OpenDS: An Open Source Implementation of Discriminant Saliency on Center-surround Hypothesis

Yen Nhi Vuong
EECS Department
York University
Toronto, Canada
ynhi97@my.yorku.ca

***Abstract*—Most existing saliency detection models tend to target the high contrast in colors, intensity and orientation channels. However, it doesn't accurately represent the visual focus of human. The saliency map most likely turns out to be a highlighted contour map of the image itself, rather than the actual focus point or region that people look at. In order to quantitatively predict human eye fixation on natural scenes, a research paper proposes saliency detection based on discriminant hypothesis. The discriminant saliency hypothesis is developed on top of the classical assumption that bottom-up saliency is a center-surround process. The methodology consists of two stages of detection, involving discriminant saliency detection, and leveraging natural image statistics to finally distinguish the target pop-out.**

***Index Terms*—saliency, bottom-up, itti-koch, discriminant hypothesis, center-surround map**

I. INTRODUCTION

In computer vision, saliency is defined as standout part(s) of the image with distinct properties that draw eye fixation, which is the attention of the viewers. There are primary 2 saliency cues: *bottom-up* - fast, stimulus-driven mechanism, and *top-down* - slower, goal-driven mechanism [1]. Bottom-up saliency is more common in development as people tend to do it naturally without following certain tasks. Bottom-up is also highly desired because it allows us to understand what typically catches our attention, what visual features falls into the retina and so on...

The original research paper is called *"On the plausibility of the* ***discriminant center-surround hypothesis*** *for visual saliency"* by Dashan Gao [2]. **Discriminant analysis** concerns with classifying set of observed data into predefined classes, giving minimum probability of expected error. The analysis requires a discriminant function. The **center-surround** hypothesis suggests bottom-up saliency as a center-surround process that derives optimal saliency architecture. In this architecture, saliency is detected from discriminating between the centre area of interest (*the center*) and the neighbourhood surrounding it (*the surround*): The larger the difference between stimuli located at the centre and its surrounding is, the more salient the location becomes. Optimal saliency detectors can serve psychophysics predictions of human saliency such as eye fixation prediction on natural scenes, background subtraction on high dynamic scenes and motion-based saliency in the presence of ego-motion.

The original research implemented a binary executable from MATLAB which is no longer working. This is an Engineering project which aims to reproduce the discriminant saliency model in Python 3 code with the help of open-source libraries such as OpenCV and Numpy.

II. METHODOLOGY AND IMPLEMENTATION

The general architecture of image processing mainly follows the Itti-Koch-Niebur (IKN) model as shown in Figure 6 [3]. Note that this is a slightly simpler and modified version compared to original Itti model, the differences will be discussed within this section.

For static imagery, the process consists of 2 main stages. First stage includes feature decomposition that results in intensity map I and four color channels (*R, G, B, and Y*). The feature maps are then convolved with three *Mexican hat wavelet filters*. Mexican filters are centered at spatial frequencies 0.02, 0.04, and 0.08 cycles/pixel, to generate nine feature channels. Mexican hat wavelet is a linear difference filter for local area and feature point detection. In additional, the intensity map is convolved with (zero-mean) *Gabor filter*. Gabor filter is a linear filter used for texture analysis; promotes the orientation/direction recognition. Gabor kernels filter at 3 spatial scales (centered at frequencies of 0.08, 0.16, and 0.32 cycles/pixel) and 4 directions (evenly spread from 0 to π). Due to challenges in implementation, some math and number notations don't match up with research description.

Next, in order to apply center-surround classification, the feature maps are worked into multi levels of Gaussian pyramid (implicitly called through OpenCV `pyrDown` function).

The second stage aims to leverage image statistics by estimating the marginal mutual information in Figure 2. In short, the saliency of location l is equated to the power of $\mathbf{X}$ to discriminate between the center and surround of l

A project proposal based on "On the plausibility of the discriminant center-surround hypothesis for visual saliency" by Dashan Gao; Vijay Mahadevan and Nuno Vasconcelos in 2008.

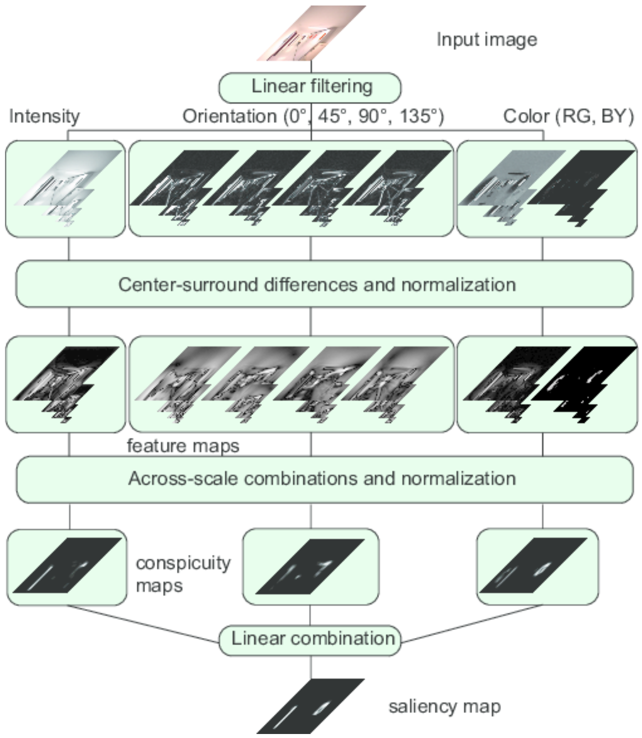

Fig. 1. Itti-Koch model with center surround map computation

based on the distributions of the feature responses estimated from the two regions. However, this approach is deemed as impractical due to extensive computation on high-dimensional feature space. Alternative solutions are to combine with Gabor or wavelet (Laplace or Mexican wavelet) coefficients to reduce complexity of high dimensional image. In other hand, it's possible to derive into new formula using Kullback–Leibler (KL) in Figure 3.

$$
\begin{aligned}
S(l) &= I_l(\mathbf{X};Y) \\
&= \sum_c \int p_{\mathbf{X}(l),Y(l)}(\mathbf{x},c)\log\frac{p_{\mathbf{X}(l),Y(l)}(\mathbf{x},c)}{p_{\mathbf{X}(l)}(\mathbf{x})p_{Y(l)}(c)}\mathrm{d}\mathbf{x}.
\end{aligned}
$$

Fig. 2. Formula for Saliency computation using Mutual Information Theory

$$
I_m(X;Y)_{(l)} = \sum_c p_Y(c)KL(p_{X/Y}(x_d/c);p_X(x_d))_{(l)}
$$

Fig. 3. Formula for Saliency computation using Kullback–Leibler divergence (KL)

III. Test Data

In order to test the model, image data was retrieved at the Toronto Dataset [4]. The images were collected under the research of Neil Bruce, John K. Tsotsos. *Attention based on information maximization*. The dataset includes 120 color images of outdoor and indoor scenes. Each image is about size: 681 x 511px which helps minimize the processing time compared to an original photograph. The observers were 20 undergrads and grads student. The eye tracker device was ERICA workstation including a Hitachi CCD camera with an IR emitting LED.

In addition, large portion of images do not contain particular regions of interest which serve the purpose of testing the natural human fixation without any predictable features.

Some other test images were collected directly from the research pdf. These images are not natural but rather synthetic data: solid, uniform background with consistent format objects (size, shape, color) on top. This data are more predictable.

IV. Results and Analysis

The project final step would be to apply the implemented model to a set of input images, which is preferably close to the research's original test set. Previously, the researchers relied on classical experiments in visual attention which were both qualitative and quantitative. For simple target pop-out test, some example images were provided directly in the research paper. And to generate progressive/dependency graph between saliency versus image feature, more static and motion data will be needed.

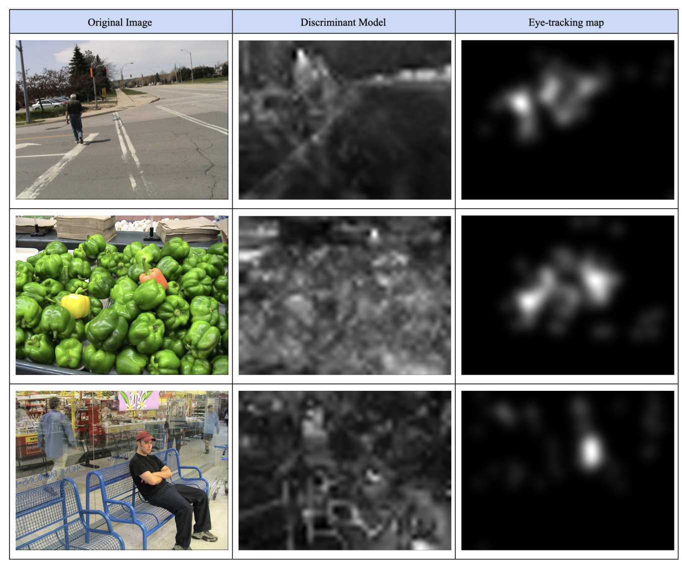

Fig. 4. Sample output table, from left to right: Original image, Discriminant map output and Actual eye-tracking data.)

Saliency model	Discriminant	Itti and Koch (2000)	Bruce and Tsotsos (2006)	Inter-subject
ROC area	0.7694	0.7287	0.7547	0.8766

Fig. 5. ROC areas for different saliency models with respect to all human fixations.)

References

[1] D. Gao and N. Vasconcelos, "Bottom-up saliency is a discriminant process," *2007 IEEE 11th International Conference on Computer Vision*, 2007. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4408851

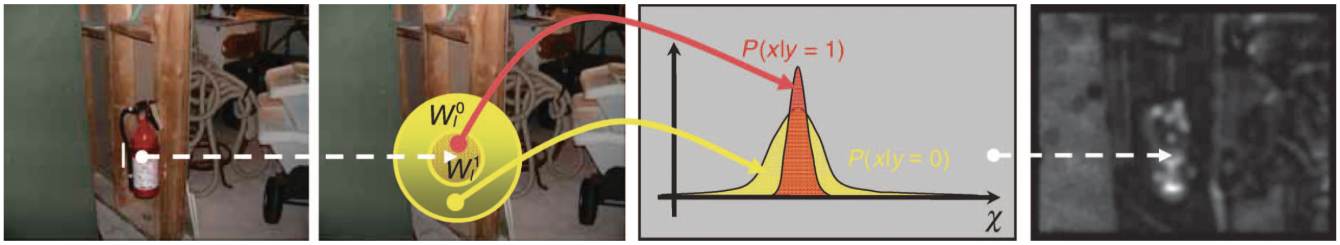

Fig. 6. Illustration of discriminant center-surround saliency

[2] D. Gao, V. Mahadevan, and N. Vasconcelos, "On the plausibility of the discriminant center-surround hypothesis for visual saliency," *Journal of Vision*, vol. 8, no. 7, p. 13, Jun 2008. [Online]. Available: https://jov.arvojournals.org/article.aspx?articleid=2193585

[3] T. Xu, Q. Mühlbauer, and S. Sosnowski, "Looking at the surprise: Bottom-up attentional control of an active camera system," Nov 2008. [Online]. Available: https://www.researchgate.net/publication/221144241_Looking_at_the_Surprise_BottomUp_Attentional_Control_of_an_Active_Camera_System

[4] 2012. [Online]. Available: http://saliency.mit.edu/results$_mit300.html$