

# A Random Center Surround Bottom up Visual Attention Model useful for Salient Region Detection

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 bottom-up saliency model which works on capturing the contrast between random pixels in an image. The model is explained on the basis of the stimulus bias between two given stimuli (pixel intensity values) in an image and has a minimal set of tunable parameters. The methodology does not require any training bases or priors. We followed an established experimental setting and obtained state-of-the-art-results for salient region detection on the MSR dataset. Further experiments demonstrate that our method is robust to noise and has, in comparison to six other state-of-the-art models, a consistent performance in terms of recall, precision and F-measure.*

## 1. Introduction

Visual attention is now being extensively investigated by the computer vision community, as it is found to be useful for various vision automation tasks such as object detection [4], image segmentation [11], image compression [10], object recognition [7], image quality assessment [10] etc. In view of this, several computational models of visual attention have been proposed in the literature.

The primary implication of a computational model of visual attention is that it outputs a saliency map for a given input image, highlighting the saliency value for each pixel in the image. They not only enable a computer vision system to generate output inspired by human cognition, but also reduce the computational overload as they eliminate regions which are not interesting [1].

From a psychological perspective, visual attention is categorized into two types. Goal-driven visual attention, where attention is directed by a prior expectation and stimulus driven attention caused by unique features, abrupt onsets and appearance of new perceptual onset. The former is called top-down attention, while the latter is referred to as

bottom-up attention. The basic paradigm of bottom-up visual attention, is that it directs attention towards salient regions and objects [2]. Various studies carried out by Itti [2], Koch [20], Tsotsos [1] and other leading groups, have shown a strong correlation between bottom-up saliency maps and their impact on improving the performance of various computer vision tasks mentioned above.

An ideal bottom-up saliency model should be designed in such a way that the operations in it can be explained in terms of visual psychology and not just image statistics. Hence a bottom-up saliency model by definition should not have any prior bases, inferences or any other training weights/tunable parameters. Finally, such a model has to be simple enough to be programmed, and should have minimal computational complexity such that it performs in real-time. In view of this, we propose a model which works on capturing the contrasts between random pixels in image. The model is explained on the basis of the stimulus bias between two given stimuli in an image as proposed by Nosofosky [29] and Ahumada [28]. Furthermore, it has a minimal set of tunable parameters. We followed the experimental settings of Achanta et al. [12] and obtained state-of-the-art results for salient region detection on the MSR [24] dataset. Automatic detection of salient regions of interest, plays an important role in most of the computer vision tasks given that there is an explosion of multimedia data.

The paper is organized as follows. In Section 2, a brief literature survey is presented. We further describe the motivations leading to the present work in Section 3. The proposed methods and the corresponding algorithms are explained in Section 4. We present the experimental results carried out during the course of this research in Section 5, and finally we conclude this article in Section 6.

## 2. Literature Survey

The classic example for a bottom-up saliency model is the hierarchical model proposed by Itti et al. [8]. This

model fuses gradient, color and orientation information to simulate attention span and spread across the image. The method is highly parametric, as it involves the computation of a linear combination of weights to fuse multiple maps in order to arrive at the final saliency map. Inspired by this method, many residual spectral based models have appeared in the literature. Ciu et al. [9] as well as Guo and Zhang [10] found new ways to utilize the fourier phase spectrum to model bottom-up attention. Orabona et al. [21] extended the model of Itti et al. [8] to work on log-polar images, as they are more robust to inplane rotation of objects. This method is specifically fine tuned to detect salient regions, as it tries to compute edges by opponent color maps and does quantization of colors with existing gradient descent algorithms. The quantized colors are subsequently fed to an associative tensor field to classify, foreground and background. Experiments have revealed that the model of Orbona et al. [21] outperforms the original model of Itti et al. [8]. However, it has a high computational overload as the learning and the classification happen simultaneously for every image. Since all of these methods involve computing the same operations at multiple scales, it is recommend to downscale the input image in order to achieve a better runtime. Downscaling of the input image, has its own impact on the performance, as interpolation errors cascade themselves adversely on the results. Some of the more complex formulations of bottom-up saliency maps employ subband analysis, wavelet decomposition, and other spectral analysis methods. This again is constrained by parameters of the spectral decomposition methods. The other issue with such methods is to determine an image-specific filter to detect the most salient frequencies, which is indeed very complex.

In order to overcome the problem of computation of the training bases for every test image, multi-scale processing, and also to address the issue of image rescaling, Achanta et al. [11] proposed a contrast detection filter, to detect salient regions. Experimental results on the Berkley image segmentation dataset have proved the efficacy of this method, but requires parameter tuning for the filter mask. In [12] the same authors propose to use the absolute difference of each image pixel with image mean as its saliency value. This method is simple and achieves results that are comparable to more complex methods, without having any of their drawbacks. Another simple approach based on the distance transform was proposed by Rosin [13]. This method computes an edge map for each threshold and fuses them to generate a saliency map. The method is truly non-parametric, but fails to perform well when the edge contrasts are weak. In order to overcome the drawbacks of such global methods, several local contrast based approaches have been proposed. The popular approach of Bruce et al. [14] based on local contrasts and maximizes the mutual information between features using ICA bases. A set of ICA bases is

precomputed using a patch size of  $7 \times 7$  pixels. Then it is used to compute the conditional and joint distribution of features for information maximization. Experiments conducted on 120 eye tracking images have proven its efficacy. But the method is constrained by its emphasis on edges and neglects salient regions. It also adds a spurious border effect to the resultant image, and requires re-scaling of the original image to a lower scale. Another similar method based on the center-surround contrast paradigm was proposed by Gao et al. [15]. It computes the entropy of the feature distribution between a center patch and its surroundings. Though popular, it is constrained by the subjectivity involved in the computation of weights for fusing the different maps. The approach of Gopalakrishan et al. [16] which is also based on entropy of local features, selects between color and orientation entropy maps and avoids the problem of combining the maps altogether. Two other successful approaches proposed by Zhang et al. [17] and Seo and Milanfar [18] also operate on local contrasts with local regression kernels without multiscale operations and also avoids fusion of multiple maps. But, both of these methods have unwieldy computational runtime, which makes them unfavourable for real-time saliency detection. A recent center-surround contrast method was proposed by Avraham and Lindenbaum [19]. They propose to identify pre-attentive segments and to compute the mutual saliency between them. They further employ a Bayesian network to learn a stochastic image model, which will be able to accept or reject segments based on saliencies. An innovative application of this saliency map has been employed for pedestrian detection. The majority of the successful bottom-up saliency models proposed so far work at the patch level by capturing local contrasts. Some utilize a bases of local contrasts to eliminate spurious contrast sensitivity. The methods belonging to the local contrast paradigm are constrained by problem of arriving at a suitable patch size to compute the saliency operations.

In order to avoid the patch size parameters, many machine learning based bottom-up saliency models are proposed. A graph based visual saliency model is proposed by Harel et al. [20]. It implements a Markovian representation of feature maps, and utilizes a psychovisual contrast measure to compute the dissimilarities between features. Hou and Zhang [22] employ sparse representation and sparse feature coding length to improve the performance of their center-surround based maps. Kienzle et al. [23] further try to learn saliency maps by studying eye movements, and bulding a SVM classifier to simulate them. To make their method robust, they train on a set of natural images, where there are no man-made objects so that the attention does not have any bias. However, this approach requires enormous amounts of data to learn saliency weights reasonably.

A few hybrid approaches try to combine the aspects of multi-scale analysis, subband decomposition and as well

as contrasting center-surround differences. An Iso-center based saliency map computation proposed by Valenti et al. [25] utilizes color boosting, edge maps and Iso-center clustering. A conditional random field based learning of Color spatial distribution and center surround contrasts with a multiscale analysis is proposed and utilized for salient region detection [24]. Meur and Chevet [26] propose a fusion of the center-surround hypothesis and oriented sub-band decomposition to develop a saliency map useful for salient region detection. Bogdan et al. [27] introduce yet another radical approach which tries to detect salient regions by finding out the probability of detecting an object in a given sliding window. They employ the concept of super pixel straddling, coupled with edge density histograms, color contrasts and the model of Itti et al. [8] saliency map. A linear classifier is trained on an image dataset to build a bag-of-features to arrive at a prior for an object in an image. The method is theoretically very attractive, but is subjected to high variations in the performance as too many features, maps and parameters are involved which require fine tuning. Keeping in view the above discussion about the strengths and weaknesses of the contemporary models we now present the motivation behind developing our proposed model.

### 3. Motivation

Hierarchical methods like the one of Itti et al. [8] are influenced by global statistics. They also involve multiple feature fusion and multi-scale image analysis. This adds to the computational run-time of the model. Hierarchical methods ignore many of the statistically significant but locally occurring patterns, because they are constrained by global thresholds. Local contrast methods overcome this problem by taking into account local features and structures. But as we have seen in Section 2, methods which work on local contrasts like the ones of Zhang et al. [18] and Seo and Milanfar [17] require large number of training priors.

Similar to the hierarchical methods many of the local contrast methods have a bias towards corner points and strong edges, which may not be of interest. Hybrid and machine learning based methods like Bogdan et al. [7], Liu et al. [24] etc. are tuned to a particular dataset. Machine learning based methods also need a large set of parameters to be fine tuned. In view of this we develop our model to overcome the aforementioned issues, and compare it with the state-of-the-art algorithms, with a blend from hierarchical, local contrast and hybrid methods.

The methodology of Achanta et al. [12] serves as an example of a global method which avoids local statistics as well as hierarchical and multi scale processing of images. The classical Itti et al. [8] is among the best known hierarchical models. The models of Bruce and Tsotsos

[14], Seo and Milanfar [18] and Zhang et al. [17] are based on local contrasts, but use different features and training bases. And Harel et al. [20] is a successful hybrid model. We employ the aforementioned six methodologies in the course of this illustration as well as the experiments presented in the subsequent section.

Majority of the existing bottom-up saliency models have a bias towards edges and corners which are statistically significant, but not necessarily semantically salient. This lets them ignore salient regions in order to highlight the edges. The first row in the Fig 1, demonstrates this effect using the image of a glowing bulb. The saliency models are expected to highlight the salient region where the possibility is high that human attention is focussed. It can be noted that with an exception of methods of Achanta et al. [12] and our proposed model, all the others highlight the boundary than the region. The method by Bruce et al. [14] which works on information maximization awards more saliency to insignificant regions than the salient region itself. The performance of Achanta et al. [12] can be explained because, this method works on global statistics and is guaranteed to work on images with high contrast.

If we now observe the second row of Fig 1, we can see an image with cutlery and biscuit with low image contrast. The method of Achanta et al. [12] fails to detect any relevant salient region. The other methods are again more biased to edges and corner points as it can be observed from the saliency map of Seo and Milanfar [18]. Only our method uniformly highlights the regions without any bias towards edges.

Edges and other local statistics are significant in images which show man-made objects. Images showing natural scenes do not necessarily have significant local gradient information. This adversely affects the performance of saliency models which exclusively use local information. The model of Bruce and Tsotsos [14] does not detect any saliency and outputs a uniformly blurred map when the natural image shown in the third row of Fig 1 is given as an input. The models of Seo and Milanfar [18] and Zhang et al. [17] which have training bases to detect locally significant patches also fail to detect any salient region. Despite the image not having any strong edge information, and also having a lot of background variation our method outputs a more visually acceptable saliency map.

In view of this aforementioned discussion our bottom-up saliency model is simple and efficient to code. The model has the ability to work in real-time like the one proposed by Achanta et al. [12] without the need to rescale the images into a smaller scale. The model has only two parameters which requires tuning. Experiments have shown that it does not require rigorous cross-validation to find optimal values for these two parameters. The proposed model also avoids competition between multiple saliency maps which

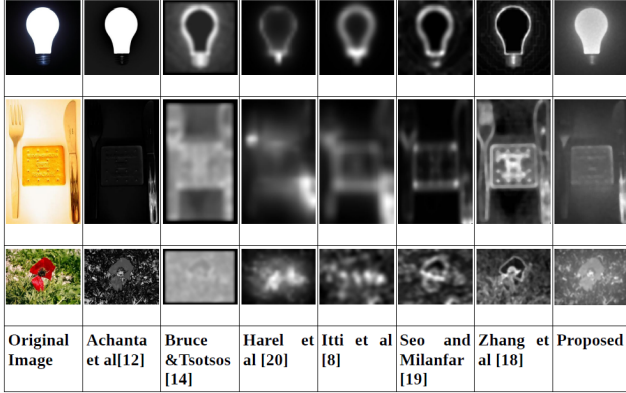


Figure 1. Illustration of the various saliency maps in comparison to the proposed saliency map. It can be observed from the last column that the proposed model highlights the region uniformly and consistently.

is in contrast to the models of Itti et al. [8], Harel et al. [20] etc and works in Lab color space which is more inspired by visual cognition. The proposed can be described in terms of visual receptive fields. Visual receptors are randomly scattered across the retina, and their density increases in the fovea [28]. Each receptive field, computes the saliency by taking into account the influence of every other stimulus present in the neighborhood, which is called stimulus bias [29]. The corresponding algorithm is presented in the next section.

#### 4. Proposed Saliency Map

Nosofosky [29] proposes that each stimulus is influenced by every other stimulus present in the attention space, which contributes to the cumulated saliency. A majority of the stimulus bias techniques have two components, the first one being the similarity function and second one being the biasing function. Based on this paradigm we propose a biasing formulation as follows.

Let  $I$  be an image of dimension  $H \times W$ . Let  $(x_1, y_1)$  and  $(x_2, y_2)$  be two distinct co-ordinate positions in  $I$ , with intensity values given by  $I(x_1, y_1)$  and  $I(x_2, y_2)$  respectively.

The attention value of  $I(x_1, y_1)$  due to  $I(x_2, y_2)$  is given by the function  $S(I, x_1, y_1, x_2, y_2)$  where,

$$S(I, x_1, y_1, x_2, y_2) = \frac{|I(x_1, y_1) - I(x_2, y_2)|}{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}} \quad (1)$$

Our input is an image and we do not consider patches or image segments as stimuli but we rather consider each pixel

as a stimuli. This is also in line with the center-surround contrast paradigm, except that we do not restrict ourselves to a pre-specified radius to choose pixels from.

Ahumada [28] proposed that the receptors of a stimulus should not be present in a regular arrangement, and that the density increases in and around fovea. Thus we propose to randomly generate a set of  $n$  random co-ordinates which act as receptive keypoints. We study the influence on the stimulus(pixel intensity value) that is present in each of these keypoints, by biasing it with a stimulus that is present in another set of  $n$  random co-ordinates in the image. The algorithm is as follows.

#### Algorithm : *Random Center Surround Saliency Map*

**Input** : (1) Image (in CIE Lab Space) of dimension  $H \times W \times 3$   
H:Height

W:Width

(2)  $\Delta_1$ , where  $0 < \Delta_1 < H \times W$

(3)  $\Delta_2$ , where  $0 < \Delta_2 < H \times W$

**Output** : Saliency Map(SM) of dimension  $H \times W$

#### Method

{

**Step 1** : Apply a Gaussian filter on image

**Step 2** : Decompose Image into L,a,b space

**Step 3** : Generate saliencies on each components

$S_L = \text{Saliency\_Map\_Compute}(L, \Delta_1, \Delta_2)$

$S_a = \text{Saliency\_Map\_Compute}(a, \Delta_1, \Delta_2)$

$S_b = \text{Saliency\_Map\_Compute}(b, \Delta_1, \Delta_2)$

**Step 4** : Generate final saliency map

$SM = \sqrt{S_L \cdot S_L + S_a \cdot S_a + S_b \cdot S_b}$

}

In order to deduce the saliency values to unattended locations we recommend to use any of the diffusion or image blurring techniques to propagate the values.

#### Algorithm : *Saliency\_Map\_Compute*

**Input** : (1) Image (in Gray Scale) of dimension  $H \times W$

H:Height

W:Width

(2)  $\Delta_1$ , where  $0 < \Delta_1 < H \times W$

(3)  $\Delta_2$ , where  $0 < \Delta_2 < H \times W$

**Output** : Interim Saliency Map (ISM) of dimension  $H \times W$

#### Method

{

**Step 1** : Set all elements of ISM to 0

**Step 2** : Start updating saliency values

for  $i=1$  to  $\Delta_1$

{

```

 $x_1 = \text{A Random Number in } 1 \leq x_1 \leq W$ 
 $y_1 = \text{A Random Number in } 1 \leq y_1 \leq H$ 
for j=1 to  $\Delta_2$ 
{
 $x_2 = \text{A Random Number in } 1 \leq x_2 \leq W$ 
 $y_2 = \text{A Random Number in } 1 \leq y_2 \leq H$ 
ISM( $x_2, y_2$ ) = ISM( $x_2, y_2$ ) + S(Image,  $x_1, y_1, x_2, y_2$ )
}
}

```

**Step 3** : Apply Median Filter on ISM

We employ Median filter to propagate saliency values because of their efficient runtime. Any other image blurring technique could also replace Median filter instead.

## 5. Experiments

We followed the experimental settings of Achanta et al. [12] and compared the considered models. The original MSR dataset for saliency consists of 5000 images with rectangular ROI annotation [24].

But Achanta et al. [12] have shown that such a rectangular ROI is inefficient. In turn Achanta et al. [12] took a subset of 1000 images, and recommended to use a mask which exactly described the segments in contrast to the rectangular ROI.

To compare our proposed model to the existing ones, we retain the six models that were used for illustration purposes in Section 4 describing the motivation. We did not implement any of these methods, but used the Matlab source codes directly released by the authors. The links for the source codes are available at the respective papers. However, we did not use the original implementation of Itti et al. [8], but instead use the one coded by Harel et al. [20], which is faster because of using Matlab MEX functions.

The  $\Delta_1$  and  $\Delta_2$  values are set to  $0.03 \cdot H \cdot W$  of an input image for our experiments.

Fig.2 shows the Recall-Precision performance of the models. It can be observed that the proposed methodology has the best curve among the other six state-of-the-art methods. Bruce and Tsotsos [14] though promising does not have a high performance because the dataset has a mix of natural images where entropy is uniformly distributed. Methods of Zhang et al. [17], Seo and Milanfar [18] also fail because they lay more emphasis on corners and edges than regions. Only the graph based method of Harel et al. [20] fares better, because it has no specific bias towards edges.

The model can be considered reliable if it performs equally well in Precision-Recall and Receiver Operating Character-

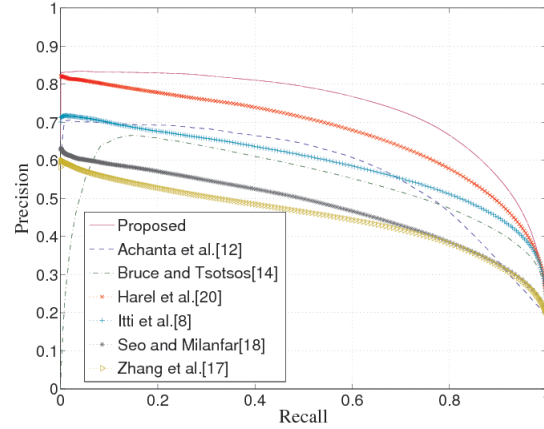


Figure 2. The Recall-Precision Plots of the proposed methodology and that of the other six state-of-the-art methods under consideration. The Proposed Model performs best as compared to the others in consideration.

istics(ROC) space consistently.

Hence we plotted the ROC curves as shown in Fig. 3 for the models under scrutiny. Our method performs consistently well even in ROC space, making the method reliable.

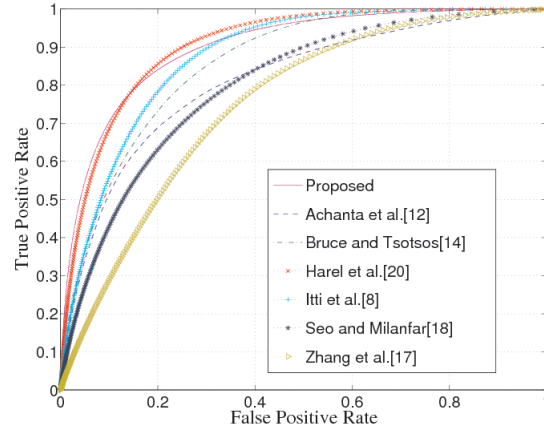


Figure 3. The ROC Plots of the proposed methodology and that of the other six state-of-the-art methods under consideration. The Proposed Model performs consistently as compared to the others in consideration.

We evaluate the performance of our model by changing the experimental parameters. It can be seen in Fig. 4, that the performance of the model dwindles when there is no distance normalization of saliency values. This confirms the assumption that influence of a stimulus over the other is inversely proportional to its distance.

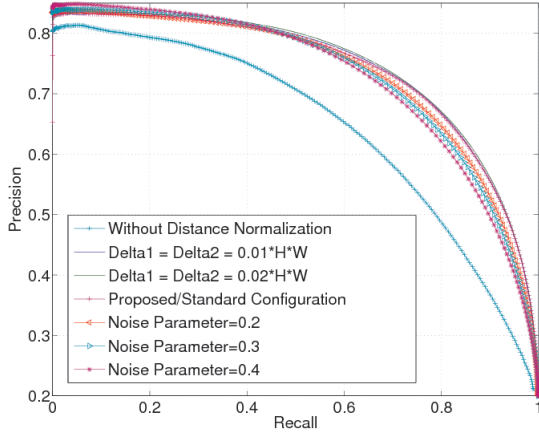


Figure 4. Performance analysis of the proposed methodology when various parameters are changed.

It is also important that the model performs well in presence of noise, and hence we added a salt and pepper noise of probability 0.2, 0.3 and 0.4 to the input images and re-ran the experiments. It can be seen that our model is robust to handle noise. We also changed the  $\Delta_1$  and  $\Delta_2$  parameters to  $(0.01 \cdot H \cdot W)$  and  $(0.02 \cdot H \cdot W)$  instead of the standard  $(0.03 \cdot H \cdot W)$  and evaluate the performance. Surprisingly they did not differ much from the standard setting that is followed in the experiments.

So far the experiments that were described in Figs. 2, 3 and 4 were from uniform thresholding of all images. But we also would want to evaluate when an image specific threshold is used, and validate the performance of the proposed methodology. In order to comprehensively evaluate the combined effect of Recall and Precision during thresholding, we employ the F-Measure. The F-Measure is defined as

$$F = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

It can be observed from Table 1, that the F-Measure assumes a maximum value of 0.73 with Iterative Means Image binarization technique. In order to evaluate the consistency of the proposed methodology over the different binarization techniques, we average all the performances, and the resultant average F-Measure is 0.62

We also conducted similar experiments with the rest of the methods in consideration. The average Recall, Precision and F-Measure due to the 12 different binarization

| Binarization            | Recall      | Precision   | F           |
|-------------------------|-------------|-------------|-------------|
| Concavity               | 0.97        | 0.37        | 0.51        |
| Entropy                 | 0.61        | 0.72        | 0.57        |
| Intermeans              | 0.85        | 0.69        | 0.72        |
| Iterative Intermeans    | 0.85        | 0.68        | 0.73        |
| Intermodes              | 0.79        | 0.68        | 0.65        |
| Maximum Likelihood      | 0.72        | 0.72        | 0.61        |
| Mean                    | 0.95        | 0.49        | 0.62        |
| Median                  | 0.97        | 0.36        | 0.51        |
| Min. Error              | 0.73        | 0.69        | 0.63        |
| Min. Error Likelihood   | 0.96        | 0.43        | 0.57        |
| Minimum                 | 0.69        | 0.74        | 0.59        |
| Moments                 | 0.84        | 0.66        | 0.71        |
| <b>Mean Performance</b> | <b>0.83</b> | <b>0.61</b> | <b>0.62</b> |

Table 1. Performance of the proposed method when different Image thresholding techniques are used from the Matlab Image Binarization Toolbox of [30]

| Saliency Models       | Recall      | Precision   | F           |
|-----------------------|-------------|-------------|-------------|
| Achanta et al. [12]   | 0.59        | 0.55        | 0.48        |
| Bruce et al. [14]     | <b>0.94</b> | 0.34        | 0.46        |
| Harel et al. [20]     | 0.78        | 0.53        | 0.55        |
| Itti et al. [8]       | 0.77        | 0.47        | 0.52        |
| Seo and Milanfar [18] | 0.61        | 0.43        | 0.42        |
| Zhang et al. [17]     | 0.62        | 0.41        | 0.42        |
| Proposed              | 0.83        | <b>0.61</b> | <b>0.62</b> |

Table 2. The average binarization effect due various methodologies on the different saliency models under consideration

techniques are presented in Table 2.

It can be observed from Table 2, that the proposed methodology scores highest among the F-measure at 0.62. The other best method with an F-measure of 0.55 magnitude is the graph-based method of Harel [20]. With this we can validate the consistency of the proposed model in both the cases where image thresholding is controlled as well as when it is independent. We now follow this up with the concluding note.

All the experiments were carried out using Matlab 7.0 installed on a DELL Precision T3400 Desktop, with Ubuntu 9.0, Intel 1.6 Ghz C2D Processor and 2 GB RAM.

## 6. Conclusions and Future Work

Treating random image pixels as receptor positions and using the Euclidean distance between two pixels as a normalizing function to weight the influence of a pixel on others is shown to model bottom-up attention effectively.

The method is not biased towards corner points or edges unlike the existing methods. Furthermore, it does not require any training bases or priors, and is shown to be most effective in salient region detection without the complexity of the many existing sophisticated methods. Due to its robustness to noise and runtime we would want to employ it to study attentional shifts in Humanoid Robots. The proposed methodology can also be conveniently plugged into the works of Bogdan et al. [27], Moosmann et al. [4], Dalal and Triggs [3] to improve their performances. We also plan to study the usage of SIFT [5] and SURF [6] keypoint locations instead of using uniform random number generated based receptor positions.

## 7. Acknowledgments

This work is supported by the RobotDoC Marie Curie Initial Training Network funded by the European Commission under the 7<sup>th</sup> Framework Programme (Contract No. 235065).

## References

- [1] A. L. Rothenstein and J.K. Tsotsos, "Attention links sensing to recognition," *Image and Vision Computing*, Vol. 26(1), pp. 114-126, 2008.
- [2] L. Elazary and L. Itti, "A Bayesian model for efficient visual search and recognition," *Vision Research*, Vol. 50(14), pp. 1338-1352, 2010.
- [3] Navneet Dalal and Bill Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, 2005, pp. 886-893.
- [4] Frank Moosmann, Diane Larlus and Frederic Jurie, "Learning Saliency Maps for Object Categorization," in *Proceedings of ECCV International Workshop on The Representation and Use of Prior Knowledge in Vision*, 2006.
- [5] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol. 60(2), pp. 91-110, 2004.
- [6] Herbert Bay, Andreas Ess, Tinne Tuytelaars and Luc Van Gool, "SURF: Speeded Up Robust Features," *Computer Vision and Image Understanding*, Vol. 110(3), pp. 346-359, 2008.
- [7] C. Kanan and G.W. Cottrell, "Robust Classification of Objects, Faces, and Flowers Using Natural Image Statistics," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [8] Laurent Itti, Christof Koch and Ernst Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 20(11), pp. 1254-1259, 1998.
- [9] X. Cui, Q. Liu and D. Metaxas, "Temporal spectral residual: fast motion saliency detection," in *Proceedings of the ACM international Conference on Multimedia*, 2009.
- [10] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," in *IEEE Trans. Image Processing*, Vol. 19(1), pp. 185-198, 2010.
- [11] R. Achanta, F. Estrada, P. Wils and S. Ssstrunk, "Salient Region Detection and Segmentation," in *Proceedings of the International Conference on Computer Vision Systems*, pp. 66-75, 2008.
- [12] R. Achanta, F. Estrada, P. Wils and S. Ssstrunk, "Frequency-tuned Salient Region Detection," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2009.
- [13] Paul L. Rosin, "A simple method for detecting salient regions," *Pattern Recognition*, Vol. 42(11), pp. 2363- 2371, 2009.
- [14] N.D. Bruce and J.K. Tsotsos, "Attention based on Information Maximization," in *Proceedings of the International Conference on Computer Vision Systems*, 2007.
- [15] Dashan Gao, Vijay Mahadevan and Nuno Vasconcelos, "The discriminant center-surround hypothesis for bottom up saliency," in *Proceedings of the Neural Information Processing Systems*, 2007.
- [16] Viswanath Gopalakrishnan, Yiqun Hu and Deepu Rajan, "Salient Region Detection by Modeling Distributions of Color and Orientation," in *IEEE Trans. Multimedia*, Vol. 11(5), pp. 892-905, 2009.
- [17] Lingyun Zhang, Matthew H. Tong, Tim K. Marks, Honghao Shan, and Garrison W. Cottrell, "SUN: A Bayesian Framework for Saliency Using Natural Statistics," *Journal of Vision*, Vol. 8(7), pp. 1-20, 2008.
- [18] Hae Jong Seo and Peyman Milanfar, "Static and Space-time Visual Saliency Detection by Self- Resemblance," *Journal of Vision*, Vol. 9(12), pp. 1-27, 2009.
- [19] Tamar Avraham and Michael Lindenbaum, "Esaliency (Extended Saliency): Meaningful Attention Using Stochastic Image Modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 32(4), pp. 693-708, 2010.
- [20] J.Harel, C.Koch, and P.Perona, "Graph-based visual saliency," in *Proceedings of the Neural Information Processing Systems*, 2007. pp. 545-552
- [21] F. Orabona, G. Metta and G. Sandini, "A Proto-object Based Visual Attention Model," in *Proceedings of the International Workshop on Attention in Cognitive Systems*, 2007.

- [22] Xiaodi Hou and Liqing Zhang, "Dynamic visual attention: searching for coding length increments," in *Proceedings of the Neural Information Processing Systems*, 2008. pp. 681-688.
- [23] W. Kienzle, F. A. Wichmann, B. Schlkopf and M. O. Franz, "Dynamic visual attention: searching for coding length increments," in *Proceedings of the Neural Information Processing Systems*, 2006. pp. 686-696.
- [24] T. Liu, J. Sun, N. Zheng, X. Tang and H. Shum, "Learning to Detect A Salient Object," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2007, pp.1-8.
- [25] R. Valenti, N. Sebe and T. Gevers, "Isocentric color saliency in images," in *Proceedings of the IEEE International Conference on Image Processing*, 2009.
- [26] O. Le Meur and J.C. Chevet, "Relevance of a feed-forward model of visual attention for goal-oriented and free-viewing tasks," *IEEE Trans. Img. Proc.*, 2010 (Inpress).
- [27] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2010
- [28] Albert J Ahumada, "Learning Receptor Position," *Computational Models of Visual Processing*, MIT Press, pp. 23-541991.
- [29] R.M. Nosofsky, "Stimulus bias, asymmetric similarity, and classification," *Cognitive Psychology*, Vol. 23(1), pp. 94-140, 1991.
- [30] <http://www.cs.tut.fi/~ant/histthresh/>.