OpenAIR @RGU ROBERT GORDON UNIVERSITY ABERDEEN

This publication is made freely available under GREEN open access.

AUTHOR(S): CHILUKAMARI, J., KANNANGARA, S. and MAXWELL, G.

TITLE: A low complexity visual saliency model based on in-focus regions and centre sensitivity.

YEAR: 2014

Publisher citation: CHILUKAMARI, J., KANNANGARA, S. and MAXWELL, G. 2014. A low complexity visual saliency model based on in-focus regions and centre sensitivity. In Proceedings of the 4th IEEE international conference on consumer electronics - Berlin (ICCE-Berlin 2014), 7-10 September 2014, Berlin, Germany. New York: IEEE [online], pages 411-414. Available from: https://doi.org/10.1109/ICCE-Berlin.2014.7034250

OpenAIR citation: CHILUKAMARI, J., KANNANGARA, S. and MAXWELL, G. 2014. A low complexity visual saliency model based on in-focus regions and centre sensitivity. In Proceedings of the 4th IEEE international conference on consumer electronics - Berlin (ICCE-Berlin 2014), 7-10 September 2014, Berlin, Germany. New York: IEEE, pages 411-414. Held on OpenAIR [online]. Available from: https://openair.rgu.ac.uk

Publisher copyright statement:

This is the AUTHOR ACCEPTED version of proceedings originally published by IEEE and presented at 4th IEEE international conference on consumer electronics - Berlin (ICCE-Berlin 2014), 7-10 September 2014, Berlin, Germany. (ISBN 9781479961665; eISBN 9781479961658; ISSN).

© 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.

OpenAIR takedown statement:

Section 6 of the "Repository policy for OpenAIR @ RGU" (available from http://www.rgu.ac.uk/staff-and-current-students/library/library-policies/repository-policies) provides guidance on the criteria under which RGU will consider withdrawing material from OpenAIR. If you believe that this item is subject to any of these criteria, or for any other reason should not be held on OpenAIR, then please contact openair-help@rgu.ac.uk with the details of the item and the nature of your complaint.

This publication is distributed under a CC BY-NC 4.0 license.

https://creativecommons.org/licenses/by-nc/4.0

OpenAIR at RGU

Digitally signed by OpenAIR at RGU
DN: cn=OpenAIR at RGU, o=Robert Gordon University, ou=Library, email=publications@rgu.ac.uk, c=GB
Date: 2017.08.31 11:33:06 +01'00'

A Low Complexity Visual Saliency Model Based on In-Focus Regions and Centre Sensitivity

Jayachandra Chilukamari, Sampath Kannangara and Grant Maxwell
School of Engineering, IDEAS Research Institute, Robert Gordon University, Aberdeen, UK

***Abstract*—A novel low-complexity visual saliency detection algorithm for detecting salient regions in images is proposed. The algorithm derives salient regions based on in-focus regions and image centre sensitivity. The performance of the algorithm in predicting human eye fixations is validated against ten state-of-the-art algorithms using a public image dataset. The results demonstrate that the proposed algorithm achieves higher prediction accuracy in saliency detection at significantly lower computational complexity compared to other algorithms.**

***Index Terms*—visual saliency, visual attention models, saliency detection**

I. Introduction

Visual saliency detection models are designed to predict or detect the regions of an image or of a video frame that the Human Visual System (HVS) typically considers to be important. These models are employed in many application areas such as target tracking [1], image and video compression [2], human robot interaction [3], navigation [4] and dynamic lighting [5]. For this reason, a significant amount of research has been done in last few years in the development of visual saliency models.

Existing saliency models can be broadly categorized into two different types: bottom-up and top-down models. Bottom-up saliency models focus on detecting specific features within images that attract human attention. Some of the bottom-up features that grab viewer attention are color, intensity and orientation. Similarly top-down features are usually the ones which deal with high level cognitive factors such as expectations, task demands and emotions. Modeling bottom-up features is relatively simple task compared to top-down features. This is because bottom-up features are generic or common features present in the images, whereas top-down influences vary among individuals and are therefore difficult to model [6].

Pioneering work on visual saliency has been done by the authors of [7] in developing a biologically inspired model based on low level features (intensity, color and orientation). High computational cost is a limitation of this model. The authors of [8] proposed Spectral Residual (SR) model that does not depend on features or categories. This model is further improved in Phase Quaternion Fourier Transform (PQFT) [9] model. PQFT is a bottom-up saliency model built using intensity, two colour channels and a motion channel for detecting saliency in static and dynamic scenes. Both SR and PQFT are extremely fast; however, they achieve low prediction accuracy. Graph Based Visual Saliency (GBVS) [10] is a bottom-up model which achieves high prediction accuracy by forming activation maps for some feature channels and implicitly optimizing the model for image centre; however, high complexity is still a weakness of this model. The authors of [11] proposed a bottom-up model based on wavelet transforms. High complexity limits its usage in real time applications. A Discrete Cosine Transform (DCT) based saliency model is proposed in [12]. A fast saliency detection algorithm proposed in [13] considers centre sensitivity, however, achieves low prediction accuracy. The authors of [14] sample the image into rectangular regions of interest and then compute the local saliencies. However, significant computational complexity is a drawback. The Saliency Using Natural statistics (SUN) [15] and Context Aware Saliency (CAS) [16], incorporates both bottom-up and top-down features to improve the performance in predicting human fixations. However, detecting a high number of features results in increased complexity.

In our earlier work we proposed a saliency model based on detecting in-focus regions using DCT coefficients [17]. In this work we present an improved algorithm where the complexity is reduced by using an Integer Cosine Transform (ICT) and the accuracy is improved by using a perceptual colour space and incorporating the centre sensitivity of the HVS.

II. Proposed Algorithm

Typically, the viewer attention is led to a specific region of an Image by bringing the region into focus (in-focus). Therefore, in our earlier model [17], the in-focus areas were assumed to be the salient regions of the image. It was observed that the 8x8 DCT spatial frequencies of the entire image were dominated by the peak frequency components of the in-focus regions of the image. Therefore, the peak frequency components of the 8x8 DCT blocks of the overall image were used to generate the in-focus region map of the image. Building on our earlier model, in this work we use Integer Cosine Transform [18] coefficients to identify the salient frequency coefficients. The use of the integer transform, compared to the DCT (which is a floating point calculation), significantly reduces the complexity of the model. When tested on an image with resolution 1024x768 images, the focus detection using DCT took 0.63 sec, whereas using ICT it takes only 0.33 sec to generate the focus map. Therefore by using ICT there is 47% of complexity reduction. This motivated us to use ICT in the current work. The previous model used the

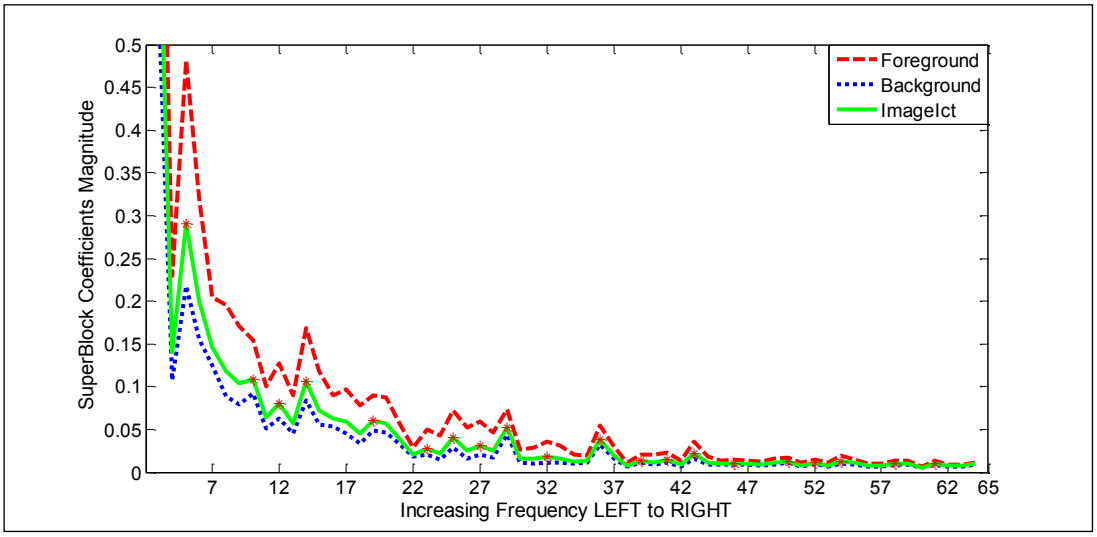
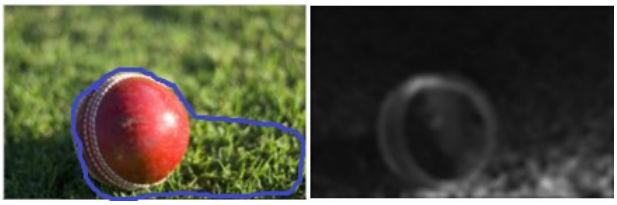


Fig.1.Superblock coefficients magnitude vs. Zig-zag scanned frequencies

Y (luminance) component of the YCbCr color space to calculate the spatial frequencies; however, this work uses the value channel of the perceptual HSV color space to calculate the integer transform coefficients. The mean of all 8x8 ICT blocks known as superblock of the in-focus region, out-of-focus region and the entire image are calculated respectively. In Fig.1 the magnitudes of all the three superblocks are zig-zag scanned and plotted against each other. Similar to the previous work the frequency coefficients corresponding to the peaks of the image ICT are summed up to generate a frequency map. Experiments across number of images have shown that the magnitude difference between in-focus and out-of-focus region tends to be significant between frequency ranges 2 and 30 and therefore only these frequencies are chosen to obtain the frequency map. This map is later Gaussian smoothed and its contrast is re-normalized in order to obtain a focus map. An example image with enclosed regions in-focus and its corresponding focus map is shown in Fig.2.

Further, we incorporate the center sensitivity of the HVS in this new model. Research on human eye saccades and fixations has shown that viewers are highly sensitive to the image center rather than the periphery [19]. The factors that are responsible for this phenomenon are photographer bias and viewing strategy. Photographer bias is the natural tendency of the



(a) (b)

Fig.2.Focus map generation shown in Fig.5.(a) Image with enclosed region in-focus (b) Contrast stretched focus map

viewers to put the objects at the centre of the image to emphasize their importance. Viewing strategy is a situation in which viewers repeatedly reorient themselves towards the centre relative to other locations of the scene. Whenever there is a scene cut or new scene available viewers start orienting towards the centre however as they become familiar with the content the orientation may change. Some of the other less influential factors are motor bias [20], orbital reserve [21], screen centre [22], low sensitivity of the HVS towards the periphery of the human eye [23] and high level influences [19]. Humans highly prefer to take short horizontal saccades rather long saccades. Due to this once the viewers look at the centre the saccade sequence tend to cluster around the starting point. Orbital reserve refers to preference of straight head position while viewing the scene. Some models such as [13], consider center sensitivity to generate saliency maps. In this work, a 2D anisotropic Gaussian intensity map is used to generate a center sensitivity map. The standard deviations σ_x and σ_y are calculated as a fixed ratio of the vertical and horizontal dimensions of the focus map as shown below.

$$\sigma_x = (c * w)/6 \quad (1)$$

$$\sigma_y = (c * h)/6 \quad (2)$$

Where c is a fraction of height (h) and width (w) of the original image. The final saliency map is generated by combining the focus and centre maps according to the following weighted (α) equation.

$$SaliencyMap = \alpha \cdot FocusMap + (1-\alpha) \cdot CentreMap \quad (3)$$

The generated saliency map is later up-sampled to the original image resolution using bi-cubic interpolation.

TABLE I: PERFORMANCE COMPARISON OF SALIENCY MODELS

Model	CC	NSS	AUC	Complexity (sec)
NVT [7]	0.23	1.09	0.76	0.76
GBVS [10]	0.29	1.35	0.81	1.18
SR [8]	0.18	0.85	0.69	0.01
SUN [15]	0.15	0.75	0.67	9.36
PQFT [9]	0.12	0.59	0.56	0.83
CAS [16]	0.22	1.05	0.74	45.96
SS [12]	0.23	1.08	0.74	0.03
RCSS [14]	0.23	1.05	0.75	5.01
SDSP [13]	0.21	0.97	0.72	0.06
WBSD [11]	0.18	0.88	0.71	24.28
Previous work [17]	**0.18**	**0.84**	**0.71**	**0.90**
PROPOSED	**0.30**	**1.36**	**0.81**	**0.34**

A. Parameter tuning

Due to the large number of parameters and value ranges involved, the parameters of this model were tuned using the hill climbing method [24] to maximize the correlation between ground truth fixation maps (GTFM) and the generated saliency maps.

The prediction accuracy is evaluated using Correlation Coefficient (CC) during parameter tuning and is later cross verified using another two metrics namely Normalized Scan path Saliency (NSS) and Area Under Curve (AUC) metrics. The parameters that need to be tuned are:

i) Gaussian kernel size n ($n \times n$ kernel) used to blur the focus map.
ii) Contrast stretching factor k of the focus map.
iii) Fraction c of the centre map, used to calculate the standard deviations of the 2D Gaussian centre sensitivity distribution.
iv) Weighting parameter α, used to combine the generated centre map and the focus map.

The tuned parameter values were estimated to be, $n = 12$, $k = 2$, $c = 0.95$, $\alpha = 0.35$.

III. RESULTS

The performance of the proposed algorithm is validated on Judd *et al.* dataset [25]. The dataset consists of 1003 images, among them 779 are landscape and 228 are portrait images. The dataset includes ground truth eye tracking data collected from 15 users. Table 1 shows the performance of the proposed model compared against 10 state-of-the-art models on this dataset. It is evident that the proposed model performs better than all evaluated state-of-the-art models when measured using CC and NSS metrics. It also performs equally as well as GBVS (next best model) in-terms of accuracy for the AUC metric. Four sample images from the dataset along with their saliency maps are shown in Fig.3. The complexity was analysed on a PC running with 16 GB RAM and Quad core 3.40 GHz Intel core i7-2600K CPU. The average time required to compute a saliency map is calculated over 100 images with resolution 1024x768. The columns under each metric shows the correlation score given by each metric (higher the score – better the correlation of the saliency model with ground truth fixation maps). The algorithm achieves 62% of complexity reduction compared to our previous model and has significantly lower complexity compared with closely performing state-of-the-art models.

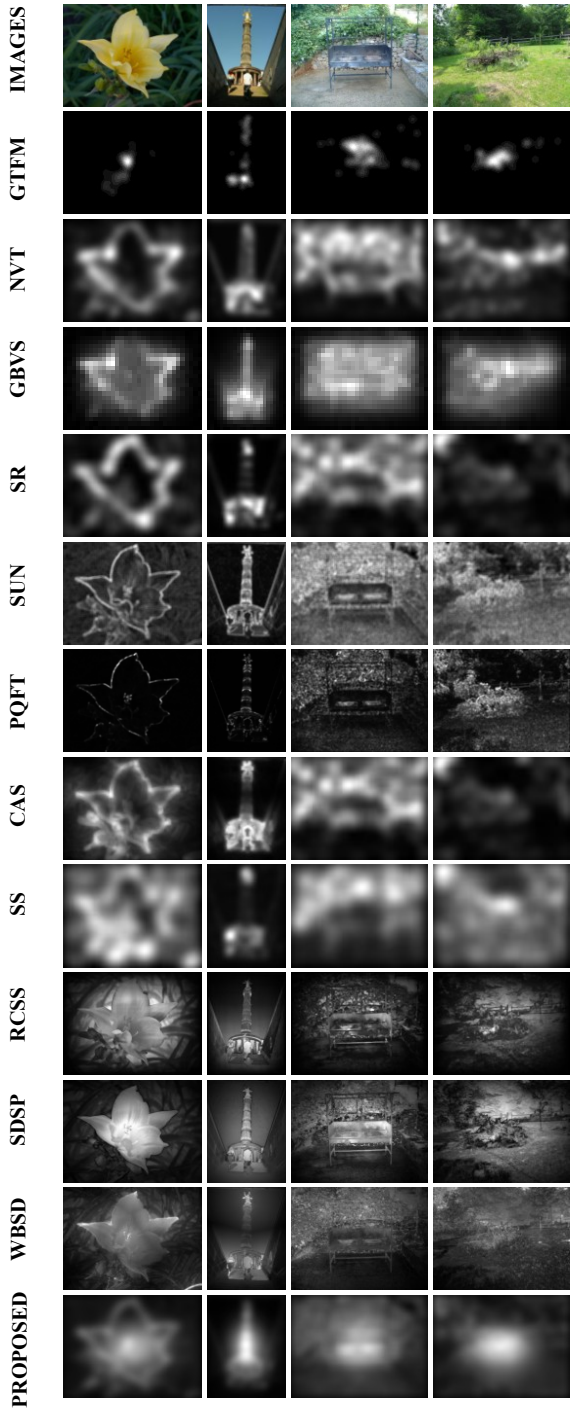


Fig.3. Qualitative comparison of four sample images from Judd's dataset

IV. Conclusion And Future work

A novel visual saliency detection algorithm based on in-focus regions and image center sensitivity is proposed. The in-focus regions are obtained using the characteristics of Integer Cosine Transform. The centre map is generated by placing a 2D anisotropic Gaussian blob at the centre of the image. The developed model is compared against ten state-of-the-art saliency models using three metrics on a publicly available dataset. The results show that our model out-performs the other models in-terms of saliency detection accuracy at a lower complexity than other closely performing models.

In the future, we plan to incorporate top-down features into our model and consider temporal characteristics of videos in order to improve the performance of the model in predicting fixations.

References

[1] S. Frintrop, "General object tracking with a component-based target descriptor," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, 2010, pp. 4531-4536.

[2] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *Image Processing, IEEE Transactions on,* vol. 13, pp. 1304-1318, 2004.

[3] M. Ajallooeian, A. Borji, B. N. Araabi, M. N. Ahmadabadi, and H. Moradi, "Fast hand gesture recognition based on saliency maps: An application to interactive robotic marionette playing," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, 2009, pp. 841-847.

[4] C. Siagian and L. Itti, "Biologically inspired mobile robot vision localization," *Robotics, IEEE Transactions on,* vol. 25, pp. 861-873, 2009.

[5] M. S. El-Nasr, A. Vasilakos, C. Rao, and J. Zupko, "Dynamic intelligent lighting for directing visual attention in interactive 3-d scenes," *Computational Intelligence and AI in Games, IEEE Transactions on,* vol. 1, pp. 145-153, 2009.

[6] A. Borji, D. N. Sihite, and L. Itti, "Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study," 2012.

[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 20, pp. 1254-1259, 1998.

[8] H. Xiaodi and Z. Liqing, "Saliency Detection: A Spectral Residual Approach," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, 2007, pp. 1-8.

[9] G. Chenlei and Z. Liming, "A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression," *Image Processing, IEEE Transactions on,* vol. 19, pp. 185-198, 2010.

[10] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," 2007.

[11] N. Imamoglu, L. Weisi, and F. Yuming, "A Saliency Detection Model Using Low-Level Features Based on Wavelet Transform," *Multimedia, IEEE Transactions on,* vol. 15, pp. 96-105, 2013.

[12] H. Xiaodi, J. Harel, and C. Koch, "Image Signature: Highlighting Sparse Salient Regions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 34, pp. 194-201, 2012.

[13] L. Zhang, Z. Gu, and H. Li, "sdsp: A Novel saliency detection method by combining simple priors," presented at the in Proc. ICIP, 2013.

[14] T. N. Vikram, M. Tscherepanow, and B. Wrede, "A saliency map based on sampling an image into random rectangular regions of interest," *Pattern Recognition,* vol. 45, pp. 3114-3124, 2012.

[15] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "SUN: A Bayesian framework for saliency using natural statistics," *Journal of Vision,* vol. 8, 2008.

[16] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 34, pp. 1915-1926, 2012.

[17] J. Chilukamari, S. Kannangara, and G. Maxwell, "A DCT based in-focus visual saliency detection algorithm," in *Consumer Electronics Berlin (ICCE-Berlin), 2013. ICCEBerlin 2013. IEEE Third International Conference on*, 2013, pp. 1-5.

[18] H. Woong and K. Chong-Min, "A Multitransform Architecture for H.264/AVC High-Profile Coders," *Multimedia, IEEE Transactions on,* vol. 12, pp. 157-167, 2010.

[19] P.-H. Tseng, R. Carmi, I. G. Cameron, D. P. Munoz, and L. Itti, "Quantifying center bias of observers in free viewing of dynamic natural scenes," *Journal of Vision,* vol. 9, 2009.

[20] B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist, "Visual correlates of fixation selection: Effects of scale and time," *Vision research,* vol. 45, pp. 643-659, 2005.

[21] D. J. Parkhurst and E. Niebur, "Scene content selected by active vision," *Spatial vision,* vol. 16, pp. 125-154, 2003.

[22] F. Vitu, Z. Kapoula, D. Lancelin, and F. Lavigne, "Eye movements in reading isolated words: Evidence for strong biases towards the center of the screen," *Vision research,* vol. 44, pp. 321-338, 2004.

[23] M. Dorr, T. Martinetz, K. R. Gegenfurtner, and E. Barth, "Variability of eye movements when viewing dynamic natural scenes," *Journal of Vision,* vol. 10, 2010.

[24] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Upper Saddle River: NJ: Prentice Hall., 2003.

[25] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 2106-2113.