Can Neuromorphic Computer Vision Inform Vision Science? Disparity Estimation as a Case Study

Guido Maiello[1,2,3], Manuela Chessa[2], Peter J Bex[3], Fabio Solari[2]

[1]UCL Institute of Ophthalmology, University College London, London, UK

[2]Dept. of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Genoa, Italy

[3]Dept. of Psychology, Northeastern University, Boston, MA, USA

The primate visual system efficiently and effectively solves a multitude of tasks from orientation discrimination to motion detection. Computer vision approaches to the same tasks often outperform biological visual systems. However, computer vision algorithms designed for a specific task rarely generalize to other tasks. Thus, integrating multiple computer vision algorithms to perform multiple tasks is a bulky and inelegant solution. The Computer Vision community is therefore beginning to implement algorithms that mimic the processing hierarchies present in the primate visual system in the hope of achieving more flexible and robust artificial vision systems [Kruger et al., 2013]. Here, we reappropriate the neuroscience "borrowed" by the Computer Vision community and ask whether neuromorphic computer vision solutions may give us insight into the functioning of the primate visual system. Specifically, we implement a neuromorphic algorithm for disparity estimation and compare its performance against that of human observers.

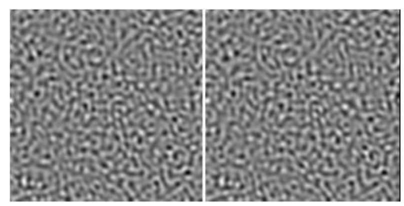

Figure 1: Example stimuli. Cross-fuse to view disparity corrugation.

The proposed algorithm mimics the processing stages occurring in primary visual cortex that feed into the ventral and dorsal pathways of the visual system. The input images to the two eyes are linearly filtered by a population of oriented Gabor filters, which approximate the simple cells of area V1 under the phase-shift model of disparity tuning [Fleet et al. 1996]. The squared responses of quadrature pairs of binocular simple cells approximate complex cells following the binocular energy model [Ohzawa et al., 1990]. The invariance of the responses of the complex cells with respect to the contrast of the input images is obtained via divisive normalization [Heeger, 1992]. Finally, a center of mass decoding strategy provides both the magnitude and the direction of disparity at each image location. Pyramidal decomposition [Burt & Adelson, 1983] is employed to mimic how the primate visual system processes multiple spatial scales. The algorithm can further mimic a foveated observer when the input images are log-polar transformed into a cortical-like representation of the visual signal to the retina [Solari et al., 2012].

We test the algorithm on the same stimuli (Fig. 1) and task as those employed by [Reynaud et al. 2015] who provide a normative dataset on human global stereopsis as a function of spatial frequency. We investigate how the algorithm's performance deviates from that of human observers as a function of its tuning parameters. The algorithm greatly outperforms human subjects when tuned with parameters to compete with non-neural approaches to disparity estimation on benchmarking stereo image datasets (Figure 2, left). Conversely, when the algorithm is implemented with biologically plausible receptive field sizes, spatial selectivity, phase tuning, and neural noise (Fig. 2, center), its performance is directly relatable to that of human observers. The receptive field size and the number of spatial scales sensibly determine the range of spatial frequencies in which the algorithm successfully operates. The algorithm's phase tuning and neural noise in turn determine the algorithm's peak disparity sensitivity. When included, the log polar transform strongly degrades disparity estimation in the model's periphery (Fig. 2, right), further closening human and algorithm performance. Hence, a neuromorphic computer vision algorithm can be reappropriated to model human behavior, and can provide interesting insights into which aspects of human visual perception have been or are yet to be explained by vision science.

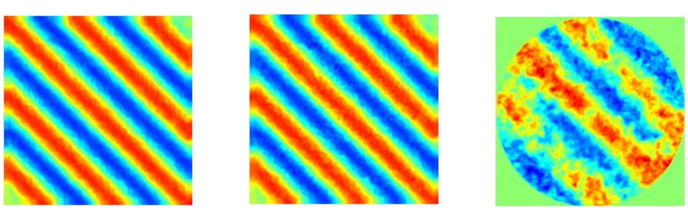

Figure 2: Disparity maps computed for the stimulus in Figure 1 by: the Computer Vision tuned algorithm (left); a non-foveated, biologically plausible tuned algorithm (center); and a foveated, biologically plausible algorithm implementation (right). Disparity range [-15:15] arcmin, mapped onto a red-green-blue colormap.