

# A Visual Saliency Map Based on Random Sub-Window Means

Tadmeri Narayan Vikram<sup>1,2</sup>, Marko Tscherepanow<sup>1</sup> and Britta Wrede<sup>1,2</sup>

<sup>1</sup>Applied Informatics Group

<sup>2</sup>Research Institute for Cognition and Robotics (CoR-Lab)

Bielefeld University, Bielefeld, Germany

{`nvikram,marko,bwrede`}@techfak.uni-bielefeld.de

**Abstract.** *In this article, we propose a simple and efficient method for computing an image saliency map, which performs well on both salient region detection and as well as eye gaze prediction tasks. A large number of distinct sub-windows with random co-ordinates and scales are generated over an image. The saliency descriptor of a pixel within a random sub-window is given by the absolute difference of its intensity value to the mean intensity of the sub-window. The final saliency value of a given pixel is obtained as the sum of all saliency descriptors corresponding to this pixel. Any given pixel can be included by one or more random sub-windows. The recall-precision performance of the proposed saliency map is comparable to other existing saliency maps for the task of salient region detection. It also achieves state-of-the-art performance for the task of eye gaze prediction in terms of receiver operating characteristics.*

**Keywords:** Bottom-up Visual Attention, Saliency Map, Salient Region Detection, Eye Fixation

## 1 Introduction

Visual saliency maps are utilized for determining salient regions in images or predicting human eye gaze patterns. Thus they have been exploited extensively for various intelligent interactive systems. There are a wide range of applications from image compression [2], object recognition [3–5], image segmentation [6] and various other computer vision tasks where saliency maps are employed. The intensity of a given pixel in the saliency map corresponds to the attention value attributed to the pixel in the original image.

The first computational model of visual attention was proposed by Koch and Ullmann [21]. They also introduced the concept of a saliency map. Subsequently, a great variety of different bottom-up visual attention models have been proposed in the literature. Methods which are employed to detect salient regions do not emphasize the semantic relevance and the opposite is true in the case of methods which are utilized to predict eye gaze patterns. Despite vast research there has not been a method which could be successfully utilized for both salient region detection as well as eye gaze pattern prediction. Therefore, in this paper we

propose a novel saliency map which performs well in salient region detection and eye gaze prediction tasks.

## 2 Literature Review

Existing saliency maps can be categorized into two fundamental groups: those which rely on local image statistics and those relying on global properties. The popular bottom-up visual attentional model proposed by Itti et al. [1] is based on local gradients, color and orientation features at different scales. It further inspired the application of contrast functions for the realization of bottom-up visual attention. The work of Gao et al. [22] was the first to employ contrast sensitivity kernels to measure center-surround saliencies. It was further improved by local-steering kernels [8] and self information [9]. The method of Bruce and Tsotsos [18] achieved the same level of performance by employing local entropy and mutual information-based features. Local methods are found to be computationally more expensive, and several global and quasi-global methods have been devised to address the issue of computational efficiency. The idea of utilizing the residual Fourier spectrum for saliency maps was proposed in [10, 11]. The authors employ the Fourier phase spectrum and select the high frequency components as saliency descriptors. These methods are shown to have high correlation with human eye-gaze pattern on an image. Frequency domain analysis for image saliency computation warrants the tuning of several experimental parameters. In order to alleviate this issue, several methods which rely on spatial statistics and features [6, 12–15] have been proposed.

Salient region detection and eye gaze prediction are the two significant applications of saliency maps. Salient region detection is relevant in the context of computer vision tasks like object detection, object localization and object tracking in videos [14]. Automatic prediction of eye gaze is important in the context of image aesthetics, image quality assessment, human-robot interaction and other tasks which involve detecting image regions which are semantically interesting [17]. The contemporary saliency maps are either employed to detect salient regions as in the case of [6, 12–14], or are used to predict gaze pattern which can be seen in the works of [1, 8–10, 15]. Though these two tasks appear similar, there are subtle differences between them. Salient regions of an image are those which are visually interesting. Human eye gaze which focuses mainly on salient regions is also distracted by semantically relevant regions [3].

Contrast has been the single most important feature for the computation of saliency maps and modelling bottom-up visual attention as it can be inferred from [6, 8, 9, 13, 14]. The method based on global contrast [6] employs absolute differences of pixels to the image mean as saliency representatives. The methods which model the distribution of contrast based on local image kernels [8, 9] need training priors and tuning of a large set of experimental parameters. The local weighting models proposed in [14, 15] are effective, but are computationally expensive. The local symmetric contrast-based method [13] overcomes the many aforementioned shortcomings. Recent research has suggested that contrast de-

tection and normalization in the V1 cortex is carried out in non-linear random local grids, rather than in linear fashion with regular grids [19, 20]. This property has been exploited in [14, 15] to compute saliency at pixel level and in [6] at global level.

We hereby propose a quasi-global method which operates by computing local saliencies over random regions of an image. This helps in obtaining better computational run-time and also captures local contrast unlike the global methods for computing saliency maps. Furthermore, it does not require any training priors and has only a single experimental parameter which needs tuning. Unlike the existing methods, the proposed saliency map is found to have consistent performance in both salient region detection and eye gaze prediction tasks. The proposed saliency map is determined as follows.

### 3 Our method

We consider a scenario where the input  $I$  is a color image of dimension  $r \times c \times 3$ , where  $r$  and  $c$  are the number of rows and columns respectively. The input image is subjected to a *Gaussian* filter in order to remove noise and abrupt onsets. This is further converted to CIE Lab space and decomposed into the three ( $L$ ,  $a$ ,  $b$ ) component images of dimension  $r \times c$ . CIE Lab space is preferred because of its similarity to the human psycho-visual space [13, 14].

Let  $n$  be the number of random sub-windows over the individual  $L$ ,  $a$  and  $b$  component images given

$$R_i = \{(x_{1i}, y_{1i}), (x_{2i}, y_{2i})\} \text{ such that } \begin{cases} 1 \leq i \leq n \\ 1 \leq x_{1i} < x_{2i} \leq r \\ 1 \leq y_{1i} < y_{2i} \leq c \end{cases} \quad (1)$$

where  $R_i$  is the  $i^{\text{th}}$  random sub-window with  $(x_{1i}, y_{1i})$  and  $(x_{2i}, y_{2i})$  being the upper left and the lower right co-ordinates respectively.

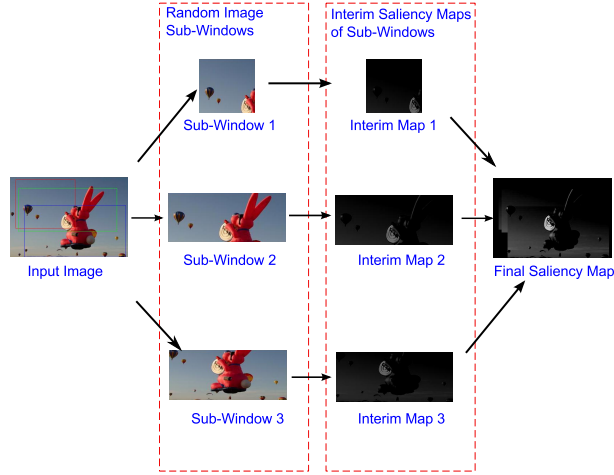
The final saliency map  $S$  of dimension  $r \times c$  is thus defined as

$$S = \sum_{i=1}^n \left( \|R_i^L - \mu(R_i^L)\| + \|R_i^a - \mu(R_i^a)\| + \|R_i^b - \mu(R_i^b)\| \right) \quad (2)$$

$\|\cdot\|$  denotes the Euclidean norm and  $\mu(\cdot)$  the mean of a given input vector, which is a two dimensional matrix in our case. To further enhance the quality of the saliency map  $S$ , we subject it to median filtering and histogram equalization. An illustration of the above paradigm is given in Fig. 1. For the sake of illustration we have considered only three random sub-windows and the resulting interim saliency map.

### 4 Results

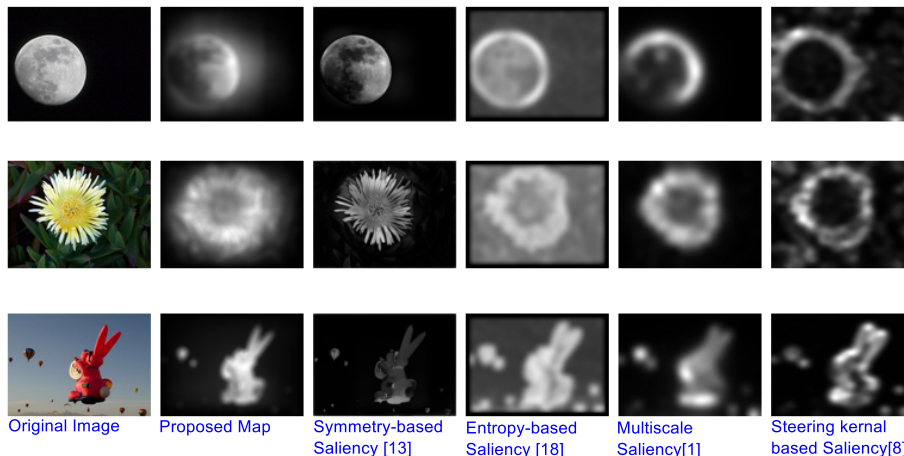
First, we illustrate the efficacy of our method on three selected images from the MSR [16] dataset in Fig. 2. The example image in Fig. 1 is also presented in



**Fig. 1.** An illustration of the proposed method where three random windows are generated to produce an interim saliency map for the given input image.

Fig. 2. It should be noted that the final saliency map shown in Fig. 2 is obtained by considering a large number of random sub-windows. It can be observed from Fig. 2 that multiscale saliency [1] and local steering kernel-based saliency [8] lay more emphasis on edges and other statistically significant points, rather than salient regions. The local steering kernel-based saliency [8], which is tuned to detect semantically relevant regions like corners, edges and local maxima ends up projecting mild image noise as salient. This can be observed on the upper right image of the moon in Fig. 2, where irrelevant regions are shown as salient. The results due to symmetry-based saliency [13], shows that the images have a sharp contrast. Images which do not consist of smooth signals have found to be bad representatives of eye gaze fixation, as eye gaze fixation function in reality is found to be smooth. The saliency provided by entropy-based methods [18] exhibit low contrast and are found to be inefficient for the task salient region detection in [14]. It can be observed that the proposed saliency does not output spurious regions as salient, has no edge bias, works well on both natural images and images with man made objects, and most importantly is also able to grasp the subjective semantics and context of a given region. The final property makes our method suitable for the task of eye gaze fixation. The experimentation carried out during the course of this research is presented in the section to follow.

In addition to the analysis of exemplar images, more comprehensive experiments were carried out on the MSR dataset [16] to validate the performance of the proposed saliency map for the task of salient region detection. In order to evaluate the performance on eye gaze prediction, experiments were conducted on the York University [18] and MIT [17] eye fixation datasets. We compared our method with reference to eight of the existing state-of-the-art methods. The selection of these methods was influenced by the impact factor of the conference



**Fig. 2.** Illustration of the proposed saliency map on three sample images of the MSR dataset [16]. From left to right: original Image from the MSR dataset [16], followed by the resultant saliency maps of the proposed method, symmetry-based saliency [13], entropy-based saliency [18], multiscale saliency [1] and local steering kernel-based saliency [8].

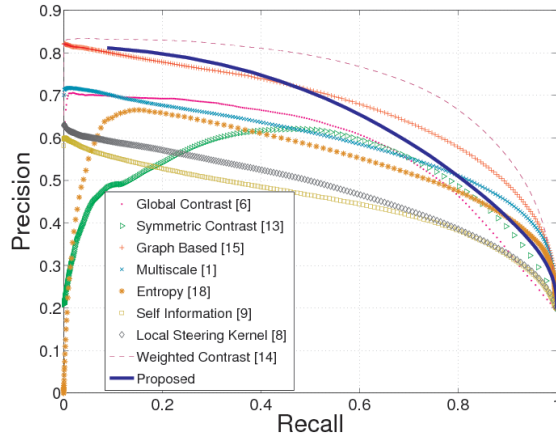
and journals in which they were published, popularity of the method in terms of citation, and the differences in their approaches. The eight methods are the global saliency-based method [6], symmetric saliency [13], entropy and mutual information-based saliency [18], graph-based saliency [15], multiscale saliency [1], local weighted saliency [14], self information-based saliency [9] and local steering kernel-based saliency [8]. The source codes for the methods were obtained from the homepages of the respective authors, whose links have been mentioned in their articles.

The following settings were used for all the experiments carried out. A Gaussian filter of size  $3 \times 3$  was used as a pre-processor on the images for noise removal. The number of distinct random sub-windows ( $n$ ) was set to  $0.02 \times r \times c$ . We arrived at this value, by varying ( $n$ ) from  $(0.005 \text{ to } 0.03) \times r \times c$  and found that the receiver operating characteristics (ROC) area under the curve (AUC) attained a level of saturation at  $0.02 \times r \times c$  on York University [18] and MIT [17] datasets. And finally a median filter of size  $11 \times 11$  was employed to smooth the resultant saliency map before being enhanced by histogram equalization. All experiments were conducted using Matlab v7.10.0 (R2010a), on an Intel Core 2 Duo processor with Ubuntu 10.04.1 LTS (Lucid Lynx) as operating system.

#### 4.1 Experiments with the MSR dataset

The original MSR dataset [16] consists of 5000 images with ground truths for salient regions as rectangular regions of interest (ROI). The problems and issues

due to such ROIs are explained in [6] and hence the same authors select a subset of 1000 images from the original set images and create exact segmentation masks. We followed the same experimental settings as described in [6]. In Fig. 3, we show the Recall-Precision performance of the models.



**Fig. 3.** The Recall-Precision performance of the methods under consideration on MSR dataset [16], with the experimental settings of [6]. Note that our method clearly outperforms the methods based on global contrast [6], symmetric contrast [13], multiscale [1], entropy [18], self information [9] and local steering kernel [8]

It can be observed that the proposed method clearly has a higher performance than the methods of [1, 6, 8, 9, 13, 18] and comparable performance with that of [14, 15], without having any of their drawbacks. The entropy-based saliency map [18] though promising does not have a high performance because the MSR dataset has a mix of natural images where the entropy is uniformly distributed. Local kernel-based methods [8, 9] also perform moderately because they are biased towards corners and edges than regions. Only the graph based method [15] and weighted distance method [14] perform well, because they have no bias towards edges.

## 4.2 Experiments using eye fixation datasets

We benchmarked the performance of the proposed method on York University [18] and the MIT [17] eye fixation dataset. The dataset of York University [18] consists of 120 images and the MIT dataset [17] consists of 1003 images. We followed the experimental method as suggested in [18] and obtained the ROC-AUC on the datasets. It can be observed from Table. 1, that our method has state-of-the-art performance. We omitted the methods of [9, 14] on the MIT

**Table 1.** The performance of the methods under consideration in terms of ROC-AUC on the York University [18] and MIT[17] datasets. It can be observed that the proposed method has state-of-the-art performance on both of the eye fixation datasets.

Saliency Map	York University [18]	MIT [17]
Global Contrast[6]	0.54	0.53
Symmetric Contrast[13]	0.64	0.63
Graph Based[15]	0.84	<b>0.81</b>
Multiscale[1]	0.81	0.76
Entropy[18]	0.83	0.77
Self Information[9]	0.67	-NA-
Local Steering Kernel[8]	0.75	0.72
Weighted Contrast[14]	0.75	-NA-
<b>Proposed</b>	<b>0.85</b>	<b>0.81</b>

dataset [17], as the corresponding Matlab codes required images to be down-sampled to a smaller size.

## 5 Discussion and Conclusion

We propose a method which has good performance on both salient region detection and eye gaze prediction tasks. The proposed method does not require training priors, has a minimal set of tunable parameters and relies only on contrast features to compute saliency maps. Our method requires minimal programming effort and achieves state-of-the-art performance despite its simplicity. Like the remaining contrast-based saliency maps, our method also fails to perform well when the color contrast is extremely low. Furthermore, the proposed saliency map fails when the task is to detect regions based on corners, orientation differences, minute differences in shapes etc. The proposed method is well suited in the scenario of human-robot interaction where eye gaze prediction and salient region detection need to be performed concurrently. Generating image specific random sub-windows to boost the proposed saliency map is another topic we wish to address in our future works.

**Acknowledgments.** Tadmeri Narayan Vikram gratefully acknowledges the financial support from the EU FP7 Marie Curie ITN RobotDoc. Contract No. 235065.

## References

1. Itti, L., Koch, C., Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Trans. Pattern Analysis Machine Intelligence*, 20(11), 1254–1259 (1998).

2. Guo, C., Zhang, L. : A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Trans. Image Processing*, 19(1), 185–198 (2010).
3. Rothenstein, A. L., Tsotsos J.K. : Attention links sensing to recognition. *Image and Vision Computing*, 26(1), 114–126 (2008).
4. Elazary, L., Itti, L.: A Bayesian model for efficient visual search and recognition. *Vision Research*, 50(14), 1338–1352 (2010).
5. Moosmann, F., Larlus, D., Jurie, F.: Learning Saliency Maps for Object Categorization. In *ECCV International Workshop on The Representation and Use of Prior Knowledge in Vision*, (2006).
6. Achanta, R., Estrada, F., Wils, P., Süsstrunk, S.: Frequency tuned Salient Region Detection. In *IEEE International Conference on Computer Vision and Pattern Recognition* (2009).
7. Buschman, T. J., Miller, E. K.: Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. *Science* 315(5820), 1860–1862 (2007).
8. Seo, H.J., Milanfar, P.: Static and Space-time Visual Saliency Detection by Self-Resemblance. *Journal of Vision*, 9(12), 1–27 (2009).
9. Zhang, L., Tong, M.H., Marks, T.K., Shan, H., Cottrell, G.W.: SUN: A Bayesian Framework for Saliency Using Natural Statistics. *Journal of Vision*, 8(7), 1–20 (2008).
10. Guo, C., Zhang, L.: A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Trans. Image Processing*, 19(1), 185–198, (2010).
11. Cui, X., Liu, Q., Metaxas, D.: Temporal spectral residual: fast motion saliency detection. In *ACM International Conference on Multimedia*, 617–620 (2009).
12. Rosin, P. L.: A simple method for detecting salient regions. *Pattern Recognition*, 42(11), 2363–2371 (2009).
13. Achanta, R., Süsstrunk, S.: Saliency Detection using Maximum Symmetric Surround. In *IEEE International Conference on Image Processing* (2010).
14. Vikram, T. N., Tscherepanow, M., Wrede, B.: A Random Center Surround Bottom up Visual Attention Model useful for Salient Region Detection. In *IEEE Workshop on Applications of Computer Vision*, 166–173 (2011).
15. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In *Neural Information Processing Systems*, 545–552 (2007).
16. Liu, T., Sun, J., Zheng, N., Tang, X., Shum, H.: Learning to Detect A Salient Object. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 1–8 (2007).
17. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In *IEEE International Conference on Computer Vision*, (2009).
18. Bruce, N.D., Tsotsos, J.K.: Attention based on Information Maximization. In the *International Conference on Computer Vision Systems*, (2007).
19. Mante, V., Frazor, R. A., Bonin, V., Geisler, W. S., Carandini, M.: Independence of luminance and contrast in natural scenes and in the early visual system. *Nature Neuroscience*, 8(12), 1690–1697, (2005).
20. Soltani, A., Koch, C.: Visual Saliency Computations: Mechanisms, Constraints, and the Effect of Feedback. *Neuroscience* 30(38) 12831–12843, (2010).
21. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology* 4(4), 219–227, (1985).
22. Gao, D., Mahadevan, V., Vasconcelos, N.: The discriminant center-surround hypothesis for bottom up saliency. In *Neural Information Processing Systems* (2007).