higher saliency value to regions with a higher local contrast and fail to give a consistent saliency mapping for complete object regions. In comparison our method is able to provide semantically meaningful object saliency maps across various types of images. This can be observed prominently in column 3 where the left hand of the boy is missing from saliency maps of existing approaches while we are able to capture the complete foreground object accurately.

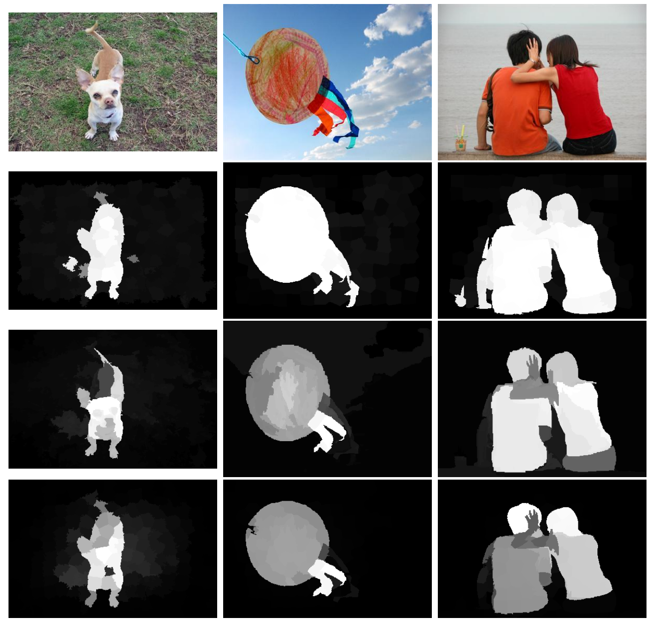


Fig. 6. Row 1: Input Images. Row 2: Our Saliency map. Row 3: Saliency map using [17]. Row 4: Saliency map using [29].

5. CONCLUSIONS

In this paper, we introduce a rather simple but yet powerful method for object saliency based on background modelling. We model each super pixel as a joint histogram in the CIE L*a*b* color space for perceptual coherency. We use the selected super pixels lying along the image border as a background model and compute the saliency of other interior super pixels as the minimum Bhattacharyya distance from the background super pixels. We tackle the fundamental problem of uniform salient value allocation to all parts of an object irrespective of it's local contrast via piece-wise linear transformation.

6. REFERENCES

[1] V. Navalpakkam and L. Itti, "An integrated model of top-down and bottom-up attention for optimizing detection speed," in *IEEE CVPR*, 2006.

[2] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?," in *IEEE CVPR*, 2004.

[3] Gaurav Sharma, "Discriminative spatial saliency for image classification," in *IEEE CVPR*, 2012.

[4] Jingdong Wang, Long Quan, Jian Sun, Xiaoou Tang, and Heung-Yeung Shum, "Picture collage," in *IEEE CVPR*, 2006.

[5] Jianru Xue, Ce Li, and Nanning Zheng, "Proto-object based rate control for jpeg2000: An approach to content-based scalability," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1177–1184, 2011.

[6] Kentaro Yamada, Yusuke Sugano, Takahiro Okabe, Yoichi Sato, Akihiro Sugimoto, and Kazuo Hiraki, "Can saliency map models predict human egocentric visual attention?," in *Computer Vision ACCV 2010 Workshops*. 2011.

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

[8] Srinivas SS Kruthiventi, Kumar Ayush, and R Venkatesh Babu, "DeepFix: A Fully Convolutional Neural Network for predicting Human Eye Fixations," *arXiv preprint arXiv:1510.02927*, 2015.

[9] P. Siva, C. Russell, Tao Xiang, and L. Agapito, "Looking beyond the image: Unsupervised learning for object saliency and detection," in *IEEE CVPR*, 2013.

[10] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari, "Measuring the objectness of image windows," *IEEE Transactions on PAMI*, vol. 34, no. 11, pp. 2189–2202, 2012.

[11] Ming-Ming Cheng, Guo-Xin Zhang, N. J. Mitra, Xiaolei Huang, and Shi-Min Hu, "Global contrast based salient region detection," in *IEEE CVPR*, 2011.

[12] Stas Goferman, Lihi Zelnik-Manor, and Ayellet Tal, "Context-aware saliency detection," *IEEE Transactions on PAMI*, vol. 34, no. 10, pp. 1915–1926, 2012.

[13] P. Siva and Tao Xiang, "Weakly supervised object detector learning with model drift detection," in *IEEE ICCV*, 2011.

[14] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Tie Liu, and Nanning Zheng, "Automatic salient object segmentation based on context and shape prior," in *Proceedings of the BMVC*, 2011.

[15] Yangqing Jia and Mei Han, "Category-independent object-level saliency detection," in *IEEE ICCV*, 2013.

[16] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Ssstrunk, "Frequency-tuned Salient Region Detection," in *IEEE CVPR*, 2009.

[17] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia, "Hierarchical saliency detection," in *IEEE CVPR*, 2013.

[18] Federico Perazzi, Philipp Krahenbuhl, Yael Pritch, and Alexander Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *IEEE CVPR*, 2012.

[19] Radhakrishna Achanta, Francisco Estrada, Patricia Wils, and Sabine Süsstrunk, "Salient region detection and segmentation," in *Computer Vision Systems*, pp. 66–75. 2008.

[20] Ran Margolin, Ayellet Tal, and Lihi Zelnik-Manor, "What makes a patch distinct?," *IEEE CVPR*, 2013.

[21] R Sai Srivatsa and R Venkatesh Babu, "Salient object detection via objectness measure," in *IEEE ICIP*, 2015.

[22] Jonathan Harel, Christof Koch, Pietro Perona, et al., "Graph-based visual saliency," *NIPS*, 2007.

[23] Xiaodi Hou and Liqing Zhang, "Saliency detection: A spectral residual approach," in *IEEE CVPR*, 2007.

[24] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and Sabine Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on PAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.

[25] Subhabrata Bhattacharya, Rahul Sukthankar, and Mubarak Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics," in *ACM MM*, 2010.

[26] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Ze Wang, "Studying aesthetics in photographic images using a computational approach," in *ECCV*, 2006.

[27] Yiwen Luo and Xiaoou Tang, "Photo and video quality evaluation: Focusing on the subject," in *ECCV*, 2008.

[28] Yu-Fei Ma and Hong-Jiang Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *ACM MM*, 2003.

[29] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, "Saliency detection via graph-based manifold ranking," in *IEEE CVPR*, 2013.