Saccade Generation for a Space-Variant Artificial Retina

Sumitha Balasuriya and Paul Siebert
Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, Scotland
{sumitha,psiebert}@dcs.gla.ac.uk

Abstract

Biological and artificial systems that use a space-variant strategy to extract visual information from a scene using a retina face the problem of targeting their sensor so that the central high acuity foveal region inspects salient regions in the scene. At the same time the coarse peripheral region of the retina must extract visual information over a wide field of view to find new interesting locations for future detailed examination with the fovea. This paper describes the saccadic exploration of an image using an artificial retina with a space-variant pseudo-random receptive field tessellation. A space-variant vision hierarchy extracts visual information and accumulates space-variant saliency data to determine the location for the next saccadic fixation.

1. Introduction

This paper reports an investigation into a vision architecture that supports machine sensors which resemble the space-variant sampling characteristics found in human retinae. These artificial retinae have a very high acuity in their central or foveal region and have increasingly reduced acuity towards the retina's periphery. Such a retina will therefore have a wide field of view but only a limited high resolution centre. In a human retina only a tiny fraction of the field of view is sampled with the fovea. Ballistic eye movements called saccades are used to target different scene locations such that we perceive a seamless integrated whole and are rarely consciously aware that our visual system is based on a space-variant sensor.

Biological and artificial systems that use a space-variant strategy to extract visual information from a scene face the problem of targeting the retinal sensor so that the central high acuity foveal region inspects important or salient regions in the field of view. This is not a trivial task, as it is not possible to know *a priori* with confidence whether a region is useful before looking at it in detail with the fovea. An effective attention model that can concentrate limited sampling and processing resources on useful scene points is an integral part of a space-variant vision system.

2. Background

Previous work on machine attention and saccadic control using space-variant visual information can be found in the computer vision literature. Swain et al. (1992) used low resolution colour cues to drive attention of their system in an object search task. They did not use a retina but instead used a coarse version of the image to mimic the low resolution periphery of a retina. The colour cues in the coarse image were used to search for the object. Rao (1994) also did not explicitly use a retina but instead used the log-polar mapping (Schwartz, 1977) to represent the sampling of a retina. Gaussian derivatives at five different scales were used and the goal (target) image was used to create a saliency map in an object search task.

Itti (2000) developed an attention model for focus of attention using centre-surround and double-opponent receptive fields to generate saliency maps, and inhibition-of-return was used to move the focus of attention. While Gaussian pyramids were used to process multiple scales, each layer of the pyramid was not space-variant, i.e. the whole image was sampled independent to the focus of attention. This paper differs from Itti's work by using a pseudo-random space-variant retina to extract visual information and compute space-variant saliency. In our work, inhibition-of-return does not just suppress the saliency of the current fixation point but also causes the artificial retina to fixate on another salient location of the image, thereby extracting novel visual information with the high acuity fovea.

Recently a face authentication system was developed by Smeraldi and Bigun (2002) using a coarse log-polar tessellation with Gabor receptive fields. Support Vector Machines were used as a classifier for the detection and authentication of facial landmarks. However space-variant receptive fields were not used and the system lacked biological plausibility as orientated receptive fields were placed directly on the retinal sensor itself.

3. The Artificial Retina

A space-variant retinal sampling of a scene or image provides a dramatic reduction in the dimensionality of the visual information that must be processed and reasoned with by a biological or machine vision system. We generated an artificial retina by creating a retinal tessellation using a self-organisation methodology called Self-Similar Neural Networks (Clippingdale and Wilson, 1996), and placing overlapping space-variant receptive fields on the retinal mosaic to sample visual information from the image at varying spatial resolutions depending on retinal eccentricity. Difference of Gaussian filters were used to extract achromatic contrast information and filters resembling colour opponent type II retinal ganglion cell receptive fields were used to extract two channels of chromatic contrast information from the image. The reader is referred to Balasuriya and Siebert (2003a,2003b) for details about the construction of the self-organised artificial retina and its receptive fields.

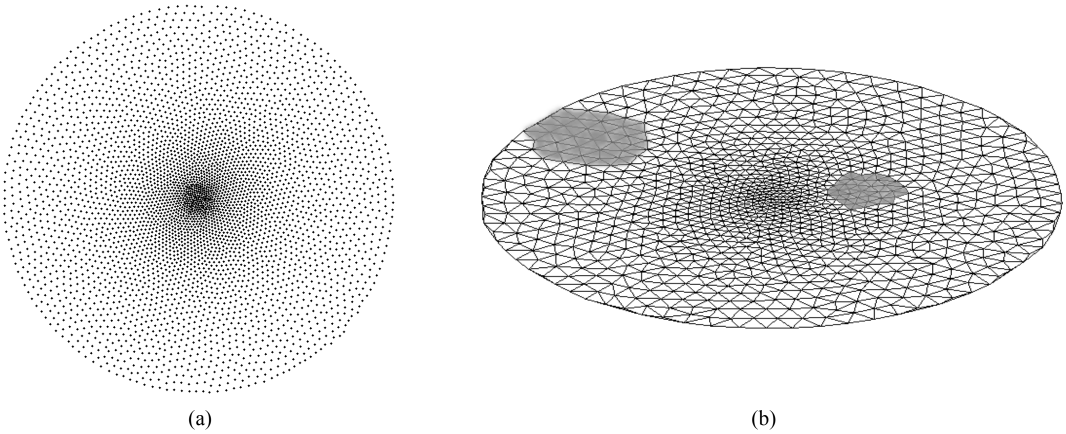

Figure 1. (a) Tessellation of a pseudo-random artificial retina with 4096 receptive fields. The coarsely sampled peripheral region of the retina gives the system a wide field of view, while the fovea provides a high resolution detailed sampling at the centre of the retina. (b) Cortical graph receptive fields associated with a retina consisting of 1024 receptive fields. There is a cortical receptive field centred at each node over the retinal tessellation resulting in a cortical graph with an equal number of nodes. Two cortical receptive fields with constant radius of three edges on the cortical graph yet varying spatial support in the image domain are highlighted.

4. Higher Level Feature Extraction

The features extracted by the circularly symmetric filters on the artificial retina were analysed further for complex features suitable for reasoning. A cortical graph structure (Figure 1b) was used to define cortical filters that are uniform in size in the artificial cortex and space-variant in the image domain. Delaunay triangulation was used to create topological relationships between nodes in the graph. Edge features were extracted by processing the achromatic difference of Gaussian output from the retina with Gabor cortical filters. The chromatic opponent information output from the retina was processed using chromatic double opponent cortical filters. Details about the cortical graph and the construction of cortical filters can be found in Balasuriya and Siebert (2003b).

A winner-take-all scheme was used to select the dominant Gabor response orientation from each location on the cortical graph. These responses were accumulated within a Gaussian neighbourhood at each orientation and were cycled to a canonical orientation (orientation with the largest response) to make the extracted information invariant to rotation in image plane. Processing resources of the system were economised by only calculating cortical filters at locations (nodes on the cortical graph) which were co-located with a significant retinal response.

5. Space-Variant Saliency and Saccade Generation

A saliency map was computed by aggregating the absolute values of the responses from the co-located achromatic and chromatic cortical filters. The system accumulates saliency by assimilating saliency information from the current fixation into an evolving saliency map. However the saliency values need to be incorporated into the saliency map reflecting not only the degree or significance of the saliency value but also the spatial scale of the represented salient region.

This was achieved by distributing the saliency values from the current saccade using the artificial retinal tessellation itself. Gaussians blobs were placed on a null image with amplitude corresponding to the associated saliency value and with size (i.e. Gaussian standard deviation, σ) corresponding to the co-located retinal receptive field. Therefore the saliency map, *Current_Smap*, generated solely from the current fixation would be

$$Current\_Smap(x,y) = \sum_r \frac{O_r}{2\pi\sigma^2} e^{\frac{-((x-x_r)^2 + (y-y_r)^2)}{2\sigma^2}}$$

where O_r is the aggregated response (saliency value) from cortical filter r centred on (x_r,y_r) and the accumulated saliency map *Smap* was calculated as

$$Smap(x, y) = Current\_Smap(x,y) \quad \text{if } Current\_Smap(x,y) > Smap(x,y)$$

In most applications there is no direct advantage in the fovea re-inspecting locations when examining a static image. Therefore an inhibition-of-return map was used to prevent the retina repeatedly re-fixating upon highly salient locations on the image. The inhibition-of-return map *Imap* was generated by placing Gaussians (with problem specific standard deviation σ_i and scaling factor A) at each saccadic fixation point (x_f,y_f) as the retina examines the image.

$$Imap(x,y) = Imap(x,y) + \frac{A}{2\pi\sigma_i^2} e^{\frac{-((x-x_f)^2 + (y-y_f)^2)}{2\sigma_i^2}}$$

The next fixation point in the image (x_f,y_f) for the retina was determined by,

$$Smap(x_f,y_f) - Imap(x_f,y_f) \geq Smap(x,y) - Imap(x,y) \quad \text{where } x_f \in x \text{ and } y_f \in y$$

and the system was instructed to saccade until,

$$Smap(x,y) - Imap(x,y) < 0 \quad \text{for } \forall x \, \forall y$$

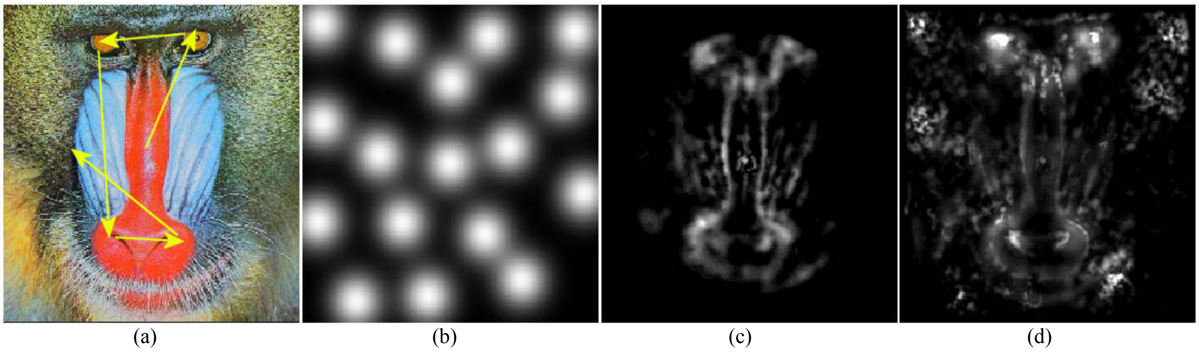

Figure 2. (a) Mandrill colour image with the first five retinal fixations. The retina was initially fixated upon the centre of the image. (b) Inhibition-of-Return map after 17 fixations. (c) Saliency information from a fixation at the centre of the image. The retina almost spans the whole image when it fixates on the centre and space-variant saliency information, detailed in the fovea and coarse in the periphery, can be observed. (d) Accumulated saliency map after 17 fixations. It is interesting to note that the system found the mandrill's eyes and nostrils to be highly salient.

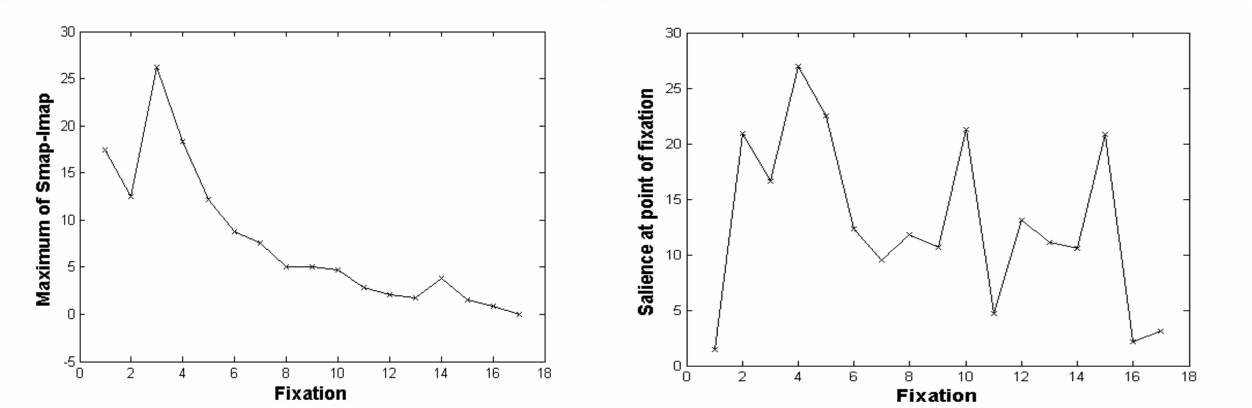

Figure 3. (a) A plot of *Smap-Imap* of the next fixation point when the retina was examining the Mandrill image. The retina stopped saccading when this value was below zero. (b) Salience value at the point of fixation. The system does not know the true salience of a location until it is examined with the fovea. Therefore the salience value at the point of fixation may not monotonically decrease as the retina examines the image.

6. Conclusion and Future Work

An approach for generating saccades for a pseudo-randomly tessellated space-variant retina was presented. The retina would serially fixate on highly salient image locations, accumulating saliency information reflecting the degree of saliency and the space-variant characteristics of the vision system.

We plan to extend this approach to accommodate top-down attention into the saliency calculation in object recognition and object search tasks using the implemented visual hierarchy. The impact of extracting visual information using a pyramid of retinae that samples space-variant information at several scales will also be investigated. A simplified implementation of lateral inhibition will be used to sparcify low level visual information.

References

Balasuriya, L. S. and Siebert, J. P. (2003a). An artificial retina with a self-organised retinal receptive field tessellation. Biologically-inspired Machine Vision, Theory and Application symposium, AISB, Aberystwyth.

Balasuriya, L. S. and Siebert, J. P. (2003b). A low level vision hierarchy based on an irregularly sampled retina. CIRAS, Singapore.

Clippingdale, S. and Wilson, R. (1996). "Self-similar Neural Networks Based on a Kohonen Learning Rule." *Neural Networks* **9**(5): 747-763.

Itti, L. (2000). Models of Bottom-Up and Top-Down Visual Attention, California Institute of Technology.

Rao, R. P. N. (1994). Top-Down Gaze Targeting for Space-Variant Active Vision. ARPA.

Schwartz, E. L. (1977). "Spatial mapping in primate sensory projection: Analytic structure and relevance to perception." *Biological Cybernetics* **25**: 181-194.

Smeraldi, F. and Bigun, J. (2002). "Retinal vision applied to facial features detection and face authentication." *Pattern Recognition Letters* **23**: 463 - 475.

Swain, M. J., Kahn, R. E. and Ballard, D. H. (1992). Low Resolution Cues For Guiding Saccadic Eye Movements. CVPR.