Implementing inhibition of return: embodied visual memory for robotic systems
Hülse, Martin; McBride, Sebastian; Lee, Mark

Implementing inhibition of return; embodied visual memory for robotic systems

Martin Hülse   **Sebastian McBride**   **Mark Lee**

Dept. of Computer Science, Aberystwyth University, SY23 3DB, Wales, UK

Abstract

Based on the biological phenomenon of inhibition of return, we introduce an architecture developed for an active robotic vision system where continually updated global information is used to modulate the action selection process for saccadic camera movements. This facilitates, in an extremely efficient way, the fundamental process of avoiding re-saccading to objects previously visited and, thus, is considered to have a wide-ranging application within active vision systems.

Inhibition of return (IOR) refers to the suppression of stimuli (objects and events) processing where those stimuli have previously (and recently) been the focus of spatial attention (Lupianez et al., 2006). In this sense, it forms the basis of attentional (and thus visual) bias towards novel objects. Although the neural mechanism underpinning IOR is not completely understood, it is well established that the dorsal frontoparietal network, including frontal eye fields (FEF) and superior parietal cortex are the primary structures mediating its control (Mayer et al., 2004). These are some of the many modulatory and affecting structures of the deep superior colliculus (optic tectum in non-mammals), the primary motor structure controlling saccade. Although visual information from the retina starts at the superficial superior colliculus, and there are direct connections between the superficial and deep layers, the former cannot elicit saccade directly (Stein and Meredith, 1991). This information has to be subsequently processed by a number of cortical and sub-cortical structures that place it: 1) in context of attentional bias within egocentric saliency maps (posterior parietal cortex) (Gottlieb, 2007), 2) the aforementioned IOR inputs from other modalities (Stein et al., 2002), 3) overriding voluntary saccades (frontal eye fields) (Stein et al., 2002) and 4) basal ganglia action selection (McHaffie et al., 2005). Thus, biologically there exists a highly developed, context specific method for facilitating the most appropriate saccade as a form of attention selection. All of the above saccade-affecting attributes have valuable robotic application but inhibition of return is potentially the most useful in the earlier stages of constructing a saccade system that is attention rather than visual-input driven. For example,

within the most basic of active vision system tasks where static objects of the same shape and color are systematically saccaded to (i.e. brought to the centre of image), there is a consistent need for a mechanism whereby objects already scanned are ignored (i.e. inhibition of return). The primary issue here is that similar image data can emerge in very different image locations, thus the only way of knowing whether an image feature has previously been saccaded to or not, is to store that information at the global level. In the following we introduce an architecture developed for a robotic active vision system where that architecture enables the system to integrate and update global information which can in turn modulate the action selection process for saccadic camera movements.

The active vision system consists of two cameras (both provide RGB 1032x778 image data) mounted on a motorised pan-tilt-verge unit. Three degrees of freedom (DOF) are used: one verge movement for each camera and one tilt which moves both cameras. Each motor is controlled by determining its position in radians (*rad*) where the state of the active vision system is fully determined by the motor positions of the tilt, left and right verge axis, $(p_{tilt}, p_{vL}, p_{vR})$.

The overall computational architecture is illustrated in Figure 1. It consists of three main parts implementing: 1) filtering image data, 2) action selection and execution and 3) the operation of the visual memory. The latter is the central feature of this architecture and main objective of this paper. Without the visual memory, action selection and the resulting saccadic eye movements are determined solely by the current retina image data. Hence, similar visual inputs (RGB image) lead to the same saccade, no matter how often this specific saccade has been executed before. With a visual memory in place, however, specific motor positions $(p_{tilt}, p_{vL}, p_{vR})$ resulting from a successful saccadic camera movement can be stored. This information can then be used to merge the camera image data with the data representing the items present in the visual memory (i.e. those previously saccaded to). The inhibition of return process can then be simply carried out by subtracting the latter from the former, essentially transforming the original camera input into a 'retina-based saliency map' where, objects in the visual memory have been inhibited leaving unsaccaded objects to compete for
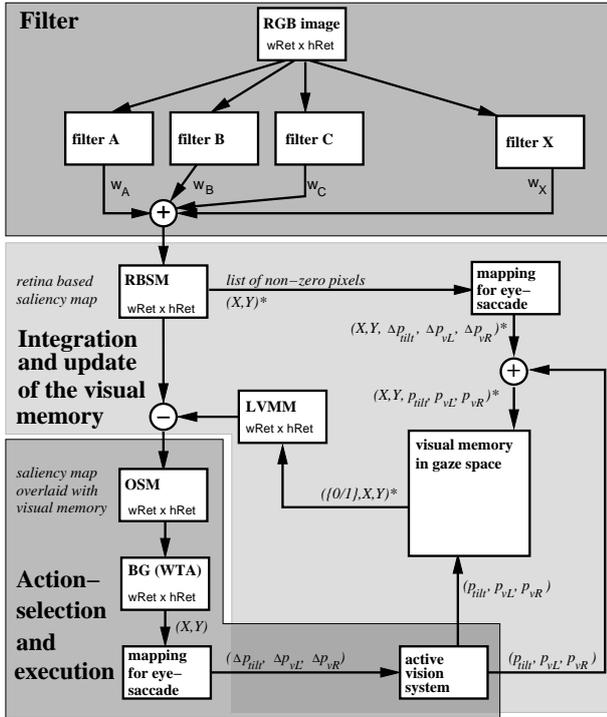
Figure 1: Architecture for embodied visual memory.

the next saccade. In the following the core function of this architecture shall be described in more detail.

A visual buffer (*local visual memory map* or LVMM). and the mapping for the saccadic eye movement (*retina based saliency map* or RBSM) are the essential elements necessary to create the so called *overlaid saliency map* (OSM), see Figure 1. The OSM then feeds into an action selection process: *Basal Ganglia*, (BG). The LVMM represents stimuli which have corresponding entries in the visual memory. The creation of the LVMM is, thus, a crucial part of the architecture. This process starts with RBSM where, for each no-zero pixel in RBSM, the corresponding Δ values (Δp_{tilt}, Δp_{vL}, Δp_{vR}) are derived. These Δ values are learnt beforehand through a mapping process previously described (Lee et al., 2007). Hence, for each non-zero pixel in RBSM we get the relative motor positions (Δp_{tilt}, Δp_{vL}, Δp_{vR}) which drives the particular pixel into the image center. The result of this step is stored as a list where each entry is written as: ($X, Y, \Delta p_{tilt}, \Delta p_{vL}, \Delta p_{vR}$). Notice, in Figure 1 an asterisk signifies a list. Adding these Δ-values to the current absolute motor positions (p_{tilt}, p_{vL}, p_{vR}) provided by the active vision system delivers the final absolute motor positions of the active vision system if a saccade movement was executed. This is again represented as a list: ($X, Y, p_{tilt}, p_{vL}, p_{vR}$). Thus, the Δ-values are replaced by the final absolute motor positions. With this global information the system can now easily ask if a specific pixel (X, Y) in the current RBSM has a corresponding item in the visual memory. If the derived absolute motor positions of pixel (X, Y) can be found in the visual memory

then this pixel is labelled with value of 1 otherwise it is labeled as 0. Thus, all list entries appear as: ($X, Y, \{0, 1\}$). From this list we can then create the LVMM which has the same dimensions as RBSM. Since LVMM contains all previously saccaded to pixels (value 1.0), substraction from RBSM results in the aforementioned 'retina-based saliency map' and an accurate mapping of objects that have not yet been saccaded to.

Although several computational models of inhibition of return of have been put forward e.g. (Sun et al., 2008), the robotic implementation of such a process has, until now, not been fully described. It is noted, however, that the implementation of the architecture presented here, by using lists to store and process data, is not biologically plausible. However, this provides a good real-time performance, and thus is considered as a pragmatic balance between biological constraint and robotic efficiency.

Acknowledgment

References

Gottlieb, J. (2007). From thought to action: The parietal cortex as a bridge between perception, action, and cognition. *Neuron*, 53, 1, 9-16.

Lee, M., Meng, Q., and Chao, F. (2007). Developmental learning for autonomous robots. *Robotics and Autonomous Systems*, 55 (9), 750-759.

Lupianez, J., Klein, R., and Bartolomeo, P. (2006). Inhibition of return: Twenty years after. *Cognitive Neuropsychology*, 23, 7, 1003-1014.

Mayer, A., Seidenberg, M., Dorflinger, J., and Rao, S. (2004). An event-related fmri study of exogenous orienting: Supporting evidence for the cortical basis of inhibition of return? *Journal of Cognitive Neuroscience*, 16, 7, 1262-1271.

McHaffie, J., Stanford, T., Stein, B., Coizet, W., and Redgrave, P. (2005). Subcortical loops through the basal ganglia. *Trends in Neurosciences*, 28, 8, 401-407.

Stein, B. and Meredith, M. (1991). Functional organization of the superior colliculus. In A.G., L., (Ed.), *The neural bases if visual function*, pages 85–100. Macmillan,, Hampshire.

Stein, B., Wallace, M., Stanford, T., and Jiang, W. (2002). Cortex governs multisensory integration in the midbrain. *Neuroscientist*, 8, 4, 306-314.

Sun, Y., Fisher, R., Wang, F., and Gomes, H. (2008). A computer vision model for visual-object-based attention and eye movements. *Computer Vision and Image Understanding*, 112, 2, 126-142.