# On Biologically-Motivated Saliency Detection

Milan A. Verma
Queen Mary, University of London
milan@dcs.qmul.ac.uk

*Abstract*— **Saliency detection has been a long-standing challenge in computer vision. Successfully detecting interesting regions of an image has a range of potential applications, such as detecting a military vehicle amongst bushes and trees, or moving the eyes of a house robot. In this paper we detail the progress of improvements to a widely-used saliency model, in order to control the amount of rendering to be done on an artistic depiction of a photograph. The intention is to render highly relevant regions with a fine brush stroke and not so relevant regions with a course brush stroke, thereby controlling the level of detail presented to the viewer.**

## I. Introduction

HUMAN visual systems have a remarkably precise and rapid ability to guide the viewer's attention when viewing a photograph to the most relevant areas. Indeed, although this behaviour to some extent is dependent on what is regarded as *relevant* to the viewer, there is a vast amount of research advocating the theory that there is some hard-wired primitive mechanism which performs this task at an early stage in the visual processing pipeline. During this stage of involuntary visual input processing, there are minimal if any influencing factors in deciding which part of the photograph the viewer's eyes will move to. With time, further processing of the visual input is performed using supplementary high-level information in order to guide the attention of the viewer through the photograph. Such high-level information is homogeneous in nature and since experience and perceptions vary from person to person, their influence in guiding the viewer's attention through the photograph can also vary from person to person. In contrast, since covert early vision operates more or less in the same manner for all humans, understanding how these mechanisms work help us to understand our human visual system. The reproduction of the early vision behavior using computers have a variety of uses such as security and surveillance [1], image compression [2], automatic target detection [3] and artistic rendering [4] to name but a few.

The objective of the work to be detailed herein is to improve a widely-used early vision model developed by Itti and Koch [12] for predicting gaze locations. The enhanced model will be validated using a proposed framework designed to compare saliency results with the Itti and Koch model [12]. Using our model, we aim to render a photograph using synthesized paint strokes of varying brush size.

## II. Saliency Detection

Saliency is a term used to describe an interesting region in an image. But how can we use computers to compute a measure of saliency? Finding measurements for detecting saliency in an image has been a long standing challenge in the computer vision community and has stimulated much research. Essentially, models for saliency detection can be categorized as either using an Engineering or a Biological approach. Biologically-oriented approaches [10] infer that tasks that we accomplish as humans have a direct genetic influence. In addition to this, complementary social and environmental influences aid the learning process and help us adapt to different situations. Biologically-oriented computation tends to draw influence from evolution and psychophysics thus forming a firm logic and theory for derived computations. Engineering approaches [11] on the other hand, use technology-based techniques for computation and problems are solved using practical solutions. Neither technique is correct, but a biological approach is more methodological in helping answer interesting evolutionary and psychophysical research questions. For this reason, the work described in this paper has a direct biological motivation.

Crucial to saliency detection is the evaluation of local image differences but equally well an image sequence showing a flickering object in a scene would highlight temporal or even spatiotemporal differences. When we view such salient occurrences with our eyes they attract attention due to a series of complex neural activity. This process involves the visual input being handled initially in the V1, the primary visual area, followed by processing in the V2, V4 and Inferior Temporal cortex (IT) [6]. The stream of visual processing involves beginning with analysis of very low-level information such as delineating colour, through to the recognition of objects. At the same time this visual pathway interacts with information from short term memory, which functions in the Lateral Prefrontal cortex [6].

To compute such methods of processing, bottom-up saliency detection deals with the cortical early visual processing, such as the processing of local image discontinuities to form saliency maps. The use of low-level visual cognition information for creating bottom-up saliency maps was first proposed by Koch and Ullman [7] in 1985. The idea was to use a set of features maps (such as those for detecting changes in pixel intensity, pixel contrast and detecting orientation) and to use these to derive a *master map*, which was later coined as the *saliency map* by Itti [12] in 1998. These basic ideas have been accepted as plausible explanations for aspects of visual cognition, thus forming the basis of subsequent computer vision research in this area. Consequently, the Neuromorphic Vision Toolkit (NVT) developed by Itti et al [12] is one of the most popular approaches for saliency in the spatial and temporal domains. It uses linear combinations of center-surround filters [9] to measure local contrasts in colour, intensity and orientation. This is a biologically plausible approach as it is possible that this is how saliency is computed across the visual field [19]. Their bottom-up approach models the human visual system when analysing a scene, including the modelling of saccades. Probabilistic approaches have also been formulated [13,14], attributing each feature with a probability of interest.

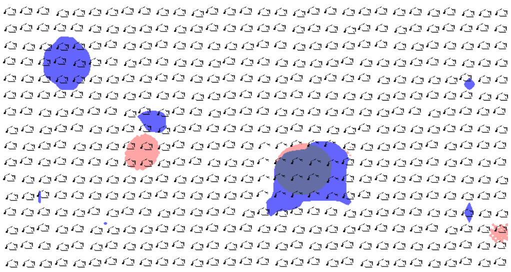
## III. PROPOSED IMPROVEMENTS

The Itti and Koch model [12] for saliency subsamples the original image to produce an image pyramid. A fixed sized Gabor filter is then convolved over the various scales in the pyramid to produce filter responses before center-surround differences are computed across scales. Although this approach efficiently produces results over multiple scales, the drawback of the approach is the artifacts that are produced due to the subsampling and the consequent resizing of the images in the image pyramid. As an improvement, we have shown that varying the filter size whilst maintaining the original image size without subsampling is a far more effective approach for producing multiscale results.

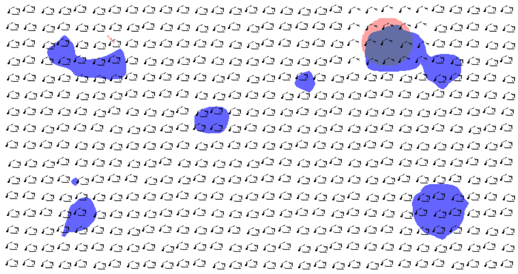
We have also shown that when combining the features maps across various scales and types, it is far more effective to use a weighted combination strategy rather than a linear one. To this end, we have successfully used the Hurst Exponent [15] for approximating the spread and dispersion of the filter responses in the feature maps. Using this approach we have been able to produce a weighting strategy for the combination of feature maps, ensuring that the highly localized filter responses are given greater importance.

## IV. EVALUATION TECHNIQUE

To evaluate the success of the improvements made to the Itti and Koch model [12] for saliency detection, we have proposed a biologically-inspired framework for comparing saliency results. Since our objective is to work with low-level stimuli which will have minimal context bias, our first challenge is to produce images with simple features. To achieve this, we have used Darwin's Biomorphs [16] as textons to form images with popout [18]. Figure 1 shows the results of the Itti and Koch model [12] overlayed with our results.



(a)



(b)

**Figure 1 (a) and (b) show the results of the Itti and Koch model (blue) [12] overlayed with the results of the improved model (pink). The Itti and Koch [12] model results show far more false-positives due to the filter responses being widely dispersed.**

Using a Genetic Algorithm [17] we then produced various generations of images, using the improved saliency algorithm as a fitness function. The results of the work conducted so far have enabled us to not only quantify saliency results but also to provide a framework with which results can be compared across saliency models.

## REFERENCES


[1] B. Abdulhai T. Rabie and A. Shalaby. Mobile active-vision traffic surveillance system for urban networks. In *Computer-Aided Civil and Infrastructure Engineering*, pages 20(4):231–241, Seoul, July 2005.

[2] F. W. M. Stentiford. An estimator for visual attention through competitive novelty with application to image compression. In *Picture Coding Symposium*, Seoul, 2001.

[3] A. Chianese G. Boccignone and A. Picariello. Using renyi's information and wavelets for target detection: An application to mammograms. In *Pattern Analysis and Application*, pages 3:303–313, 2000.

[4] A. Santella. The art of seeing: Visual perception in design and evaluation of non-photorealistic rendering. In *PhD thesis*, 2005.

[5] L. Itti and C. Koch. Computational modelling of visual attention. In *Nature Reviews Neuroscience*, pages 2(3):194–203, May 2001.

[6] G. Deco and E. T. Rolls. Attention, short-term memory, and action selection: A unifying theory. In *Progress in Neurobiology*, pages 76: 236–256, August 2005.

[7] C. Koch and S. Ullman. Shifts in selective visual attention: underlying neural circuitry. In *Human Neurobiology*, pages 4:219–227, 1985.

[8] L. Itti. Models of bottom-up and top-down visual attention. *PhD thesis*, *California Institute of Technology*, January 2000.

[9] J. Duncan. Integrated mechanisms of selective attention. In *Curr. Opin. Biol.*, pages 7:255–261, 1997.

[10] A.E. Eiben and J.E. Smith. *Introduction to Evolutionary Computing*. Springer; 1 edition, 2003.

[11] R. S. Pressman. *Software Engineering: A Practitioner's Approach*. McGraw-Hill Science/Engineering/Math; 5th edition, November 2001.

[12] C. Koch L. Itti and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 20(11):1254–1259, 1998.

[13] R. Rosenholtz. A simple saliency model predicts a number of motion popout phenomena. In *Vision Reseacrh*, pages 39:3157–3163, 1999.

[14] T. Kadir, A. Zisserman, and M. Brady, An affine invariant salient region detector. In *ECCV* pages 404-416, 2004.

[15] Hendrik J. Blok, On the Nature of the Stock Market: Simulations and Experiments. *Ph.D thesis, University of British Columbia*, 2000.

[16] Dawkins, Richard. The Blind Watchmaker. *W.W. Norton and Company, Inc*. 1987.

[17] M. A. Verma and P. W. McOwan. An adaptive methodology for synthesizing mobile phone games using genetic algorithms. In *Congress on Evolutionary Computation*, pages 864–871, Edinbourgh, Scotland, September 2005

[18] A. M. Treisman and G. Gelade. A feature-integration theory of attention. In *Cognitive Psychology*, pages 12(1):97–136, 1980.

[19] G. Rees Zhaoping L, L. Itti and J.K. Tsotsos. The primary visual cortex creates a bottom-up saliency map. In *Neurobiology of Attention*, pages 93:570–575, 2005.