Visual Search: Guided Eye Movements for Foveated Sensory Systems

MIGUEL ECKSTEIN
Department of Psychological and Brain Sciences
University of California, Santa Barbara
Email: eckstein@psych.ucsb.edu

Find the parking spot. Find your backpack in the car’s trunk. Find the blue garbage bin to throw the recyclable cup of coffee you were drinking in the car. Find the elevator. Find the button to operate the elevator. Find the key that opens the office. Find the keyhole. Find the outlet to plug in your laptop. Find the power button on the laptop. Find the icon for the email. Life is comprised of short visual searches. Each of these searches often involves moving the eyes to point the central area of the human retina (the fovea) to regions of interest in the scene to extract information related to the search. The fovea processes visual information with high spatial detail. Visual processing away from the fovea (peripheral processing) is mediated by a lower density of cone photoreceptors, higher convergence of cones onto retinal ganglion cells, and fewer associated neurons in primary visual cortex per millimeter of retina. Its consequence: Fine spatial discriminations are not possible with peripheral processing and thus humans often must make eye movements to utilize the fovea to explore a scene and search. But why have many animals evolved this varying resolution (foveated) visual system? Why not have, instead, a visual system that supports fine spatial detail across the entire visual field? The answer is likely related to the high metabolic cost of running a full high resolution system over the entire visual field. Over a ¼ of the brain is already dedicated to vision, and visual processing at full resolution across the entire visual field would increase that substantially. The density of cones in the fovea is approximately 20 times larger than at 10 degrees in to the periphery and 90 times at the far visual periphery (Curcio, Sloan, Kalina, and Hendrickson, 1990). The fovea which occupies 0.01 % of the retina utilizes approximately 10 % of the neuronal machinery in primary visual cortex (Azzopardi and Cowey, 1993). A high resolution processing system across the entire visual field matching the fovea’s ratio of primary visual cortex (V1) neurons per mm of retina would result in roughly a one thousand size increase in the primary visual cortex. Instead, many animals have evolved a varying resolution visual system where a central area is given preferential processing and representation in the brain. However, a foveated visual system relies critically on the guidance of eye movements to efficiently explore
and extract information from scenes. Humans perform these eye movement searches effortlessly and automatically. The brain uses peripheral processing to extract critical information and guides the eyes across the scene. Eye movements are guided by information about the searched target including basic features including color, size, orientation and shape (Eckstein, Beutter, Pham, Shimozaki, and Stone, 2007; Findlay, 1997; Malcolm and Henderson, 2009). The brain is able to acquire information in the visual periphery to guide eye movements concurrent with analyses of information at the foveal region (Ludwig, Davies, and Eckstein, 2014). Yet often times a target can be small and difficult to detect in the visual periphery and the human brain must rely on other visual cues to guide eye movements. Objects in the visual environment are typically not randomly located: fruits tend to be on or under trees, plates are placed on tables, and chimneys on houses. Humans have a remarkable ability to learn statistical relationships among objects (Fiser and Aslin, 2002). The brain uses peripheral processing of a scene and learned relationships among objects or basic features to rapidly guide eye movements towards likely target locations (Figure 1; Eckstein, Drescher, and Shimozaki, 2006; Torralba, Oliva, Castelhano, and Henderson, 2006; Zelinsky, 2008). In searching for a coffee cup, the 1st eye movement (blue squares) is directed within 200-300 ms to countertops and locations that are in agreement with what subjects judge to be likely coffee cup locations (red circles: explicit reports about expected object locations in a scene).

The ability to use scene context to guide eye movements can be an indication of visual expertise in many perceptual tasks. For example, traditional fishing, still practiced in many places in the world, requires a watchman to sit for hours at the top of a hill and visually inspect the ocean surface for the presence of a particular type of fish (e.g., sardines). When the watchman detects the school, he will signal to his fellow fishermen to boat out to the ocean to surround the school with the appropriate nets. Because the presence of the school is often difficult to see on the ocean (Figure 1) with the visual periphery, the watchman typically relies on the presence of certain birds on the ocean surface to indicate the possible presence of the school. The birds’ motion, easily visible in the periphery, guides the watchman’s eye movements to the likely school location for further scrutiny with the fovea.

Where is the information guiding eye movements represented in the brain? In recent years, studies using functional magnetic resonance imaging (fMRI) have suggested that object selective cortex (OSC), previously thought to be only important in the recognition of objects, seems to play a critical role representing searched target in cluttered scenes (Peelen and Kastner, 2011, 2014). Activity in
these areas, prior to the presentation of the visual stimuli, seems to mimic the visually evoked activity by the currently searched target. Prior to search for a person, OSC activity becomes more person-like and while searching for a car, it becomes more car-like. These findings suggest that OSC might represent a target template used to guide search. In addition, the lateral occipital complex (LOC), an area in OSC represents the likely location of a target object in the scene (Preston, Guo, Das, Giesbrecht, and Eckstein, 2013, Figure 2).

References: