Understanding Eye Movements: Psychophysics and a Model of Primary Visual Cortex

David Berga

Department of Computer Science, Universitat Autònoma de Barcelona, Bellaterra, Spain
Computer Vision Center, Edifici O Campus UAB, Bellaterra, Spain

26 July 2019

Abstract

Humans move their eyes in order to learn visual representations of the world. These eye movements depend on distinct factors, either by the scene that we perceive or by our own decisions. To select what is relevant to attend is part of our survival mechanisms and the way we build reality, as we constantly react both consciously and unconsciously to all the stimuli that is projected into our eyes. In this thesis [1] we try to explain (1) how we move our eyes, (2) how to build machines that understand visual information and deploy eye movements, and (3) how to make these machines understand tasks in order to decide for eye movements.

(1) We provided the analysis of eye movement behavior elicited by low-level feature distinctiveness with a dataset of 230 synthetically-generated image patterns [2]. A total of 15 types of stimuli has been generated (e.g. orientation, brightness, color, size, etc.), with 7 feature contrasts for each feature category. Eye-tracking data was collected from 34 participants during the viewing of the dataset, using Free-Viewing and Visual Search task instructions. Results showed that saliency is predominantly and distinctively influenced by: 1. feature type, 2. feature contrast, 3. temporality of fixations, 4. task difficulty and 5. center bias. From such dataset (SID4VAM), we have computed a benchmark of saliency models by testing performance using psychophysical patterns [3]. Model performance has been evaluated considering model inspiration and consistency with human psychophysics. Our study reveals that state-of-the-art Deep Learning saliency models do not perform well with synthetic pattern images, instead, models with Spectral/Fourier inspiration outperform others in saliency metrics and are more consistent with human psychophysical experimentation.

(2) Computations in the primary visual cortex (area V1 or striate cortex) have long been hypothesized to be responsible, among several visual processing mechanisms, of bottom-up visual attention (also named saliency). In order to validate this hypothesis, images from eye tracking datasets have been processed with a biologically plausible model of V1 (named Neurodynamic Saliency Wavelet Model or NSWAM)[4]. Following Li's neurodynamic model, we define V1's lateral connections with a network of firing rate neurons, sensitive to visual features such as brightness, color, orientation and scale. Early subcortical processes (i.e. retinal and thalamic) are functionally simulated. The resulting saliency maps are generated from the model output, representing the neuronal activity of V1 projections towards brain areas involved in eye movement control. We want to pinpoint that our unified computational architecture is able to reproduce several visual processes (i.e. brightness, chromatic induction and visual discomfort) without applying any type of training or optimization and keeping the same parametrization. The model has been extended (NSWAM-CM)[5] with an implementation of the cortical magnification function to define the retinotopical projections

---

towards V1, processing neuronal activity for each distinct view during scene observation. Novel computational definitions of top-down inhibition (in terms of inhibition of return and selection mechanisms), are also proposed to predict attention in Free-Viewing and Visual Search conditions. Results show that our model outpeforms other biologically-inpired models of saliency prediction as well as to predict visual saccade sequences, specifically for nature and synthetic images. We also show how temporal and spatial characteristics of inhibition of return can improve prediction of saccades, as well as how distinct search strategies (in terms of feature-selective or category-specific inhibition) predict attention at distinct image contexts.

(3) Although previous scanpath models have been able to efficiently predict saccades during Free-Viewing, it is well known that stimulus and task instructions can strongly affect eye movement patterns. In particular, task priming has been shown to be crucial to the deployment of eye movements, involving interactions between brain areas related to goal-directed behavior, working and long-term memory in combination with stimulus-driven eye movement neuronal correlates. In our latest study [6] we proposed an extension of the Selective Tuning Attentive Reference Fixation Controller Model based on task demands (STAR-FCT), describing novel computational definitions of Long-Term Memory, Visual Task Executive and Task Working Memory. With these modules we are able to use textual instructions in order to guide the model to attend to specific categories of objects and/or places in the scene. We have designed our memory model by processing a visual hierarchy of low- and high-level features. The relationship between the executive task instructions and the memory representations has been specified using a tree of semantic similarities between the learned features and the object category labels. Results reveal that by using this model, the resulting object localization maps and predicted saccades have a higher probability to fall inside the salient regions depending on the distinct task instructions compared to saliency.

Key Words: saliency, eye movements, attention, visual cortex, horizontal connections, visual search, free-viewing, psychophysics, firing rate, neural networks.

References

[1] David Berga. *Understanding Eye Movements: Psychophysics and a Model of Primary Visual Cortex*. PhD thesis, Universitat Autonoma de Barcelona, Campus UAB, 08193 Bellaterra, Barcelona, July 2019.

[2] David Berga, Xose R. Fdez-Vidal, Xavier Otazu, Victor Leboran, and Xose M. Pardo. Psychophysical evaluation of individual low-level feature influences on visual attention. *Vision Research*, 154:60 – 79, 2019.

[3] David Berga, Xose R. Fdez-Vidal, Xavier Otazu, and Xose M. Pardo. Sid4vam: A benchmark dataset with synthetic images for visual attention modeling. In *2019 IEEE International Conference on Computer Vision*. IEEE, oct 2019.

[4] David Berga and Xavier Otazu. A neurodynamic model of saliency prediction in v1. *arXiv preprint arXiv:1811.06308*, 2018.

[5] David Berga and Xavier Otazu. Modeling bottom-up and top-down attention with a neurodynamic model of v1. *arXiv preprint arXiv:1904.02741*, 2019.

[6] David Berga, Calden Wloka, and John K Tsotsos. Modeling task influences for saccade sequence and visual relevance prediction. *Journal of Vision*, 19(10):106c–106c, 2019.