Primate Structures in Synthetic Dynamic Active Visual Saliency

Andrew Dankers[1,2] **Nick Barnes**[1,2] **Alex Zelinsky**[3]

[1]National ICT Australia[4], Locked Bag 8001, Canberra ACT Australia 2601
[2]Australian National University, Acton ACT Australia 2601
{andrew.dankers,nick.barnes}@nicta.com.au

[3]CSIRO ICT Centre, Canberra ACT Australia 0200
alex.zelinsky@csiro.au

Abstract

We implement biologically plausible early vision processes on a distributed vision system centered around an active stereo vision mechanism. We develop visual saliency for active analysis of real, dynamic scenes. We optimise real-time performance by minimising network traffic and maximising CPU loads in the distributed synthetic vision system. We see that the structures and functional pathways of the synthetic system form an architecture broadly similar to that observed from neural correlates in primates. The correspondence to biology has emerged naively, as a result of performance optimisation, and not by directly modeling the known or hypothesized architecture and functional pathways of primate visual centers.

1. Introduction

A vision system able to adjust its visual parameters to aid task-oriented behavior – an approach labeled *active vision* (Aloimonos et al., 1988) – can be advantageous for scene analysis in realistic environments. We develop an architecture for real-time saliency analysis of realistic, dynamic scenes using active vision. In line with the goals of epigenetic robotics, we work towards a flexible active attention system useful for visual scene analysis in arbitrary environments, rather than optimising a gaze arbitration scheme for a specific environment and/or task.

We begin by implementing cues known to contribute to the perception of saliency in the primate visual cortex. Cues are processed in real time on a network of computers. We adopt biologically plausible techniques to incorporate these cues into a saliency map for the task of actively scanning the scene for regions of interest. We use a deliberately naive approach, that is, we don't require any particular model or structure when distributing processing tasks over the network. The only requirement is real-time performance. The structure of the distributed processing system should emerge naively from real-time optimisation of the saliency task.

Figure 1: CeDAR, active vision apparatus.

We first present the active vision platform (Section 2.). We implement an active rectification step (Section 3.), and describe cues useful for bottom-up saliency (Section 4.). We present an approach to gaze arbitration for use with active cameras and dynamic scenes that inhibits the saliency of previously attended objects, despite their motion, such that uniquely salient scene events are attended (Section 5.). The functional structure of the framework (Section 6.), and results (Section 7.) are subsequently discussed.

2. Platform

CeDAR (Fig.1), the Cable-Drive Active-Vision Robot (Truong et al., 2000), incorporates a common

[4]National ICT Australia is funded by the Australian Department of Communications, Information Technology and the Arts and the Australian Research Council through *Backing Australia's ability* and the ICT Centre of Excellence Program.

Figure 2: Online output from active rectification process: mosaic of rectified frames from right CeDAR camera.

tilt axis and two independent pan axes separated by a baseline of $30cm$. All axes exhibit a range of motion of greater than 90^o, speed of greater than $600^o s^{-1}$ and angular resolution of 0.01^o. Synchronised images with a field of view of 45^o are obtained from each camera at $30fps$ at a resolution of 640x480 pixels, and distributed to a processing network. The mechanical status of the viewing apparatus and acceptance of motion control commands are handled by a dedicated motion control server.

3. Active Rectification

In (Dankers et al., 2004) we described a method to rectify camera barrel distortions and to actively enforce *parallel epipolar geometry* (Hartley and Zisserman, 2004) using camera geometric relations, independent of the contents of the images. The mechanical rectification process, an extension of similar work in (Fusiello et al., 2000), enables online epipolar rectification of the image sequences and the calculation of the shift in pixels between consecutive frames from each camera, and between the current frames from the left and right cameras. We are able to construct globally epipolar rectified mosaics of a scene as the cameras move. Fig.2 shows a snapshot of online output from the mosaic process for a single camera operating at $27fps$. As with the original images, processed cues can be assembled into mosaics. Using such mosaics, the relative location of attended regions can be retained accross saccades.

4. Early Visual Cues

Neurons at the earliest stages in the visual brain are tuned to simple features like intensity contrast, colour opponency, orientation, motion, and stereo disparity. These low level feature maps contribute to the perception of saliency, different features contributing with different strengths (Braun and Julesz, 1998). Relative feature weighting can be influenced by top-down modulation and training (Itti and Koch, 2000). Pre-attentive feature computation occurs continually in primates across the entire visual field. A neuron will fire vigorously even if the animal is attending away from that neuron's receptive field, or if the animal is anesthetized (Treue and Maunsell, 1996). Early visual processing takes around 25-50ms. Little biological evidence exists for strong interactions across different visual features such as colour and orientation (Trieisman and Gelade, 1980). Within a broad feature dimension, strong local interactions between filters (eg, various orientations within the general orientation feature) have been characterised via neuronal correlates (Itti and Koch, 2000). Less evidence exists for within-feature competition across different spatial scales (Itti and Koch, 2000).

In consideration of these findings, we choose conceptually relevant and biologically plausible early visual cues; including depth, optical flow and depth flow, colour, intensity, orientation, collision path criticality, and attended object contouring. Spatial uniqueness in each cue (except the last two) is determined for incorporation into saliency perception. Cue synthesis is subject to real-time performance constraints, so cues are implemented with processor economy in mind. We process images in YUV[1] colour space. In obtaining uniqueness maps, the borders of the image equate to a step that would otherwise produce a significant response. Before processing, we therefore enforce a smooth edge transition by multiplying each image by a windowing function that gradually reduces the intensity values at the edges of the image to zero.

4.1 Intensity Uniqueness

A dark spot in the context of a light background is conceptually unique. That is, intensity contrast is important in saliency, not local absolute intensity (Nothdurf, 1990). In primates, early visual neurons are tuned to spatial contrast and neuronal responses are strongly modulated by context (Allman et al., 1985). Neurons tuned to intensity centre-surround produce a response that can be synthesized using a *difference-of-Gaussian* (DoG) approximation (Itti and Koch, 2000). In a manner similar to (Ude et al., 2005), we create a Gaussian pyramid from the intensity image. Successive images in the pyramid are down-sampled by a factor of two (n times), and each is convolved with the same Gaussian kernel. To obtain DoG images, the Gaussian pyramid images are upsampled (with bi-linear interpolation) to the original size and then combined. Combination involves subtracting pyramids at coarser scales C_n from those at finer scale C_{n-c}. We consider two levels of interaction, immediate neighbours

[1] YUV: Y represents the intensity channel, U and V are colour chrominance channels.

Figure 3: Intensity uniqueness.

Figure 4: Colour uniqueness.

$C_n - C_{n-1}$, and second neighbours $C_n - C_{n-2}$, to obtain a DoG pyramid with $n-3$ entries. Finally, the $n-3$ entries are added to obtain a map where the most spatially unique region emerges with the strongest response. The intensity uniqueness operation is done for both left and right image feeds at frame rate. Fig.3 shows sample output.

4.2 Colour Uniqueness

Colour channels are sent to a separate server for processing in parallel with intensity information. In fact, there exists evidence to suggest that colour is treated in separate regions in the brain to intensity (Dacey, 1996). Colour centre surround uniqueness is computed as per intensity. We process the U and V chrominance opponents separately and combine the result by addition. In this manner, the region with the most unique colour chrominance emerges with the strongest response. Colour uniqueness is calculated for both left and right image feeds at full frame rate. Fig.4 shows output for the colour uniqueness response.

4.3 Optical Flow

The rectification and mosaicing process removes the view-frame effect of any encoded camera geometry changes (pan,tilt). Once the location of the current and previous frame in the mosaic for each camera is known, we calculate optical flow only on the overlapping region of consecutive view frames in the mosaic. This process allows estimation of horizontal and vertical scene flow independent of the motion of the cameras (rather than flow relative to the camera image frame). A *sum of absolute differences* (SAD) flow operation (Banks and Corke, 1991) is used. We obtain four maps from the two cameras: horizontal and vertical flows in each camera. Centre-surround

Figure 5: Optical flow, horizontal direction. The hand (white) moves left, the body (dark) moves right.

Figure 6: Left and right input, and resulting depth map.

uniqueness is determined for all four maps. We down-sample images before computing flow for processor economy. Fig.5 shows sample horizontal flow estimation.

4.4 Depth and Depth Flow

The stereo disparity cue, like flow, involves a SAD disparity search over a small response field. The epipolar rectified mosaics allow us to search for pixel disparities along horizontal scan-lines only. We search only the neighboring ± 16 pixels in the second image for a correspondence to the candidate pixel location in the first image. We conduct the disparity search in the overlapping region of current left and right current frames only. The velocities of visual surfaces in the depth direction are calculated using an approach similar to that of (Kagami et al., 2000). The centre-surround uniqueness algorithm is applied to depth and depth flow outputs. Figs 6 and 7 show depth and depth flow output respectively.

4.5 Orientation Uniqueness

Eye trackers were used to observe that humans preferentially fixate upon regions with multiple orientations (Zetzsche, 1998). Within the broader orientation feature dimension, strong local interactions between separate orientation filters have been characterised via neuronal correlates (Itti and Koch, 2000). A winner-take-all competition is activated amongst

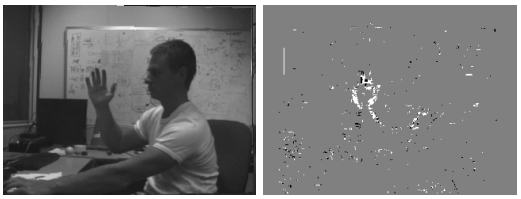

Figure 7: Depth flow, hand moves towards cameras.

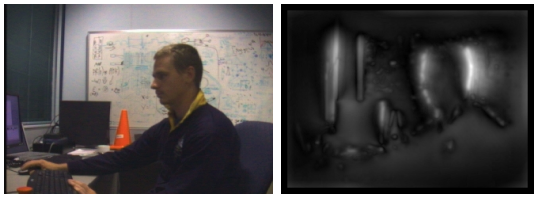

Figure 8: Orientation response,horizontal direction.

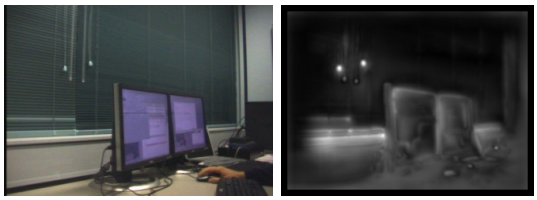

Figure 9: Orientation uniqueness. The multiple orientation responses of the two bright dots stand out from the predominantly horizontal orientation of the blinds.

neurons tuned to different orientations and spatial frequencies within one cortical hypercolumn (Carrasco et al., 2000).

The synthetic response is achieved using complex log-Gabor convolutions over multiple scales within each orientation. The log-Gabor response is akin to the impulse response observed in the orientation sensitive neurons in cats (Sun and Bonds, 1994). The log-Gabor kernel provides a broader spatial frequency response than the Gabor kernel, so fewer scale convolutions are necessary for the same spatial sensitivity. We compute the convolutions in Fourier space and obtain orientation response maps for each orientation and scale. Within each orientation, we sum all scale responses. Processing each orientation is a heavy operation, and because we have four virtual CPUs per server, we limit the operation to four orientations. The associativity of convolution means that the subsequent orientation uniqueness operation (involving a series of convolutions) need not be done for each orientation separately. We can simply sum the orientation maps, and apply the centre-surround uniqueness operation to the result. We obtain orientation response maps for each orientation, a single map of the regions that respond to the most orientations (like corners and edges, Fig.8), and an orientation uniqueness map, Fig.9 where the strongest response occurs at regions that contain orientations not typical to the rest of the image, regardless of scale.

4.6 Critical Collision Cue

The critical collision cue detects pixels on visual surfaces in the scene that are on an instantaneous trajectory leading towards the visual aparatus. A similar neural response has been observed in pigeons

Figure 10: Critical collision cue. The head is moving towards CeDAR. Disparity and flow estimates on the sides of the head (response is better in textured/contrasting regions) enhance the critical collision cue in these regions.

(Wylie et al., 1998). At each pixel where the required measurements exist and are valid, we obtain a position vector $p = (x, y, depth)$ and a velocity vector $v = (flowx, flowy, depthflow)$. We obtain the collision criticality cue according to:

$$\frac{||p||}{||v||}(1 - (-nv \cdot np)), \tag{1}$$

where the dot represents the dot product, and $nv = v/|v|$, and $np = p/|p|$ are unit vectors. That is, the component of the velocity vector associated with a scene point in the direction of the negative distance vector to that scene point is calculated and modulated by the time ($||p||/||v||$) that the scene point would take to get to the origin (the mid point between the cameras) if it were to maintain the current trajectory.

4.7 Zero Disparity and Object Contours

Long range excitatory connections in V1 appear to enhance responses of orientation selective neurons when stimuli extend to form a contour (Gilbert et al., 2000). The result is that monkeys exhibit sparse neuronal activity when viewing complex natural scenes, compared to the vigorous response elicited by small laboratory stimuli in isolation. Accordingly, we develop a cue that responds to attended contoured objects regardless of background clutter. We define a synthetic fovea approximately the size of a fist held a distance of $60cm$ from the camera. For our cameras, this corresponds to a region of about 100x100 pixels.

For humans, the boundary of an attended object emerges effortlessly because the object is centered in, and appears near identically in our left and right retinas, whereas the rest of the scene usually does not. That is, it will be at *zero disparity*. For the synthetic system, the approach is the same; the attended object will appear with identical pixel coordinates in the left and right images. A robust *zero disparity filter* (ZDF) has been formulated (Dankers et al., 2005) to identify objects that map to image frame pixels at the same coordinates in the left and right camera foveas, and their contour. Fig.11

Figure 11: MRF ZDF output (right) with left and right input(respectively), showing foveal processing regions.

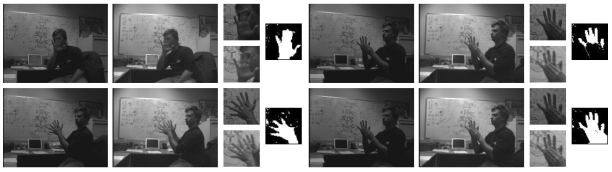

Figure 12: Robust performance in difficult situations: Segmentation of the attended hand from a face in the near background (top left); from a second distracting hand in the background (bottom left); and from a distracting occluding hand in the immediate foreground, a distance of $3cm$ from the tracked hand at a distance of $2m$ from the cameras (top right). Once closer than $3cm$, they are segmented as the same object (bottom right).

shows sample output of the MRF ZDF cue. Fig.12 shows examples of segmentations of the subject (in this case, a hand), where subject-like distractors such as skin, nearby objects, and other hands are present. This cue operates continually because, like primates, the system can only fixate on visual surfaces, not free space, so there will usually be an extractable object centered in the foveas, or the entire fovea will be centred at zero disparity (e.g. attending a large flat object like a brick wall).

The MRF approach can also be used for object-based refinement of cues. For example, to assign the collision criticality cue to an entire ball coming towards us, rather than just the few pixels on the ball that have been detected as such. Object-based contextual refinement of cues and saliency is plausible because humans associate a cue with the entire object, rather than a few points on the object (Pasupathy and Connor, 1999). Object-based cue refinement is beyond the scope of this paper.

4.8 Inter-dependency of Cues

Fig.13 shows cue interdependencies. Serialisation of cue computation can be read off the graph. For example, the collision criticality cue depends on rectification, depth, flow, and depth flow ordered serial pathway.

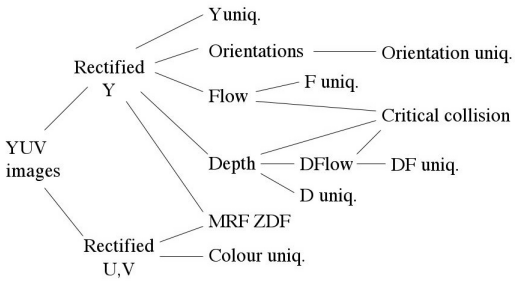

Figure 13: Synthetic cue dependencies.

5. Active Saliency and Gaze Arbitration

We combine centre-surround cues in a fashion similar to the winner-take-all method (Itti and Koch, 2000). In monkeys, salient locations are retained across saccades by transferring activity among spatially-tuned neurons within the intraparietal sulcus (Merriam et al., 2003). Mechanisms of spatial updating maintain accurate representations of visual space across eye movements. Accordingly, our method introduces three separate maps such that inhibition of return (IOR) can be determined for the active cameras with dynamic scenes. The three maps include a Bayesian saliency mosaic, an IOR mosaic, and an image frame fixation map.

5.1 Bayesian Saliency

We incorporate the centre-surround cues into a saliency mosaic using the Bayesian update equation. For each camera, the response level of each pixel in each centre-surround cue for each image is used to increment the probability that the corresponding pixel in the mosaic is salient. Let $s[x, y]$ denote the cue response at pixel location $[x, y]$. Given a cue response measurement M at a pixel $[x, y]$, we use the incremental log-likelihood form of Bayes' Law (Elfes, 1989) to update the saliency map at each pixel. We introduce cue weight W_c corresponding to an empirical weighting of the cue compared to all other cues:

$$\log L(salient) \leftarrow \log L(M \mid salient) + W_c \log L(s) \tag{2}$$

Log-likelihoods provide an efficient implementation for incorporating new data into the saliency map by reducing the update to an addition. Gain W_c could be autonomously updated by higher level operations, representing top-down modulation. In this experiment, the cue weights are declared empirically and remain static.

All entries in the Bayesian saliency map are decayed over time, so that a permanent perception of salience is not anchored to previously attended regions. This decay rate (S_d) describes how easily the

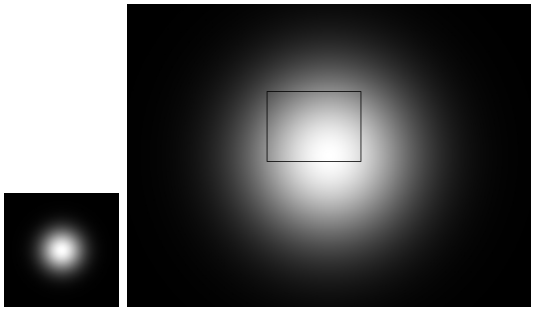

Figure 14: Gaussian IOR increment pattern (left) applied to each view frame. Mosaic-sized radial TSB (right) showing current view frame position. The gradient across the view frame induced by the TSB enhances saliency towards the centre of the mosaic.

system's attention can be distracted. Again, rate S_d could be modulated by higher level processes, depending on the level of concentration required for a particular task. The decay rate also prevents the saliency grid implementation from saturating.

5.2 Inhibition of Return

Inhibition of return (IOR) represents the notion that once we have assessed a particular point in a scene, we are less inclined to look there again. The system evaluates IOR every frame. A Gaussian IOR distribution is applied to the region around the current fixation point (left, Fig.14). Expanding upon this for dynamic scenes, we propagate IOR according to the estimated current optical flow. In this manner, IOR accumulates at the scene location we are attending, but it remains attatched to objects that have been attended as they move. In propagating IOR, it is spread and reduced according to Gaussian uncertainty in the region's new location.

We decrement the entire IOR mosaic over time according to decay rate I_d, so that previously inhibited locations eventually become uninhibited. As with saliency decay rate S_d, faster I_d decay means more frequent saccades to distractors around the scene. Again, this rate can be modulated by higher level operations, though we declare it empirically. With this technique, once we have attended an object such that it is no longer interesting and it then moves to another location in the scene, it is not immediately salient because it carries its inhibition of salience with it until the IOR decay rate, or the uncertainty in its location, reduces the inhibition of its saliency.

5.3 Fixation Map and Gaze Arbitration

To achieve fixation upon salient regions in dynamic scenes with moving cameras, we first modulate (multiply) the saliency map by the IOR map. It is now known that the prefrontal cortex implements attentional control by amplifying task-relevant information rather than inhibiting distracting stimuli (Nieuwenhuis and Yeung, 2005). Accordingly, we modulate the result by introducing a task-dependent spatial bias (TSB) map (right, Fig.14) tailored to the specific task. For example, if we are trying to drive a car, we know that we should tend to keep our gaze upon the road, and as such we bias the lower half of the mosaic where we would expect to find the road. For a forwards search task, we might like to use a radial TSB, such that the system does not tend to divert its gaze too far away from forwards. The TSB may be dynamically updated as appropriate for the current task.

After modulating saliency by IOR and TSB, we find the global peak of the resulting fixation map, and attend the scene point it corresponds to. We can bias the system for specific tasks. For example, by weighting to the saliency of skin-coloured regions, the system preferentially attends to hands and faces, but still attends briefly to other distracting stimuli. Similarly, we are experimenting with biasing the system to attend to the road, road signs, and road lines in the road scene. While preferentially "*keeping it's eyes on the road*", the system briefly evaluates other salient events in the road scene.

6. Functional Structure

We adopt a client-server architecture to allow concurrent serial and parallel functional processing. At the lowest level, the rectification server distributes rectified images and rectification parameters to all other nodes (servers) that require this information. U and V colour chrominance images for both the left and right images are sent to the colour centre-surround server for processing. Intensity images are sent to the other servers. To minimise network bandwidth, to cope with the processing load of each frame, and to prevent repetition of computations, nodes in the structure are configured simultaneously as clients of processes preceding them in the functional serial pathway (Fig.13), and as servers to nodes following them. Each node corresponds to a physically separate PC and all are dual CPU hyper-threaded 3GHz machines, with two physical CPUs amounting to four virtual processors. Trade-offs exist between splitting tasks into sub tasks, passing subtasks to additional nodes, and minimising network traffic. The best performing solution involves grouping of serialised tasks on each server, and that as many operations are done on the image data on the same server as possible, so that there is minimal CPU idle time and minimal network traffic between servers. The serial nature of cue computations means that there is often no gain possible in distributing the task – in fact further network transfer of data between servers would slow performance

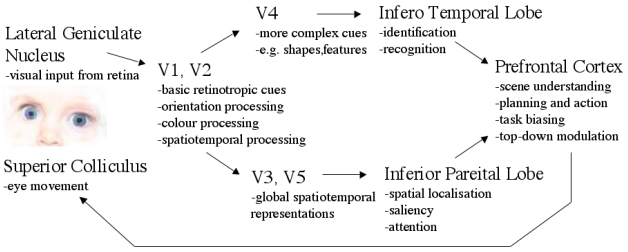

Figure 15: Broad interactions in primate visual brain.

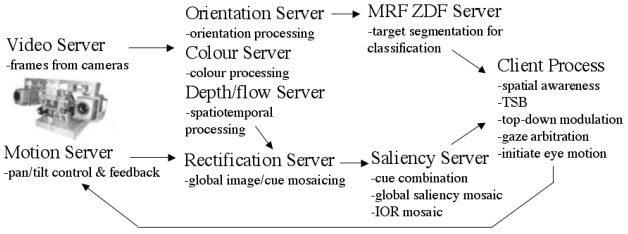

Figure 16: Interactions in synthetic vision system.

significantly. Fig.15 shows a broad model of the major interactions in the primate visual brain. Fig.16 shows the broad structure of the synthetic vision system. It is noted that the synthetic structure bears a good resemblance to the broad interactions between visual centers in the primate brain. Analogies can be drawn between the function of the lateral geniculate nucleus and the video server. Similarly, the motion control server respnds to motion commands in a manner analogous to the superior colliculus. The global representation of space across saccades that occurs in the intraparietal sulcus in V3 performs a function similar to that or the rectification server. The MRF ZDF server extracts attended objects, potentially for identification, in a fashion analogous to the recognition and identification functions of the infero temporal cortex. The orientation, depth and flow, intensity and colour functions are analogous to those occurring in brain areas V1, V2 and V3. The saliency server processes cues in a manner analogous to the inferior parietal lobe. At the highest level, a client process modulates relative cue weightings and updates spatial biasing according to the desired task, which are functions generally considered to occur within the prefrontal cortex. Modulation feedback pathways, such as the ability of the prefrontal cortex to modulate neuronal responses in V1 (or the ability for the client process to modulate cue weightings) have been omitted from the diagrams.

7. Results

The synthetic vision system preferentially directs its attention towards previously unattended salient objects/regions. Upon saccading to a new target, the MRF ZDF cue extracts the object that has grabbed

Figure 17: Objects entering (left) and leaving (right) inhibited region (camera motion disabled). After time, IOR inhibits saliency of the cone near fixation (top left). The previously unattended plate moves in front of the cone, bringing associated low IOR (middle left). IOR accumulates over the plate, reducing its saliency (bottom left). As an inhibited location (top right) leaves the fixation point (bottom right), it takes the associated IOR with it. Behind it, the cone is uninhibited and is initially salient. IOR is grounded within the mosaic reference frame (not the view frame), so as the cameras move, IOR remains associated with objects in the scene.

the system's attention, maintaining stereo fixation on that object (smooth pursuit), regardless of its shape, colour or motion. Attention is maintained until a more salient scene region is encountered, or until accumulated uncertainty in IOR propagation allows previously attended objects to become salient again. Fig.17 shows the interaction between IOR and saliency as an object enters or departs the currently attended region. See *Demonstration Footage* for video sequences of the system actively attending salient scene regions according to this framework.

8. Conclusion

The emergence of attentive behaviors that appear intelligent, and that can be biassed for specific real-time tasks in arbitrary real environments is in accord with the objectives of epigenetic robotics. By implementing biologically plausible early visual cues we have developed a synthetic visual system able to actively divert its attention to salient regions of real scenes in real time. Indeed, the specific processing algorithms may not (and probably do not) reflect what actually happens in the primate brain. By assembling basic cues for the task of detecting saliency, and by optimising the system for performance as a distributed processing network with no prior structure requirement, we note that the distributed structure has naively emerged to form pathways and processing centers broadly similar to those known to exist in the primate visual cortex.

Demonstration Footage

Footage of the system in operation is available at:

`http://rsise.anu.edu.au/~andrew/epirob06`

References

Allman, J., Miezin, F., and McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisions in visual neurons. In *Annu. Rev. Neurosci.*, pages 8:405–430.

Aloimonos, J., Weiss, I., and Bandyopadhyay, A. (1988). Active vision. In *IEEE International Journal on Computer Vision*, pages 333–356.

Banks, J. and Corke, P. (1991). Quantitative evaluation of matching methods and validity measures for stereo vision. *IEEE International Journal of Robotics Research*, 20(7).

Braun, J. and Julesz, B. (1998). Withdrawing attention at little or no cost: detection and discrimination of tasks. In *Percept. Psychophys.*, pages 60:1–23.

Carrasco, M., Penpeci-Talgar, C., and Eckstein, M. (2000). Spatial convert attention increases contrast sensitivity across the csf: support for signal enhancement. In *Vision Res.*, pages 40:1203–1215.

Dacey, M. (1996). Circuitry for color coding in the primate retina. In *Proc. Nat. Acad. Sci.*, pages 93:582–588.

Dankers, A., Barnes, N., and Zelinsky, A. (2004). Active vision - rectification and depth mapping. In *Australian Conf. on Robotics and Automation*.

Dankers, A., Barnes, N., and Zelinsky, A. (2005). Active vision for road scene awareness. In *IEEE Intelligent Vehicles Symposium*.

Elfes, A. (1989). Using occupancy grids for mobile robot perception and navigation. *IEEE Computer Magazine*, pages 46–57.

Fusiello, A., Trucco, E., and Verri, A. (2000). A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22.

Gilbert, C., ito, M., Kapadia, M., and Westheimer, G. (2000). Interactions between attention, context and learning in primary visual cortex. In *Vision Res.*, pages 40:1217–1226.

Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision, Second Edition*. Cambridge University Press.

Itti, L. and Koch, C. (2000). Feature combination strategies for saliency-based visual attention systems. In *J. Electronic Imaging*.

Kagami, S., Okada, K., Inaba, M., and Inoue, H. (2000). Realtime 3d depth flow generation and its application to track to walking human being. In *IEEE International Conf. on Robotics and Automation*, pages 4:197–200.

Merriam, E., Genovese, C., and Colby, C. (2003). Spatial updating in human parietal cortex. In *Neuron*, pages 39:351–373.

Nieuwenhuis, S. and Yeung, N. (2005). Neural mechanisms of attention and control: losing our inhibitions? In *Nature*, pages 8:1631–1633.

Nothdurf, H. (1990). Texture discrimination by cells in the cat lateral geniculate nucleus. In *Exp. Brain Res*, pages 82:48–66.

Pasupathy, A. and Connor, C. (1999). Responses to contour features in macaque area v4. In *Journal of Neurophysiology*, pages 82:2490–2502.

Sun and Bonds (1994). Two-dimensional receptive field organization in striate cortical neurons of the cat. In *Vis Neurosci.*, pages 11: 703–720.

Treue, S. and Maunsell, J. (1996). Attentional modulation of visual motion processing in cortical areas mt and mst. In *Nature*, pages 382:539–541.

Trieisman, A. and Gelade, G. (1980). A feature-integration theory of attention. In *Cogn. Psychol.*, pages 12:97–136.

Truong, H., Abdallah, S., Rougeaux, S., and Zelinsky, A. (2000). A novel mechanism for stereo active vision. In *Australian Conf. on Robotics and Automation*.

Ude, A., Wyart, V., Lin, M., and Cheng, G. (2005). Distributed visual attention on a humanoid robot. In *Report*.

Wylie, D., Bischof, W., and Frost, B. (1998). Common reference frame for neural coding of translational and rotational optic flow. In *Nature*, pages 392:278–282.

Zetzsche, C. (1998). Investigation of a sensorimotor system for saccadic scene analysis: an integrated approach. In *5th Intl. Conf. Sim. Adaptive Behav.*, pages 5:120–126.