Validating the Visual Saliency Model

Ali Alsam and Puneet Sharma

Department of Informatics & e-Learning (AITeL),
Sør-Trøndelag University College (HiST),
Trondheim, Norway
er.puneetsharma@gmail.com

Abstract. Bottom up attention models suggest that human eye movements can be predicted by means of algorithms that calculate the difference between a region and its surround at different image scales where it is suggested that the more different a region is from its surround the more salient it is and hence the more it will attract fixations. Recent studies have however demonstrated that a dummy classifier which assigns more weight to the center region of the image out performs the best saliency algorithm calling into doubt the validity of the saliency algorithms and their associated bottom up attention models. In this paper, we performed an experiment using linear discrimination analysis to try to separate between the values obtained from the saliency algorithm for regions that have been fixated and others that haven't. Our working hypothesis was that being able to separate the regions would constitute a proof as to the validity of the saliency model. Our results show that the saliency model performs well in predicting non-salient regions and highly salient regions but that it performs no better than a random classifier in the middle range of saliency.

Keywords: Saliency, fixations.

1 Introduction

A salient image region is defined as an image part that is clearly different from its surround [1]. This difference is measured in terms of a number of attributes, namely, contrast, brightness and orientation [2–6]. By measuring these attributes, visual saliency algorithms aim to predict the regions in an image that would attract our attention under free viewing conditions [4], i.e., when the observer is viewing an image without a specific task such as searching for an object. Finally, the output of the visual saliency algorithms is a so called saliency map which is a two dimensional gray scale map where the brighter regions represent higher saliency.

To evaluate the performance of visual saliency algorithms, the two dimensional saliency maps are compared with the image regions that attract observers' attention [7–14]. This is done by displaying to the observers a set of images and using an eye tracker to record their eye fixations. Further, it is thought that a higher number of fixations correspond to salient image regions. The recorded

J.-K. Kämäräinen and M. Koskela (Eds.): SCIA 2013, LNCS 7944, pp. 153–161, 2013.

fixations are thus compared with the associated visual saliency maps in a pair wise manner [8, 15–17]. Unfortunately, most studies have shown that while the saliency algorithms do predict a certain percentage of fixations they are far from being able to fully account for observers' visual attention [18, 19]. In fact, in a recent comprehensive eye tracking study by Judd et al. [20], it was shown that a dummy classifier defined by a Gaussian blob at the center of the image was better at predicting the eye fixations than any of the visual saliency models [1, 21, 22]. In other words, assuming that the eye fixations fall at the center of the image results in better prediction than an analysis of the image content. This finding is surprising and raises the question of whether our attention is indeed guided by salient image features.

In this paper we set about validating the saliency algorithm by means of an experiment in which we divided 200 images into regions which have received fixations and others that didn't. By collecting the values returned by the saliency algorithm local to those regions into two matrices we were able to use discrimination analysis to determine whether the data of the two matrices is separable. Our working hypothesis was that being able to separate the data using linear discrimination analysis would constitute a proof that the saliency algorithm is indeed in good correspondence with the eye fixations while failing to separate the data would constitute a proof that the saliency algorithm is a poor predictor of eye fixations.

In our experiment we found that the saliency algorithm predicts eye fixations almost perfectly in regions that don't attract any fixations and also in regions that attract many fixations. It is, however, a poor estimator of fixations in regions with middle saliency where the algorithm performs as a random classifier.

2 Brief Description of the Saliency Algorithm

Input to the algorithm is provided in the form of static color images. Three early features: color, intensity, and orientation are calculated from the input image. From these features several spatial scales are created using dyadic Gaussian pyramids [1]. Salient features are detected by using center-surround differences which are grounded in vision studies. The center-surround differences are calculated between the fine and the coarse scales followed by normalization. For details see [1]. Finally the resultant feature maps are combined linearly to form a so-called saliency map.

3 Experiments and Results

3.1 Data Set

The images and the associated fixations data used in the analysis were obtained from the comprehensive study by Judd et al. [20]. The data-set [20] includes 1003 images which were shown to 15 different observers with normal vision under free viewing conditions, i.e., the observers viewed the images without a specific task such as searching for an object.

Validating the Saliency Theory. In this experiment we set about validating the claim that the eye fixates on regions in the image that are salient or different with respect to their surround. To achieve an objective validation we chose to divide each image into two different sets of regions, in the first we have image regions which have attracted observers fixations and in the second set we have image regions that didn't attract fixations. The data was based on a subset of the images and corresponding fixations obtained by Judd et al. [20] where we used 200 landscape images and all the fifteen observers. The images were 1024 by 768 pixels in dimension and a fixated area was defined as a square region of dimensions 100 by 100 pixels where the center was located at the fixation point. Non-fixated areas were chosen randomly from parts of the image that had a region of a 100 by 100 pixels without any fixations. As an example, the fixated and the non-fixated regions for an image and the corresponding feature maps obtained by the saliency algorithm [1] are shown in figure 1.

By dividing the image into square regions that are classed as either fixated or not fixated we were able to assign a value to each square part that corresponds to the average of the intensity of the corresponding pixels in the saliency map obtained by Itti et al. [1]. In so doing we obtained two matrices, F and N_f where the elements in the vectors of F were the values of the averages of the feature maps based on the square regions centered at the fixation points while the vectors of N_f were the average values for non-fixated areas. Further we chose the number of non-fixated areas to be equal to that of the fixated regions, thus, the size of F was $n \times k$ were n was the number of fixations in all the 200 images and k was the number of feature maps was defined by the algorithm to be three maps pertaining to intensity, color and orientation.

Our main objective with the creation of the matrices F and N_f was to determine whether we can separate between the data of the two matrices using discrimination analysis or not. The basic idea was that being able to separate the data would constitute a proof that the fixations are indeed driven by low level features such as contrast and lightness as is the claim by researchers supporting the bottom up attention model. We further believe that the level of separation achieved between the fixated and non-fixated regions would offer us a clear view as to the goodness of the saliency algorithm in predicting the fixations. Thus if the prediction is random we can conclude, based on the available data set, that the idea that salient regions attract attention is false while a perfect separation would indicate that salient image regions dictate our visual attention.

We chose a simple discrimination method which involves calculating the difference vector between the averages of F and N_f and then projecting the vectors of F and N_f onto the difference vector to judge whether the data is separated along that vector or in other words whether F and N_f are significantly different. Mathematically, the operation are:

$$w = \mu_F - \mu_{N_f}, \tag{1}$$

where the size w corresponds to the number of feature maps (3 for the saliency algorithm), and μ_F and μ_{N_f} are the means along the columns of F and N_f.

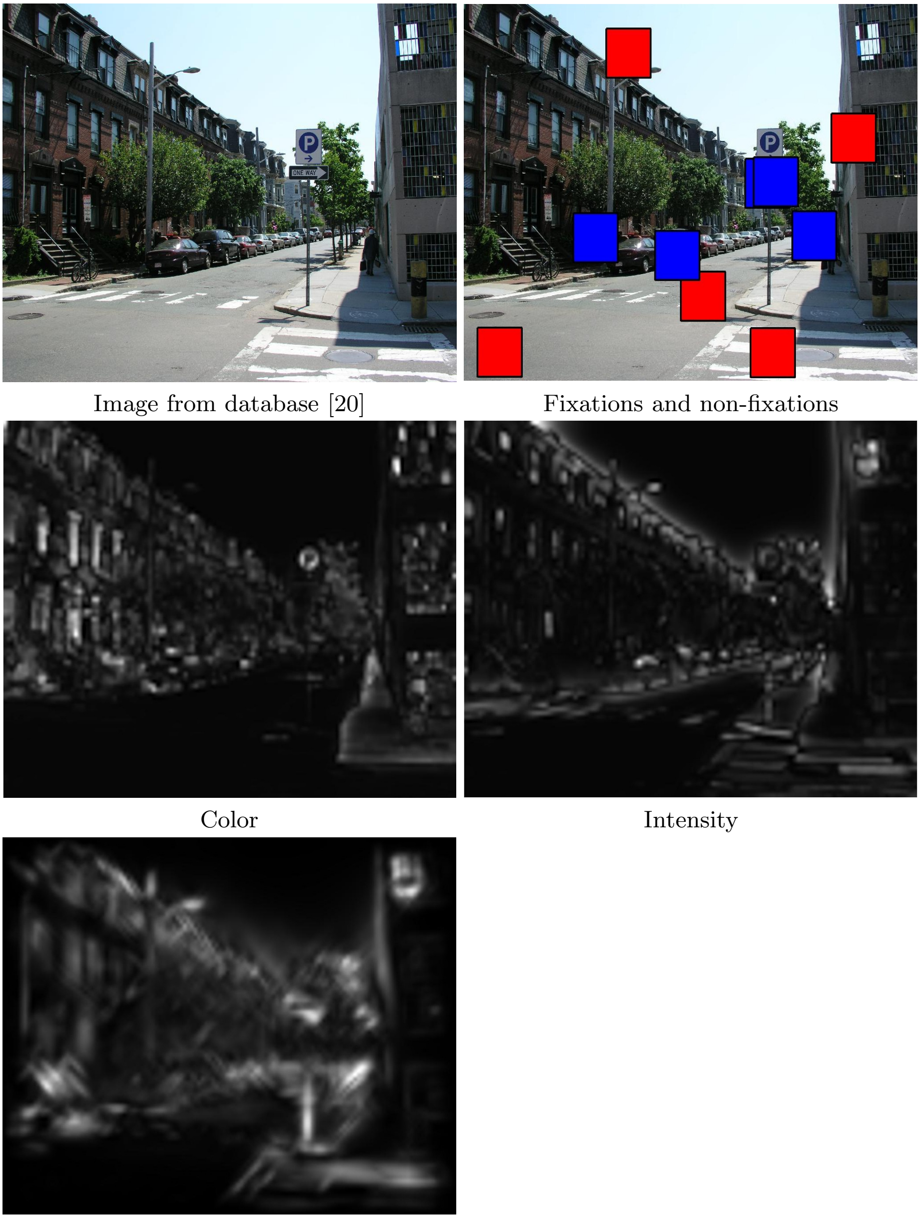


Image from database [20]

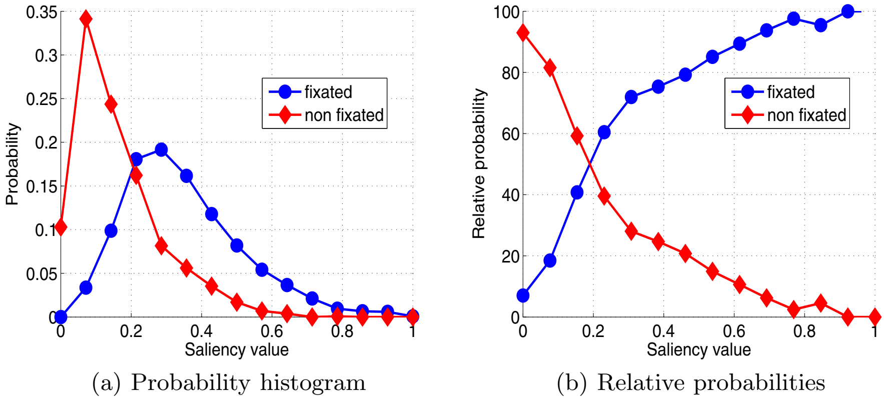
Fixations and non-fixations

Color

Intensity

Orientation

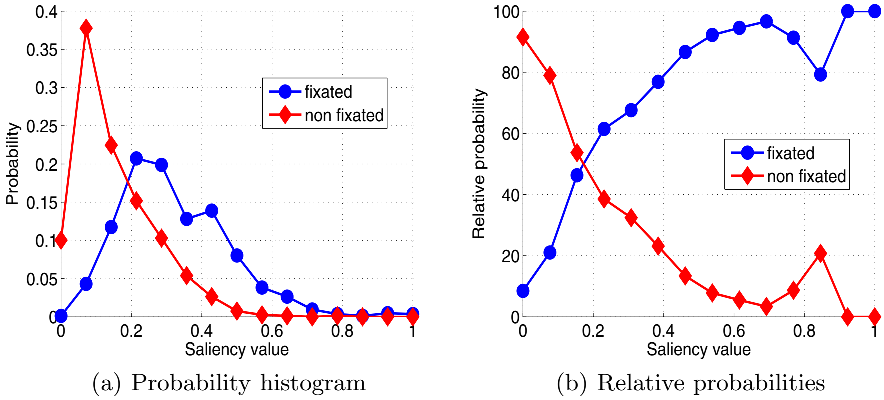
Fig. 1. Figure shows the fixated and the non-fixated regions for an image and the corresponding feature maps obtained by the saliency algorithm [1]. The fixated regions are marked as blue and the non-fixated regions are marked as red



(a) Probability histogram

(b) Relative probabilities

Fig. 2. Probability histograms and relative probabilities for the fixated and non-fixated regions for observer no 1. X-axis shows the saliency values obtained by using the visual saliency algorithm [1].



(a) Probability histogram

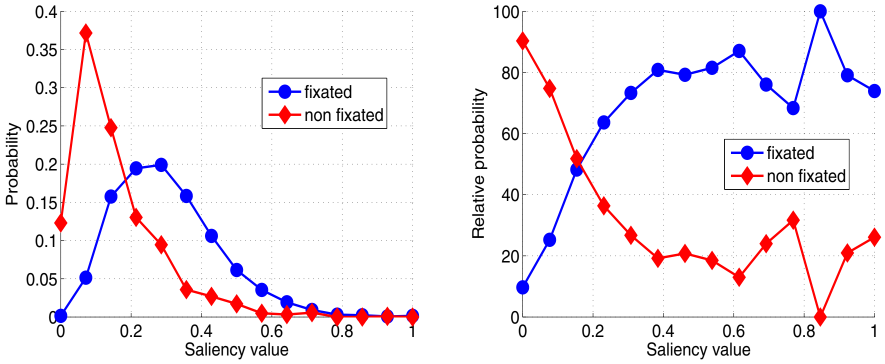
(b) Relative probabilities

Fig. 3. Probability histograms and relative probabilities for the fixated and non-fixated regions for observer no 2. X-axis shows the saliency values obtained by using the visual saliency algorithm [1].

$$p_F = wF \tag{2}$$

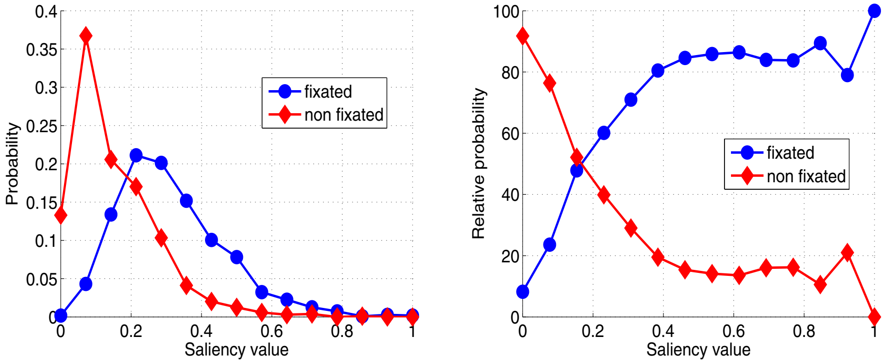
$$p_{N_f} = wN_f, \tag{3}$$

where the number of elements of the vectors p_F and p_{N_F} are 1 by k. The distribution of p_F and p_{N_F} provides a mathematical description of whether the fixated and non-fixated regions are indeed different as predicted by the saliency algorithm.



(a) Probability histogram (b) Relative probabilities

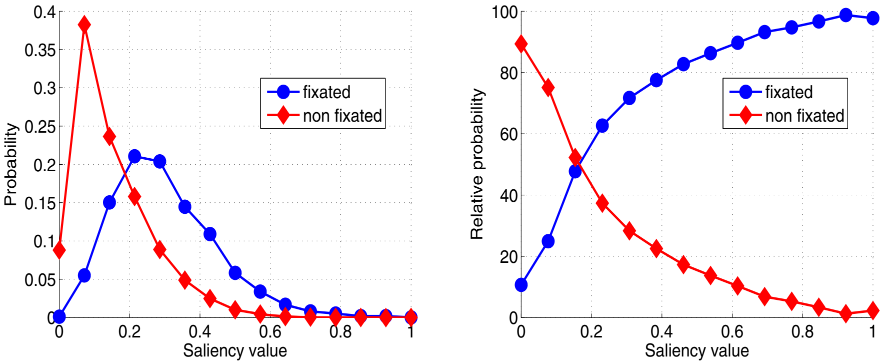
Fig. 4. Probability histograms and relative probabilities for the fixated and non-fixated regions for observer no 3. X-axis shows the saliency values obtained by using the visual saliency algorithm [1].



(a) Probability histogram (b) Relative probabilities

Fig. 5. Probability histograms and relative probabilities for the fixated and non-fixated regions for observer no 4. X-axis shows the saliency values obtained by using the visual saliency algorithm [1].

In figure 2, we plotted the probability histograms of p_F and p_{N_f}. Here, the histogram was normalized such that the area under the curve is one. We note that the separation between the two sets is not ideal but rather we find a considerable overlap between the two histograms specifically in the middle range. We further note that there is a clear separation between the two sets for regions of the images that received no fixations indicating that the method is good at predicting non-salient regions of the images. At a value of 0.3 the classification of the two sets is random.



(a) Probability histogram (b) Relative probabilities

Fig. 6. Probability histograms and relative probabilities for the fixated and non-fixated regions for an average observer. X-axis shows the saliency values obtained by using the visual saliency algorithm [1].

To gain better insight into the ability of the algorithm to separate the image regions into fixated and non-fixated, we plotted the relative probabilities of the histograms. For the non-fixated histogram, the relative probabilities were obtained by dividing the area under the non-fixated probability histogram curve of a specific bin i by the area under the fixated histogram curve for the same bin. For the relative probability of the fixated histogram the reciprocal value was calculated. Based on the fixation data of observer number one, this curve is plotted in figure 2 where we observe that for low salience values the separation of non-fixated regions is ideal and that the goodness of the separation declines to a level that is random. We also note that the separation of the highly salient regions, is nearly ideal. Based on this we can conclude that the algorithm is good in predicting non-salient and highly salient regions but its performance drops in the middle range. Assuming that the algorithm is a good representation of the way in which the human vision system functions we can state that flat regions which are almost never fixated while middle range contrast attracts fixations though not in every part and regions with very high saliency almost always attract fixations. This interpretation is of course dependent on the total number of fixations and the spatial distribution of the salient regions.

To generalize the analysis for the other observers, we performed the same calculations for all the observers and found similar trends in all cases. The results for observers two, three, and four shown in figures 3, and 5 respectively; and similar results were obtained for the fifteen individual observers. The results for the average observer based on all fifteen observers are shown in figure 6.

4 Discussion

In this paper, we performed a study to validate the claim that human eye fixations correspond to salient image features. We divided the image into regions which attracted fixations and others that were deemed by the observers as non-salient. By grouping the associated values for the feature maps obtained from the saliency algorithm by Itti et al. [1] into two matrices one pertaining to the fixated regions and an other to the non-fixated areas we were able to use linear discrimination to separate the regions optimally. Our working hypothesis was that being able to distinguish between the local values of the feature maps at fixated and non-fixated regions would indicate that the algorithm is indeed useful in predicting eye fixations. Our findings indicate that saliency algorithm by Itti et al. [1] is nearly ideal at predicting non-salient and highly salient regions with a considerable confusion in the mid saliency region.

References

1. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 1254–1259 (1998)
2. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology 4, 219–227 (1985)
3. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research 40, 1489–1506 (2000)
4. Itti, L., Koch, C.: Computational modelling of visual attention. Nature Reviews Neuroscience 2, 194–203 (2001)
5. Walther, D., Koch, C.: Modeling attention to salient proto-objects. Neural Networks 19, 1395–1407 (2006)
6. Underwood, G., Humphreys, L., Cross, E.: Congruency, Saliency and Gist in the inspection of objects in natural scenes. In: Eye Movements: A Window on Mind and Brain, pp. 563–579. Elsevier (2007)
7. Walther, D.: Interactions of Visual Attention and Object Recognition: Computational Modeling, Algorithms, and Psychophysics. PhD thesis, California Institute of Technology, Pasadena, California (2006)
8. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Proceedings of Neural Information Processing Systems (NIPS) (2006)
9. Cerf, M., Harel, J., Einhauser, W., Koch, C.: Predicting human gaze using low-level saliency combined with face detection. In: Advances in Neural Information Processing Systems (NIPS), vol. 20, pp. 241–248 (2007)
10. Henderson, J.M., Brockmole, J.R., Castelhano, M.S., Mack, M.: Visual Saliency Does Not Account for Eye Movements during Visual Search in Real-World Scenes. In: Eye Movements: A Window on Mind and Brain, pp. 537–562. Elsevier (2007)
11. Rajashekar, U., van der Linde, I., Bovik, A.C., Cormack, L.K.: Gaffe: A gaze-attentive fixation finding engine. IEEE Transactions on Image Processing 17, 564–573 (2008)
12. Meur, O.L., Callet, P.L., Barba, D., Thoreau, D.: A coherent computational approach to model bottom-up visual attention. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 802–817 (2006)
13. Borji, A., Itti, L.: State-of-the-art in visual attention modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 185–207 (2013)
14. Borji, A., Sihite, D.N., Itti, L.: Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. IEEE Transactions on Image Processing 22, 55–69 (2013)

15. Parkhurst, D., Law, K., Niebur, E.: Modeling the role of salience in the allocation of overt visual attention. Vision Research 42, 107–123 (2002)
16. Oliva, A., Torralba, A., Castelhano, M.S., Henderson, J.M.: Top-down control of visual attention in object detection. In: Proceedings of the 2003 International Conference on Image Processing, ICIP 2003, vol. 1, pp. 253–256 (2003)
17. Henderson, J.M.: Human gaze control during real-world scene perception. Trends in Cognitive Sciences 7, 498–504 (2003)
18. Tatler, B.W., Baddeley, R.J., Gilchrist, I.D.: Visual correlates of fixation selection: effects of scale and time. Vision Research 45, 643–659 (2005)
19. Tatler, B.W.: The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. Journal of Vision 7, 1–17 (2007)
20. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: International Conference on Computer Vision (ICCV) (2009)
21. Rosenholtz, R.: A simple saliency model predicts a number of motion popout phenomena. Vision Research 39, 3157–3163 (1999)
22. Cerf, M., Frady, E.P., Koch, C.: Faces and text attract gaze independent of the task: Experimental data and computer model. Journal of Vision 9, 1–15 (2009)