A bioinspired model of early visual processing with hue-feature saliency for a cognitive architecture

Cynthia Avila-Contreras, Daniel Madrigal,
Félix Ramos, and Juan Luis del Valle-Padilla

CINVESTAV del IPN, Guadalajara, Jalisco, México
cavila, dmadrigal, framos@gdl.cinvestav.mx

Abstract

We present a computational model that describes the early stages of visual processing and the within selective attention mechanisms to generate feature-based (hue or color) activations of salient localizations based on neurophysiology evidence of selective responses in the visual pathway. The model identifies related brain areas, the feasible computations of each one, and proposes the type of data generated and shared among the components. This work is part of the selective aspect of an attention system designed for a broader cognitive architecture for virtual creatures.

Keywords: visual processing, hue feature salience, lateral inhibition, double opponent

1 Introduction

With vision we could perceive a great diversity of features (color, movement, distance, etc.) from our environment. A huge quantity of neurons is needed to acquire and process the visible light, only in the retina exist millions of photoreceptors. However, not all the information is completely processed: we can only notice the prominent parts of a scene, what we *pay attention* to them in a selective manner.

Selective attention involves allocation of processing resources to the most relevant or salient sensory events, filtering out the unimportant data. Concurrent mechanisms in the brain may help to determinate the focus of attention. They could be *bottom-up* inhibition of neural activity based on purely physical peculiarities (stimulus-driven), which makes it salient; *top-down* enhancement of neural activity based on the congruence of the stimulus with current thoughts and task-related information (goal-driven), which makes it relevant; and finally *motivational or emotional values* related to the stimuli (value-driven) could make an stimulus more salient. There are three major paradigms for focusing attention: attention to specific visual properties such color or motion (feature-based), attention to spatial regions like left or right (space-based), and attention to the segmentation which compounds a perceptible object (object-based). It is possible that the final deployment of attention is achieved by integrations among all of them during different stages of visual processing.

Biological evidence along with theoretical and computational models agreed that early stages of visual processing (from the entry of information in the retina to the first layers in the occipital cortex) works in a bottom-up manner to extract salient features, enhanced by feature and space based top-down signals.

In visual processing models, there are works like [16, 6], which focus on modeling energy transduction as realistically as possible for the preservation of chromatic luminosity, but appear to be isolated from the attentive mechanisms. Related with salient areas detection, Itti & Koch model [7] is a relevant reference to generate bottom-up saliency map using a low level features (color, intensity and orientation), although it leaves aside the top-down influences on space and characteristics. Some approaches use and/or extend that model to include or improve other untreated aspects of attention[3]. Even so, it is remarkable that it is more computational-oriented than bio-inspired modeling.

In our perspective, it is needed a model that gathers more biological plausible operations in the visual processing pathway in order to achieve realistic results. The forementión models set the stage for us to glimpse the potential of a proposal of this characteristics.

Our work focuses on the creation of a cognitive architecture that has several cognitive functions, such as attention, sensory processing, perception, decision making, planning, etc. Therefore, we believe that modeling the first stage of sensory processing with attention must be sufficiently flexible to continue growing. Our approach includes the identification of independent modular components that interact with each other (and not a global neural network) to achieve a function. In a previous paper, we present a model that describes the interaction between visual processing, attention and novelty detection[2], but we did not include detailed information about in the possible computation of each component.

2 Model of early visual processing with attention

In the beginning stages of visual processing, original inputs (stimulus intensity sensed by the receptors) are acquired and transformed into an internal representation. Afterwards, during later phases, the internal representation is treated to extract the features that are used to describe the environmental objects. The visual pathway covered by the model is the blue area depicted in figure 1. The visual processing starts with activation of receptors in retina, then information pass through thalamus and ascends to visual areas in occipital cortex. Besides this feedforward connections, there are feedback projections which may help to modulate the activity across areas.

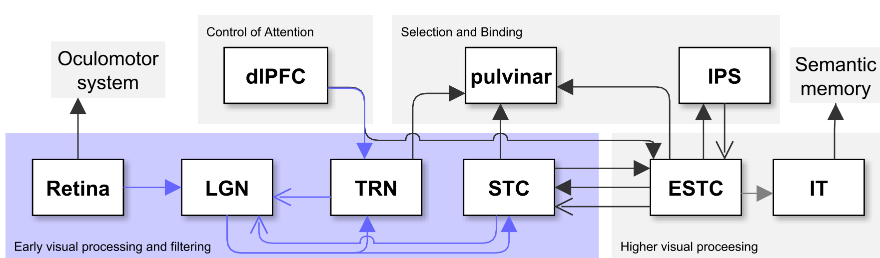

Figure 1: Bla

Next, we will describe the computation done in each brain area and its correspondence in our model, which is summarized in 2 using a representation similar to activity diagrams defined in UML 2.0.

2.1 Initial transductions

Retina (RET) is the layer sensitive to the light due to photoreceptors that are located there. There are two types of photoreceptors: cones and rods. The cones are excited by bright light and mediate the color vision [12]. On the other hand, rods can not handle the bright light, usually at daylight, the activation levels of those receptors are saturated [17]. The photoreceptors are physically distributed over the retina in a mosaic way [1].

In this work, the RET responses are modeled as a transformation of an image in RGB color space to LMS values, representing long, medium and short cone responses. It is calculated using the Hunt-Pointer-Estevez (HPE) model, which use as input an image codified in XYZ format. The following equation describes the HPE model: $\begin{bmatrix} l & m & s \end{bmatrix}^T = [\mathbf{M}_{HPE}][\mathbf{M}_{XYZ}] \begin{bmatrix} R & G & B \end{bmatrix}^T$. Where l, m, and s denote the output cone responses, R, G, and B is a pixel codified in RGB color space, $\mathbf{M}_{HPE}$ and $\mathbf{M}_{XYZ}$ are transformation matrices; the first one is used to calculate the activation of the cones and the second one is used to transform from RGB to XYZ. The values of the transformation matrices were taken from [5]. The equation is applied to each pixel of the input image, and the output are the three matrices $\mathbf{l}$, $\mathbf{m}$, and $\mathbf{s}$.

The primary input to **lateral geniculate nucleus (LGN)** of the thalamus is sensory information generated in the retina. One of their functions is to connect optical nerve with visual processing areas (occipital lobe) [11]. LGN performs a color opponent process, which can be seen as color space conversion from LMS (cones color space) *to DKL* (color difference space). This task is performed by:

$SO_D(x,y) = \sum_{p=-N}^{N} \sum_{q=-N}^{N} [\mathbf{l}(x+p,y+q)g(p,q|\sigma_l) - \mathbf{m}(x+p,y+q)g(p,q|\sigma_m)]$

$SO_K(x,y) = \sum_{p=-N}^{N} \sum_{q=-N}^{N} [\mathbf{s}(x+p,y+q)g(p,q|\sigma_s) - \frac{\mathbf{l}(x+p,y+q)+\mathbf{m}(x+p,y+q)}{2} g(p,q|\sigma_y)]$

$SO_L(x,y) = \frac{\mathbf{l}(x,y)+\mathbf{m}(x,y)}{2}$

Where g denotes a Gaussian function, and the sigma values (σ_l, σ_m, σ_s, and σ_y) are used to control his amplitude. The parameter N determine the receptive field size.

During visuo-spatial task, activations in LGN are modulated (incremented) to favor the cued region (space-based attention) [8].Also, neurons in the **thalamus reticular nuclei (TRN)** are highly responsive to shifts in spatial attention, and shows the opposite pattern of activation than LGN: directing spatial attention to a stimulus decreases the firing rate of correspondent TRN neurons. In our model, TRN responses are dependent on LGN excitatory input and a relevant region ρ_r received from the attentional set AS in dorsolateral prefrontal crotex (DLPFC). The function of *Relevant-region inhibition* performed in TRN is given by $T(x,y) = SO(x,y) \cdot (1 - (\tau \cdot R(x,y)))$, where $\tau = .04 \pm .005$[9] is the maximum decrement when location is fully relevant, and R is a matrix of the spatial relevance calculated after ρ_r. The update of LGN is $SO(x,y) = SO(x,y) \cdot (1 + \lambda \cdot R(x,y))$ but only when $T(x,y) < SO(x,y)$, where $\lambda = .09 \pm .005$[9] is the maximum increment in LGN due inhibitory input from TRN.

2.1.1 Feature extraction

The next step in visual processing is achieved in **striate cortex (STC)**, located in the occipital lobe. This cortex is also known as "primary visual cortex" or V1. Its neurons are tuned only

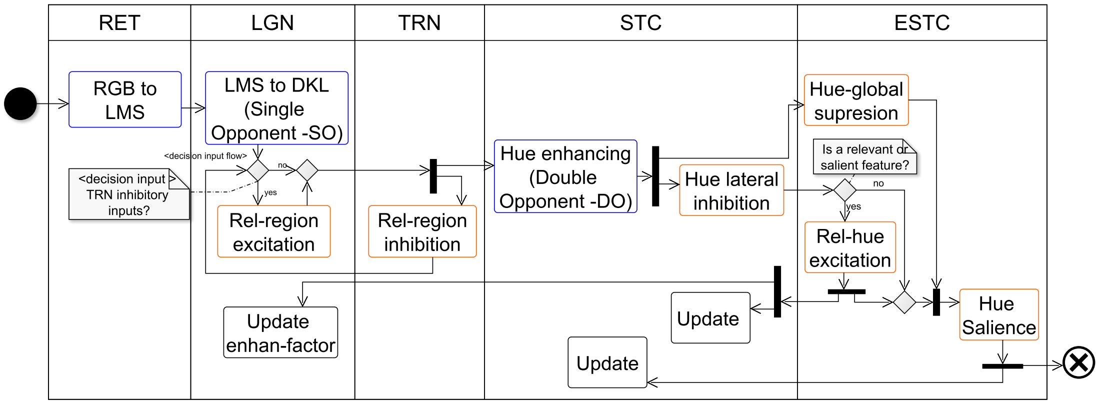

Figure 2: The model represented in a activity diagrama. Vertical divisions group operations executed by each component. Blue boxes are just visual processing operations, while orange boxes are attention-related operations.

to primitive features rather than complex objects, and they respond to stimulus features that are even invisible to awareness: it responds to orientations, spatial frequencies, and colors [13]. For this reason, V1 can be seen as a biological system that detects textures and edges. On the other hand, there is not a lot of evidence about the color processing performed in this area. However, some theories suggest a color-tuning process based on the color comparison between adjacent parts of the visual scene. The first step performed in our model is a process of *Hue enhancing*; biologically this process is carried out by double-opponent cells. The computation of this cells are described by the following equations: $DO_D(x,y) = H_R(x,y) - H_G(x,y)$ and $DO_K(x,y) = H_B(x,y) - H_Y(x,y)$, here DO denotes the activation of a Double Opponent Cell and

$H_R(x,y) = \sum_{p=-N}^{N} \sum_{q=-N}^{N} [u(SO_D(x+p,y+q))g(p,q|\sigma_{r1}) - u(SO_D(x+p,y+q))g(p,q|\sigma_{r2})]$
$H_G(x,y) = \sum_{p=-N}^{N} \sum_{q=-N}^{N} [u(-SO_D(x+p,y+q))g(p,q|\sigma_{g1}) - u(-SO_D(x+p,y+q))g(p,q|\sigma_{g2})]$
$H_B(x,y) = \sum_{p=-N}^{N} \sum_{q=-N}^{N} [u(SO_K(x+p,y+q))g(p,q|\sigma_{b1}) - u(SO_K(x+p,y+q))g(p,q|\sigma_{b2})]$
$H_Y(x,y) = \sum_{p=-N}^{N} \sum_{q=-N}^{N} [u(-SO_K(x+p,y+q))g(p,q|\sigma_{y1}) - u(-SO_K(x+p,y+q))g(p,q|\sigma_{y2})]$
being u a threshold to zero function, g indicates a Gaussian function, and the sigma values are used to control his amplitude. The parameter N determine the receptive field size.

Also, in STC exists neuronal modulation due to features contrast [10]. It is suggested that maps representing saliency for basic features (i.e. orientation, direction, color or spatial frequency) [15] are created in parallel in this area by bottom-up mechanisms [14]. This process could be achieved by inhibitions between V1 neurons via lateral connections, leading to mutual suppression between neurons tuned to similar input features (iso-feature suppressions)[18]. Based on that, our *Hue lateral inhibition* function acts over the individual maps of represented in the double opponent responses. To do it, we first need to separate DO_D in red and green maps, following that if $DO_D(x,y) < 0$ then $DO_D(x,y) \in G$, $DO_D(x,y) \in R$otherwise. The equivalent operation is done for yellow and and blue maps. To obtain the inhibited activation in R', we have $R'(x,y) = R(x,y) - \sum_{p=-N}^{N} - \sum_{q=-N}^{N} \cdot \Delta_1(R(x,y), R(x-p,y-q), d_R) g(p,q)$, and Δ_1 is a function for the similarity between two intensities of one color and d_R is the maximum difference for red colors, such that $\Delta_1(i_1,i_2) = d - (i_1 - i_2)^2$ (only when $(i_1 - i_2)^2 \geq d$ and 0 otherwise). All the other maps (G', B' and Y') are calculated using its equivalent functions.

The later stage of visual processing are executed in **extrastriate cortex (ESTC)**. It comprises several visual structures located at the front of V1, as V2, V4 and MT. It is believed that ESTC participate in more complex visual processes, as color perception. There is evidence of increase in activations due top-down signals of relevant color, so the proposed function *relevant-hue excitation* evaluates if the individual hue lateral-inhibited map received is currently relevant (via signals of $\rho_h \in AC$ from DLPFC) and updates its values.

For hue-feature salience detection, the hue lateral inhibitions proposed in STC only affects local intensity, this means just close surrounding areas.A function to determinate the global hue activations and to suppress them according to its repetitiveness could be performed in ESTC, because its receptive fields are bigger than the ones in STC. Also, ESTC activations related to color are more frequent, so we can relay in a two dimensional color representation in this area using the double opponent responses in STC. The proposed is a *Global-hue activations* operation $G(x, y) = (C(DO_D, k(x, y))/\gamma)) \cdot \Delta_2(DO_{D,K}(x, y), CA_{D,K}(DO_D(x, y)))$, where $DO_{D,K}(x, y)$ is the two axis (D and K) color in (x, y), C is the global color count of all colors similar to $DO_{D,K}(x, y)$, CA_D, K is the global average of the all colors similar to $DO_D, K(x, y)$ and $Delta_2(c1_{D,K}, c2_{D,K}) = \sqrt{(c1_D - c2_D)^2 + c1_K - c2_K)^2}$ is the difference between the two colors.

Finally, hue-salience is the combination of the is the linear addition of each of the individually modulated hue maps $S_h = R' + G' + B' + Y'$ and global activations in G, in the way of $S_H = 1 - (1 - G)/S_h$, which decreases the global activations map G to reflect the salience of lateral inhibitions S_h by increasing the contrast between the two maps.

3 Implementation and Experiments

The proposed model was implemented in Java with the use of the OpenCV library. The application use as input any JPEG image and give as an output a set of matrices and a HUE saliency image. The output of each component in the model are 2D matrices that are used to the next operations.

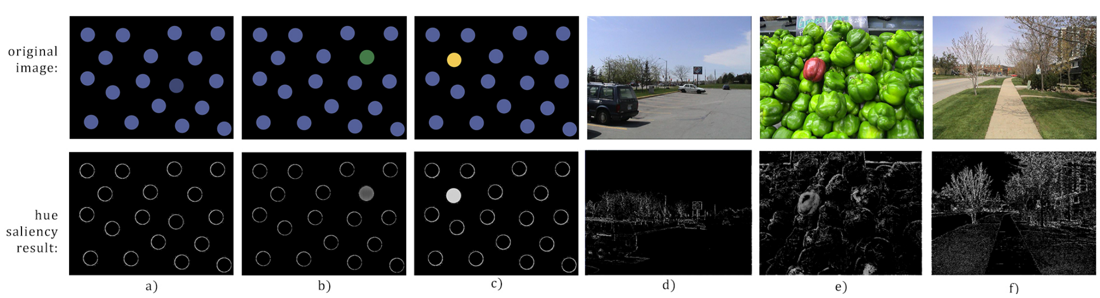

Figure 3: The HUE saliency images are grayscale images, where the whitest pixels indicates that this point is more different in color.

To exemplify the feasibility of the proposed model we show the results of six images. In figure 3a), 3b) and 3c) we show the results of our experiments over synthetic images that resembles psychophysical test of salient pop-out effects. We can see the differences between a barely salient color region in 3a) (fairly pale blue circle among darker circles) against the highly salient color in 3b) (bright yellow circle among blue circles). The effect in 3b) is also as expected, due to perceptual similarity between dark green and dark blue.

Despite we are developing this system to be a part of a cognitive architecture for artificial creatures immerse in a virtual 3D environments, we also wanted to experiment with images of natural scenes (which are different in hue distyribution than synthetic, more contrasting images in virtual simulations) taken from the Toronto data set[4], that are used in benchmarks of saliency detection. The original image in figure 3e) has a good contrast between color dark red and bright green (with some areas of green being more yellowish green), our results shows more salience in the region of the red pepper, but also in some other regions where yellow light and bluish shadows are concentrated. For figure 3d) and 3e), there is no a easily pop-out hue-salient region, and our results depicts that larger areas of the same hue (like sky and pathways) where diminished while some objects where salient (mainly trees).

Its worth to mention that salient areas where not normalized to expose just a certain percent of the maximums values, and that adjustment could lead to more segregated patches of salient hues.

4 Summary and Discussion

In this paper we have proposed a model of visual processing with attention-related mechanisms to obtain basic salience, specifically with bottom-up feature inhibition due contrast of hues (color) and top-down signals of relevant-hue and relevant-region to space and feature gain. The computations and connectivity of its components are based on findings in neurophysiology of visual hierarchy, starting with photo-receptors in retina and reaching to a representation of feature-dimension salience in extrastriate-cortex.

The particular set of operations described is simple, despites its computational complexity. So far, that is not a limitation because this work is aimed to be part of a cognitive architecture for human-like virtual creatures, in which other cognitive functions interact and share information among the models. Because of that, it is mainly designed to be implemented as a distributed system, with the operations separated in components and its outputs can be received or requested by other components in any given stage and moment. This is perhaps the major differentiation of our approach from other computational, biological or even cognitive models.

Our experiments shows that the model performs as expected comparing to other important models of the state of the art. Besides, it reproduces some psychophysical results of contrasting colors, as well as natural scenes from Tsotsos data set [4]. Results are acceptable as first part of the visual process. Along the different components, there are several parameters that may be adjusted to improve the salient area, depending on characteristics of input image. Future work is needed to do it automatically via feedback connections.

Finally, we believe in the flexibility of this model to be extended with more features (orientation, movement) using similar principles. Furthermore, this model compounds the pre-attentive (or automatic) stage of visual attention, and can be upgraded including more selective aspects, as final selection, binding of features and orientation, as well as attentional control relying in higher brain areas.

Acknowledgements

The authors would like to thanks to CONACyT for the financial support.

References

[1] P. K. Ahnelt. The photoreceptor mosaic. *Eye*, 12(3b):531–540, May 1998.

[2] Cynthia Avila-Contreras, Ory Medina, Karina Jaime, and Flix Ramos. An agent cognitive model for visual attention and response to novelty. In *Agent and Multi-Agent Systems: Technologies and Applications*, pages 27–36. Springer International Publishing, 2014.

[3] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(1):185–207, Jan 2013.

[4] Neil Bruce and John Tsotsos. Attention based on information maximization. *Journal of Vision*, 7(9):950–950, 2007.

[5] M.D. Fairchild. *Color Appearance Models*. The Wiley-IS&T Series in Imaging Science and Technology. Wiley, 2005.

[6] S. Gao, K. Yang, C. Li, and Y. Li. A color constancy model with double-opponency mechanisms. In *2013 IEEE International Conference on Computer Vision*, pages 929–936, Dec 2013.

[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*, 20(11):1254–1259, 1998.

[8] J Patrick Mayo. Intrathalamic mechanisms of visual attention. *J. Neurophysiol.*, 101(3):1123–1125, 2009.

[9] Kerry McAlonan, James Cavanaugh, and Robert H Wurtz. Guarding the gateway to cortex with attention in visual thalamus. *Nature*, 456(7220):391–394, 2008.

[10] Hans-Christoph Nothdurft. *Neurobiology of Attention*, chapter 38. Salience of Feature Contrast. Elsevier, 2005.

[11] Andrew Preston and Allison S. Evans. *Encyclopedia of Clinical Neuropsychology*, chapter Lateral Geniculate Nucleus of Thalamus, pages 1435–1436. Springer New York, New York, NY, 2011.

[12] R. C. Reid and W. M.Usrey. *Fundamental Neuroscience*, chapter Chapter 27 Vision, pages 637 – 659. ElSevier, 2008.

[13] Robert Shapley and Michael J. Hawken. Color in the cortex: single- and double-opponent cells. *Vision Res.*, 51(7):701 – 717, 2011. Vision Research 50th Anniversary Issue: Part 1.

[14] Stewart Shipp. The brain circuitry of attention. *Trends in cognitive sciences*, 8(5):223–230, 2004.

[15] Rufin VanRullen. Visual saliency and spike timing in the ventral visual pathway. *Journal of Physiology-Paris*, 97(23):365 – 377, 2003. ¡ce:title¿Neurogeometry and visual perception¡/ce:title¿.

[16] Kaifu Yang, Shaobing Gao, Chaoyi Li, and Yongjie Li. Efficient color boundary detection with color-opponent mechanisms. In *CVPR*, pages 2810–2817. IEEE Computer Society, 2013.

[17] Andrew J. Zele and Dingcai Cao. Vision under mesopic and scotopic illumination. *Front. Psychol.*, 5:1594, December 2014.

[18] Xilin Zhang, Li Zhaoping, Tiangang Zhou, and Fang Fang. Neural activities in v1 create a bottom-up saliency map. *Neuron*, 73(1):183–192, 2012.