CogEye: An online active vision system that disambiguates and recognizes objects

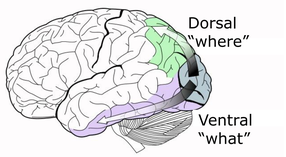
1 Center for Computational Neuroscience and Neural Technology, Boston University
2 Department of Speech Language Pathology & Audiology, Northeastern University, Boston, MA

Lena Sherbakov[1], Gennady Livitz[1], Aisha Sohail[1], Anatoli Gorchetchnikov[1], Ennio Mingolla[2], Heather Ames[1], Massimiliano Versace[1]

Background

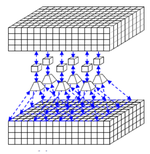
Background information:
○ Computational studies tend to focus on supervised or unsupervised classification algorithms for object recognition disregarding the role of eye movements all together
○ Psychophysics/eye-tacking studies tend to focus on eye movements during object learning (not recognition) or during visual search tasks (which impose a different goal-based objective on saccades)

Our approach: Object recognition (the *what* system) and eye movement (the *where* system) should be studied as a single integrated system.



We developed a bio-inspired active vision computational model, CogEye, that autonomously makes decisions about where to look, acts on those decisions in the form of saccades, learns view-invariant representations of salient objects, and builds mental spatial maps of its surroundings.

What is Cog: CogEye is built on Cog Ex Machina[1] – a GPU-based computing platform designed for simulating large-scale integrative brain systems. Cog is co-developed by Hewlett-Packard Labs and the Boston University Neuromorphics Lab.

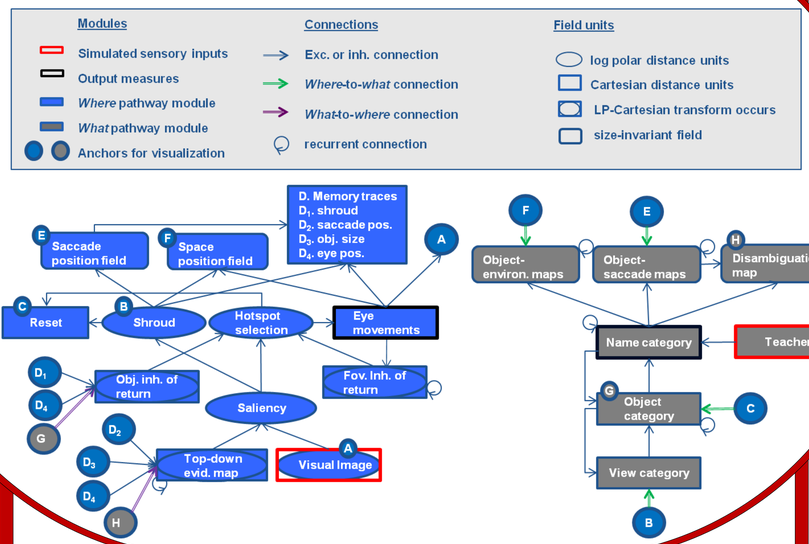

Parallel processing …


on GPUs

Ongoing and future work

○ Eye-tracking study on human subjects, with a particular focus on saccades during the object disambiguation phase
○ CogEye with moving animat
○ Robust figure-ground segmentation based on current foveation
○ Efficient visual search
○ Integration with navigation and path planning
○ Implementation of CogEye on a robotic platform

CogEye: system diagram



Application

Standard object recognition benchmarks: MNIST Digits

Fig 2: After 25,000 training views, CogEye achieves 97.33% correct on a testing set that is 15% the size of the training set.

Studying the role of top-down feedback on object disambiguation

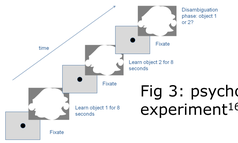
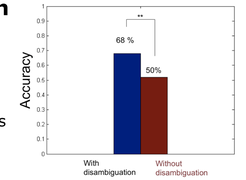
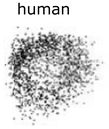


Fig 3: psychophysics experiment[16]




human

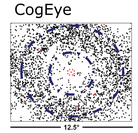
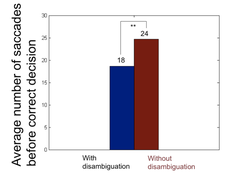
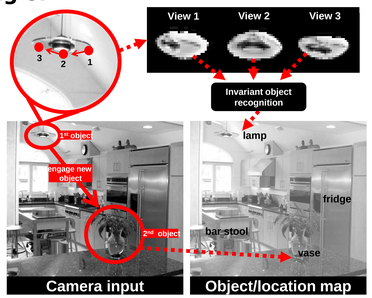

CogEye

Fig 4: CogEye shows similar distribution of foveations to human observers

Fig 5: provides a testable hypothesis for the role of eye movements during object learning and recognition. Disambiguation strategy increases both accuracy of classification and efficiency of saccade allocation.



Continuous learning, recognition, and mapping of a dynamically changing scene



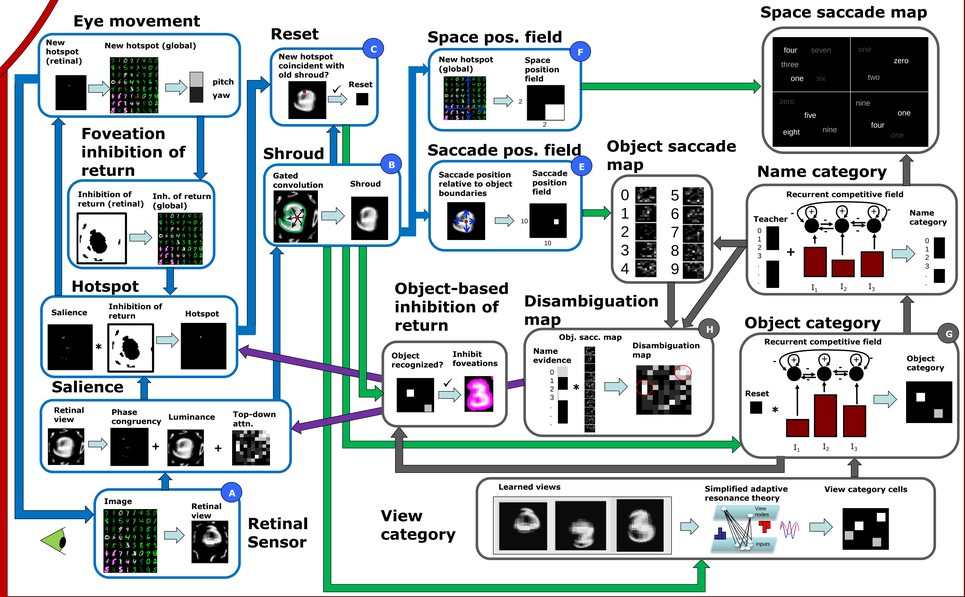
CogEye: biology and function

Module	Brain Area	Function	Algorithm	Approx # Neurons
Visual input	Retina[2]	Space-variant sampling of image	Log-polar transformation[2]	2500
Saliency	V2, V3, V4[3]	Bottom up saliency extraction (edges, corners, phase congruency, luminance)	Gabor filter bank/energy-based phase congruency[4]	217500
Hotspot	V1, V4, PP[5]	Selects location of next foveation	Spatial winner take all	2500
Shroud	PP[6,7,10]	Forms a surface-fitting attentional shroud that suppresses non-foveated surround	Convolution/recurrent excitation	2500
Reset	mSP[8]	Sends a signal when the foveated hotspot is no longer on the previously formed shroud. Binding signal from views to object category.	Coincidence detection	2500
Eye movement	SC, FEF[9]	Moves the eye to the next selected hotspot	Reverse log-polar to linear transformation	22801
Inhibition of return/working memory	SC?,hNTg?	Maintains visual memory of previously foveated locations to prevent repeated foveations to same spot	Reverse log-polar, recurrent memory with leak	274390
Saccade position field	aIPA?	Calculates current saccade position relative to the size and shape of the object	Spatial integration	402
Space position field	PP?	Calculates current saccade position relative to the size and shape of visual environment	Spatial integration	400
Memory traces/corollary discharge	PP?	Maintains memory traces of previous shroud, eye movement, and saccade position field activities for feedback interactions with what pathway	Recurrent memory with leak	2604
Top-down evidence map	PFC,PP[10]	Stores top-down saliency: informative foveation locations based on object ambiguity in what stream	Recurrent memory with leak	298915
Object inhibition of return/WM	SC?, hNTg?	Prevents previously identified objects by the what stream from being redundantly foveated	Reverse log-polar, recurrent memory with leak	294691
View category	pIT[11]	Unsupervised view classification	Simplified fuzzy adaptive resonance theory[12]	1875
Object category	aIT[13]	Accumulates evidence and groups multiple view categories into view-invariant object categories	Recurrent competitive field[14]	400
Name category	PFC[15]	Accumulates evidence and groups object evidence and teacher cues to produce name category	Recurrent competitive field	40
Object-saccade map	PP?	Heatmap of previous within-object foveations and their evidence contribution to object identity	Recurrent spatial integration with leak	1000
Object environment map	PP? HP?	Heatmap of within-scene foveations and their evidence contribution to object and scene identity	Recurrent spatial integration with leak	100
Disambiguation map	PFC?	Calculates object disambiguation power of contending locations of foveation	Greatest variance/mismatch detection	100

V1-V4: primary visual cortex, PP=posterior parietal, mSP=medial superior parietal, SC=superior colliculus, FEF=frontal eye fields, hNTg=habituative neurotransmitter gates, aIPA=anterior intraparietal area, PFC=pre-frontal cortex, pIT=posterior inferotemporal, aIT=anterior inferotemporal, HP=hippocampus

Total: lower bound on number of neurons ≈1.12 million

CogEye details



Acknowledgements and References

Acknowledgements

Supported in part by CompNet, Hewlett-Packard (HP), Neurala Inc., National Science Foundation (NSF) Science of Learning Center CELEST (Center of Excellence for Learning in Education, Science and Technology, NSF SBE-0354378 and NSF OMA-0835976)

REFERENCE

1. Snider, G., Amerson, R., et al. (2011). From synapses to circuitry: using memristive memory to explore the electronic brain. IEEE., 37-44.
2. Schwartz, E.L. (1980). Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. Vision Research, 20 (8), 645-669.
3. Li, W. Piech, V, Gilbert, D. (2006). Contour Saliency in Primary Visual Cortex. Neuron, 50(6), 951-962.
4. Peter Kovesi. (2003). Phase Congruency Detects Corners and Edges. The Australian Pattern Recognition Society Conference: DICTA 2003. Sydney. 309-318.
5. Andersen, R.A., Essick, G.K., & Siegel, R.M. (1985). Encoding of spatial location by posterior parietal neurons. Science, 230 (4724), 456-458.
6. Tyler, C.W., & Kontsevich, L.L. (1995). Mechanisms of stereoscopic processing: stereoattention and surface perception in depth reconstruction. Perception, 24 (2), 127-153.
7. Fazl, A., Grossberg, S., and Mingolla, E. (2009). View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. Cognitive Psychology, 58, 1-48.
8. Chiu, Y., & Yantis, S. (2009). A Domain-Independent Source of Cognitive Control for Task Sets: Shifting Spatial Attention and Switching Categorization Rules. Journal of Neuroscience, 29(12), 3930-3938.
9. Klier, EM, Wang H, Crawford JD (2001). The superior colliculus encodes gaze commands in retinal coordinates. Nat Neurosci ,4 (6), 627–32.
10. Yantis, S., & Serences, J.T. (2003). Cortical mechanisms of space-based and object-based attentional control. Current Opinion in Neurobiology, 13 (2), 187-193.
11. Vuilleumier, P., Henson, R.N., Driver, J., & Dolan, R.J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. Nature Neuroscience, 5 (5), 491-499.
12. Baraldi, A., Alpaydin, E. (2002).Constructive Feedforward ART Clustering Networks – Part I. IEEE transactions on neural networks, vol. 13, no. 3.
13. Tanaka, K. (1997). Mechanisms of visual object recognition: monkey and human studies. Current Opinion in Neurobiology, 7 (4), 523-529.
14. Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. Studies in Applied Mathematics, 52, 213–257.
15. Rainer, G., & Miller, E.K. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. Neuron, 27 (1), 179-189.
16. Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. Journal of Vision, 7(3), 1–6.