SALIENCY DETECTION VIA STATISTICAL NON-REDUNDANCY

Aanchal Jain, Alexander Wong, and Paul Fieguth[*]

Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada
{a27jain,a28wong,pfieguth}@uwaterloo.ca

ABSTRACT

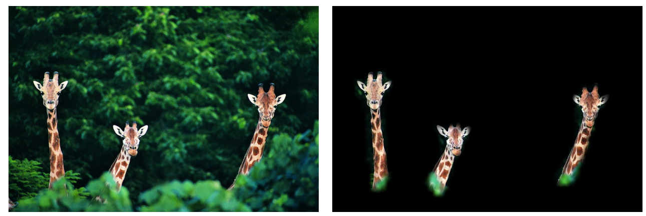
A novel algorithm based on statistical non-redundancy is proposed for saliency detection in natural images. By modeling site neighbourhoods as realizations of other site neighbourhoods under a Gaussian process, the saliency of any arbitrary site can be characterized by the statistical non-redundancy of its site neighbourhood with respect to other site neighbourhoods in a given image. Preliminary results using natural images show that the proposed method provides improved precision vs. recall characteristics over previous methods such as spectral residuals and spectral whitening.

1. INTRODUCTION

The automatic detection of saliency in images forms an important step in computer vision tasks such as image segmentation [1] and object recognition [2]. Traditional models of saliency detection are based on training, which makes them unsuitable for generalized tasks. As such, it is much desired to develop automatic methods for detecting saliency in images, without the need for prior knowledge of image content.

In recent years many general purpose saliency detection algorithms have been proposed. Itti and Koch proposed a saliency model based on the human visual search process [3, 4, 5, 6, 7]. Hou and Zhang proposed a spectral residual model [8], where the log-spectrum of an input image is analysed and the residual of the image is extracted in the spectral domain, which is then used to construct a saliency map in the spatial domain. This method simulates the behaviour of pre-attentive visual search and is an efficient method for early stage visual processing. Wang et al. proposed a context-based model [9], where saliency is viewed as an anomaly relative to a context, which could be local or global, however this method needs a database of images for obtaining the extrinsic saliency map. Murray et al. [10] proposed an efficient model of color appearance in human vision, which contains a principled selection of parameters as well as innate spatial pooling mechanism and can be generalized to obtain a saliency model, with scale integrated via a simple inverse wavelet transform over the set of weighted center-surround outputs and the reduction of ad-hoc parameters.

[*]The authors would like to thank the Natural Sciences and Engineering Research Council (NSERC) of Canada for supporting this research project.



(a) original image (b) object image

Fig. 1: In this example, the giraffes are the salient (unique) objects of interest that we wish to detect, as shown in the object image, relative to the background. In this figure the object image is based on the ground truth.

Other methods include the frequency-tuned approach proposed by Achanta et al. [11], and the context-aware approach proposed by Goferman et al. [12].

An observation that can be made with regards to saliency in images is that it is closely related to visual uniqueness in image attributes such as colors, edges, and boundaries. For example, in natural images such as that shown in Fig. 1a, the salient objects of interest (e.g., giraffes) exhibit very different visual characteristics than the rest of the scene (the forest). From an information-driven perspective, one can also say that the object of interest is unique as it has low information redundancy within the given image compared to the rest of the scene, which is highly redundant. Motivated by this observation, our proposed work aims to detect saliency within an image without prior knowledge regarding the underlying image content by quantifying this non-redundant nature of salient objects using a statistical modeling approach. This quantification is accomplished in the proposed method by computing the statistical non-redundancy of each site in the image, given the other sites in the image based on site neighbourhoods. The resulting non-redundancy gives a saliency map that provides a rough classification of the image into regions which are salient (high statistical non-redundancy) and non-salient (low statistical non-redundancy), and can be used to identify the salient objects of interest.

This paper is organized as follows. The theory behind the proposed method is presented in Section 2. The experimental results using natural images are presented and discussed in Section 3 with conclusions drawn in Section 4.

2. STATISTICAL NON-REDUNDANCY

In natural images, salient objects of interest can be characterised by their statistical non-redundancy with respect to a given image. Our method aims to quantify the statistical non-redundancy to obtain a saliency map which divides the image into salient and non salient regions, which can then be used to obtain object maps.

2.1. Statistical non-redundancy between sites

Let N_i be a set of image pixels in the neighbourhood of site i. Suppose that the neighbourhood around an arbitrary site i in an image can be seen as a realization of the neighbourhood around some other site j:

$$N_i = N_j + \eta_{ij}, \tag{1}$$

where η_{ij} is a noise process between the two site neighbourhoods following some distribution $P(\eta_{ij})$. Suppose we model $P(\eta_{ij})$ as an independent and identically distributed Gaussian process with zero-mean and variance σ. Therefore, the probability of an arbitrary site k in the neighbourhood N_i being a realization of the corresponding site in the neighbourhood N_j can be defined as:

$$P(N_i^k|N_j^k) = e^{\frac{-(N_i^k - N_j^k)^2}{\sigma^2}} \tag{2}$$

Therefore, one can quantify the statistical redundancy between the neighbourhoods as:

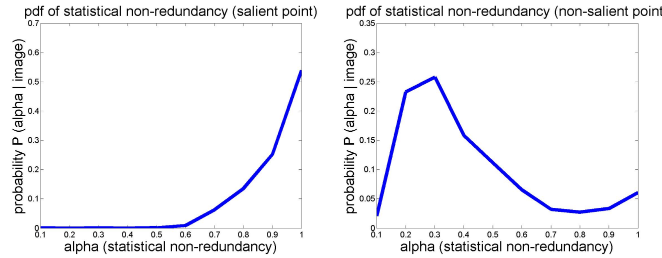
$$P(N_i|N_j) = \prod_k P(N_i^k|N_j^k). \tag{3}$$

Since we are interested in quantifying the saliency of a site, and sites associated with salient objects of interest have low statistical redundancy, one can quantify the statistical non-redundancy between two sites i and j (denoted as α_{ij}) as

$$\alpha_{ij} = 1 - P(N_i|N_j) = 1 - \prod_k P(N_i^k|N_j^k). \tag{4}$$

2.2. Statistical non-redundancy within the image

Given the aforementioned model for statistical non-redundancy between sites, we wish to determine a metric for quantifying the statistical non-redundancy of a site with respect to the given image. For a particular site i, let us compute α_{ij} over all possible sites j within the given image, so that we can compute the probability of α_{ij} given the image f (denoted as $P(\alpha_{ij}|f)$). What $P(\alpha_{ij}|f)$ gives us is an overall picture of the statistical non-redundancy of site i given the image f. Fig. 2 shows $P(\alpha_{ij}|f)$ for two different sites within the same image. We can see that the salient site has a higher degree of non-redundancy as compared to the non-salient site. This observation agrees with our expectation as the salient sites would have fewer matching neighbourhoods. Therefore, it can be observed that the overall statistical non-redundancy trend of a site gives us a good indicator of image saliency.



(a) salient site (b) non-salient site

Fig. 2: $P(\alpha_{ij}|f)$ for two sites in the same image. As expected, the distribution shows a significant saliency sensitivity.

2.3. Saliency map

We have shown that salient sites in the image have a higher overall statistical non-redundancy than non-salient sites. Salient regions in the image are unique and the neighbourhoods centred around the salient sites have fewer matching neighbourhoods as compared to the non-salient sites. We can use this fact to quantify the statistical non-redundancy of every site in the image and thus generate a saliency map, where the intensity at each site is proportional to the statistical non-redundancy of the corresponding site in the original image. In the proposed method, for each site i in image f, we compute $P(\alpha_{ij}|f)$ based on all sites j in f. For simplicity, we then define the overall statistical non-redundancy S_i of site i as the expected value of α_{ij} given image f:

$$S_i = \underset{j}{E}(\alpha_{ij}|f) \tag{5}$$

The collective map of S_i for all sites in f forms the saliency map of the image. The saliency map divides the image into high intensity and low intensity sites. The high intensity points in the saliency map correspond to the salient sites in the original image. This intuitively makes sense as a site corresponding to a unique region would have low information redundancy within the image and hence a higher value for statistical non-redundancy. For color images, the saliency map is based on the aggregate expectation across all color channels.

2.4. Generating the object image

Given the computed saliency map, one can get a rough idea of where the salient objects of interest are by applying an appropriate threshold on a saliency map, which can then be used to generate an object image. Choosing a low threshold would unnecessarily include sites from the background (false positives) in the salient object region of the object image. On the other hand choosing a high threshold would lead to sites from the object region being neglected in the object image (false negatives). For pure illustrative purposes, we generate

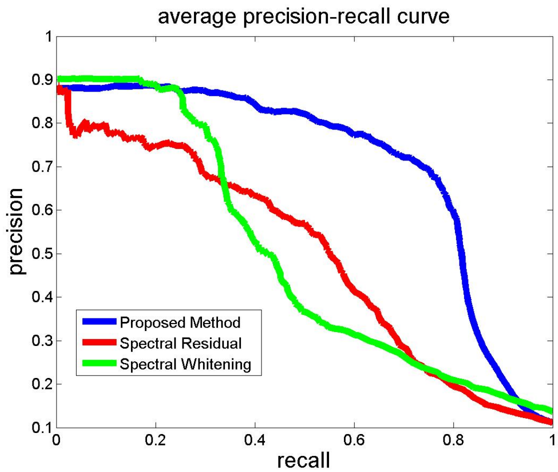


Fig. 3: Average precision-recall curves for three different methods over a set of test images. The proposed method gives substantially better results compared to spectral residual and spectral whitening methods.

an object image by thresholding the original image using the mean absolute deviation of the saliency map as the threshold (denoted by T):

$$T = E(|X - E(X)|). \tag{6}$$

For our purposes, X is the vector containing all sites in the saliency map.

3. PERFORMANCE EVALUATION

We have compared the performance of our algorithm (which we will denote as NR) against spectral whitening (SW) and spectral residual (SR) [8]. The test images used for our tests are the same as those used in [8]. For each natural scene image, there are four images which correspond to the object identification done by the four human subjects. The combination of these four images marked by the human subjects forms the ground truth.

Rather than computing an object image using an automatic thresholding method, which can give inconsistent and unreliable results depending on the method chosen, we perform a more rigorous evaluation process by computing the precision vs. recall plot for all possible thresholds. This approach provides a much better visualization on how the proposed method performs independent of arbitrary thresholds.

3.1. Results and discussion

Fig. 3 shows the average precision-recall curve taken over the test images. The precision-recall curve shows that the proposed NR method provides better performance than the SR and SW methods.

Example images with their corresponding saliency maps generated using the tested methods are shown in Fig. 4, with

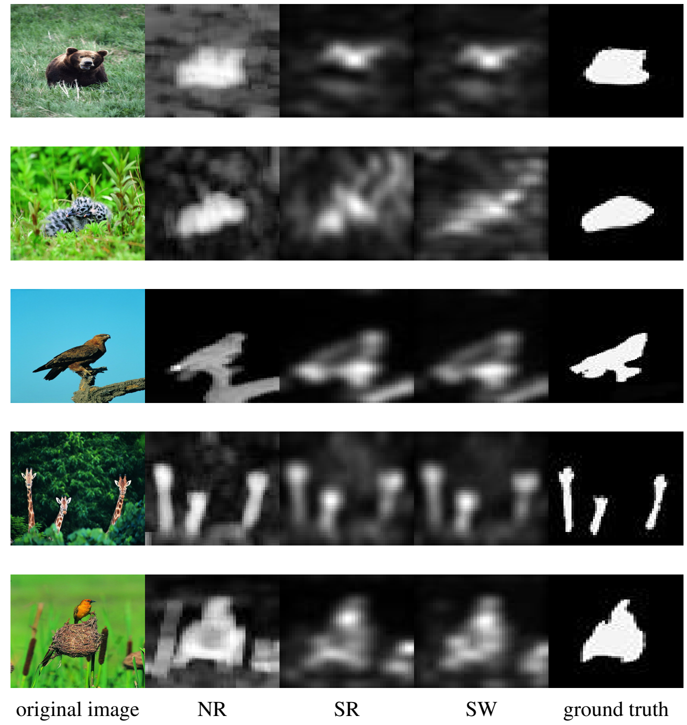


Fig. 4: Saliency maps for some tested natural images. In all the cases we can see that the saliency maps obtained from the proposed method (NR method) are closer to the ground truth as compared to the saliency maps obtained from spectral residual and spectral whitening methods.

the ground truth images also shown for reference. It can be observed that, with reference to the ground truth images, the salient objects within the images are well defined in the saliency maps obtained using the NR method when compared to the other methods. On the other hand the maps obtained from the other methods are relatively diffused and the salient regions detected in the maps focus more on the edges rather than the entire salient objects of interest.

Fig. 5 shows the object images as identified by the three algorithms under consideration. Since the proposed NR method gives detailed saliency maps, good object images that capture the objects of interest in their entirety can be obtained. Based on these preliminary findings, the proposed NR method shows great potential as the first step in object recognition and other such image processing and computer vision tasks.

4. CONCLUSIONS AND FUTURE WORK

In this paper, a novel method for saliency detection in images based on statistical non-redundancy was investigated. Preliminary results show that the proposed method holds strong potential for identifying salient objects of interest within an image without the need for prior knowledge about underlying image content, making it a good addition as a first step to many computer vision tasks.

While the proposed method shows great promise as the first step in many computer vision and image processing tasks, a much more thorough and comprehensive study needs to be performed to fully understand the theoretical underpinnings of the method, as well as discover ways to improve the initial framework. For example, the performance of the proposed method at different resolutions need to be investigated to evaluate its performance for different resolution images. Furthermore, while a Gaussian distribution is used to model the statistical relationship between sites, further study will involve investigating the use of different statistical models, which can be useful for different applications such as medical and remote sensing image processing and analysis. Finally, further investigation into how the overall statistical non-redundancy is quantified would be interesting.

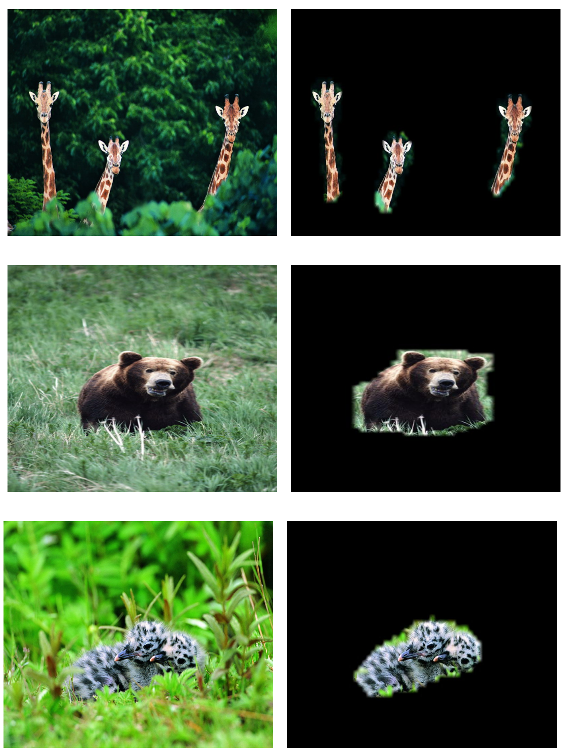


Fig. 5: Object images generated from proposed method (non-redundancy method). The algorithm very clearly marks out the regions which are visually salient with respect to the rest of the image. These form the objects of interest.

5. REFERENCES

[1] J. Han, K. Ngan, M. Li, and H. Zhang, "Unsupervised extraction of visual attention objects in color images," vol. 16, no. 1, pp. 141–145, 2006.

[2] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2004, pp. 37–44.

[3] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision research*, vol. 40, no. 10-12, pp. 1489–1506, 2000.

[4] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature reviews neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

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

[6] L. Itti and C. Koch, "A comparison of feature combination strategies for saliency-based visual attention systems," in *SPIE human vision and electronic imaging IV*. Citeseer, 1999, vol. 3644, pp. 473–482.

[7] L. Itti and C. Koch, "Feature combination strategies for saliency-based visual attention systems," *Journal of Electronic Imaging*, vol. 10, pp. 161, 2001.

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

[9] M. Wang, J. Konrad, P. Ishwar, K. Jing, and H. Rowley, "Image saliency: From intrinsic to extrinsic context," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 417–424.

[10] N. Murray, M. Vanrell, X. Otazu, and C.A. Parraga, "Saliency estimation using a non-parametric low-level vision model," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 433–440.

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

[12] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2376–2383.