An artificial retina with a self-organised retinal receptive field tessellation

L. S. Balasuriya and J. P. Siebert
Department of Computer Science
University of Glasgow
sumitha@dcs.gla.ac.uk psiebert@dcs.gla.ac.uk

Abstract

In this paper we present a means to generate a receptive field tessellation for an artificial retina using a self-organisation methodology. The approach addresses the problem of creating a receptive field tessellation that seamlessly changes from the uniform fovea to the space-variant periphery of a retina. We sampled and reconstructed images by placing overlapping space-variant receptive fields on this retinal tessellation and scaling the field size according to the local node density. The results of sampling and reconstructing images using the above self-organising retina tessellation are presented.

1 Introduction

An artificial retina that can sample an image or scene in a space-variant manner is an essential front end to a biologically inspired machine vision system. This sampling defines the information content, representation and processing of the entire vision system. Researchers have reported problems computing plausible retinae that sample an image without over-sampling the central foveal region or creating discontinuities in the retinal tessellation as the spatial distance between sampling points increases with eccentricity (refer section 3).

Conventional approaches for creating retinal tessellations have been based on analytic transforms. However, the authors question the tractability of the problem, from an analytic perspective, that meets the constraints of a continuous regular (in the fovea) to log-polar (in the periphery) sampling regimen.

This paper describes the design, implementation and initial evaluation of a space-variant, continuous retina with receptive field locations that were determined by self-organisation. This is essentially an artificial (software) sensor that can be used to sample an image. Simple difference of Gaussian receptive fields (Marr, 1982) with varying receptive field sizes were placed on the generated retinal tessellation to extract visual information channels from an input image that correspond to the distinct layers of the X visual pathway. This multi-resolution, multilayer representation of the input image has been structured to form a space-variant image pyramid (Burt and Adelson, 1983). We have demonstrated the inverse process of superimposing the retina-space image reconstructed from each layer of our pyramid to regenerate a foveated version of the original input image.

2 Motivation and Approach

The physiology of the primary visual cortex comprises uniform units of neurological machinery that processes the whole visual field of view (Hubel and Wiesel, 1979) arriving from the retina. Therefore, a biologically motivated artificial retina should be capable of extracting visual information such that it can be processed by uniform computational machinery in an artificial cortex.

Wilson (1983) showed that retinal structures could extract information such that the uniform cortical machinery would output an ‘ensemble of messages’ that does not change with differences in scale and orientation of an object for a given point of fixation. The implementation of such a system can then result in the coding of an object irrespective of its scale and orientation.

In order to achieve image coding uniformity, it is necessary to implement a retina tessellation model that similarly preserves sampling continuity. However, real retinas appear to exhibit an almost uniform hexagonal receptor packing structure in the fovea that transforms seamlessly into an exponential

sampling structure (log-polar retino-cortical transformation) towards the peripheral field. While the exponential sampling strategy of the retina periphery appears to have many desirable properties, capable of ameliorating subsequent visual information processing (Schwartz, 1977 and Wilson, 1983), it must transform into a uniform fovea without generating discontinuities to allow a single set of uniform 'coding units' to be constructed in the cortex. These coding units would then be able to represent the local (multi-resolution) image features (together with their topographic relationships) to generate Wilson's "ensemble of messages". Subsequent higher analysis might take the form of dynamic link matching (von der Malsburg, 1981), classical graph matching or more radical schemes such as spike coding (Thorpe and Gautrais, 1997).

Figure 1: The standard Lena input image that was used for all image processing operations in this paper.

While continuity in the sampling strategy is the key to achieving a uniform coding strategy, the nature of the particular form of space-variant sampling adopted is also an issue. Exponential sampling provides the means to achieving quasi-scale invariance and a conformal retina-to cortical mapping ensures that local codings retain their approximate invariance to scale changes.

The retina model itself can be implemented within a filter-sample scheme, where bandpass filters are located at locations defined by the exponential sampling strategy and these filters are scaled spatially to ensure that the output cortical image does not contain aliasing components. In other words, a frequency shift (frequency normalisation) is taking place in the sampled retinal image with respect to sampling eccentricity that results in uniform cortical (radial) spatial frequencies for exponentially increasing retinal (radial) spatial frequencies.

A foveated, space-variant extraction of visual information from an image prevents the combinatorial explosion of information and computational processing suffered by conventional approaches where the whole image is given equal processing emphasis. A retina will reduce the dimensionality and bandwidth of the image data that is being sampled and processed. When the retina is fixated at a salient object in the image the important foveal region of the image is sampled with a very high sampling density while the less important peripheral region is given less emphasis.

Researchers have not been able to define a retina with the above properties that can extract visual information from an image without over-sampling the image or creating discontinuities in the retina tessellation (section 3). In fact, there are few examples of full vision systems that classify or conduct recognition tasks with information extracted from a space-variant retina (Smeraldi et. al., 1998).

Solving the problem of generating a viable retina model will allow biologically motivated vision to progress and therefore if we can construct a continuous retinal tessellation that circumvents the mentioned limitations, this can drive attentional structures which are part of an overall vision system.

The objective of this paper is to present an approach to create a retina for such an artificial visual pathway.

3 Previous retina models

Most space-variant retinae used in conventional machine vision systems are based around a log-polar sensor. The mapping from image space to what is called a 'cortical space' is computed according to the log(z) model proposed by Eric Schwartz (1977). If $z = x + jy$ is the location (in complex coordinates) of a receptive field centre in the image plane, then the location of its output in the artificial cortex is given simply by log(z) where,

$$\log(z) = \log(|z|) + j \arg(z)$$
$$= \log(\text{eccentricity}) + j\,(\text{angle})$$

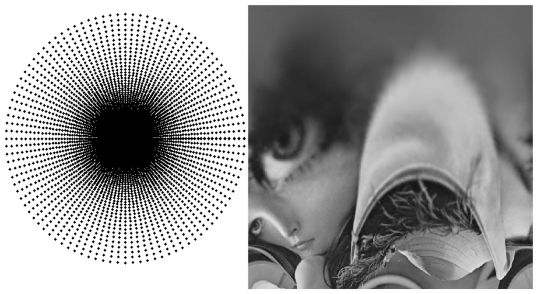

Figure 2: A retina with receptive field locations based on the log(z) model (left). A cortical images created by sampling the standard Lena image with Gaussian receptive fields placed on the retina (right).

Higher level processing can now be conducted on the cortical image (figure 2) as this is a space-variant, foveated representation of the image. Because an image has an underlying minimum spatial frequency any attempts to sample an image at higher spatial frequencies would result in a highly correlated output. This super-Nyquist sampling can be observed in the foveal region of the log(z) retina (figure 2). In fact a large percentage of the cortical image is highly correlated information and the sampling process has not optimally reduced the dimension of the extracted visual information.

The log(z+α) model by Schwartz(1980) tried to avoid the singularity at the fovea and reduce the over-sampling in the foveal region, however the highly correlated foveal region in the cortical image can still be seen for realistic values of α (figure 3). Furthermore the log(z+α) retinal tessellation has an anisotropic fovea which is elongated vertically. An interesting feature of the log(z+α) model is that the cortical image is split along the vertical meridian into two visual hemispheres as in biology.

Figure 3: A schematic representation of receptive field locations based on the log(z+α) model (left). A cortical images created by sampling the standard Lena image with Gaussian receptive fields using the log(z+α) model (right).

Researchers have also tried to separate the foveal and peripheral regions of the retina (Bolduc and Levine, 1998) and thereby uniformly sample the foveal region while sampling the peripheral region with a space-variant tessellation. However one can see that this creates a distinct discontinuity in the sampling between the fovea and periphery (figure 4).

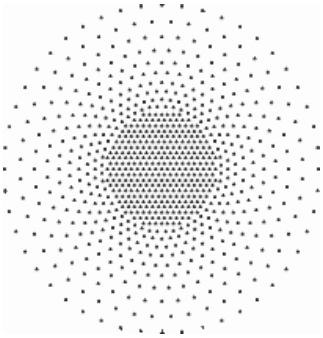

Figure 4: A plot of the receptive field centres from a retina with a uniform foveal region.

We haven't found any analytic or geometric mapping that can describe the change in topography of the receptive field centres between the fovea and periphery. However by forgoing geometric regularity in the retinal structure, a self-organisation methodology (Clippingdale and Wilson, 1996) has enabled us to build retinal tessellations that instead exhibit continuity in sampling density. The final structure of such a retina locally resembles a pseudo-regular hexagonal lattice with sight deviations in the hexagonal topology in some locations.

4 Generating a self-organised tessellation

The retinal tessellations in Section 5 were created using Clippingdale and Wilson's (1996) 'self-similar neural network' model. The tessellations derived using this model have uniform foveal regions which seamlessly merge into the space-variant periphery of the retina (figures 6 and 7). The 2-dimensional coordinates of the receptive field centres of the tessellations are represented in the network as x_i network weights which are stimulated by an input stimulus which is derived by applying a random transformation to the network weights themselves.

Therefore, for a network of N units, each characterised by a 2 dimensional weight vector, the input stimulus $y_i(n)$ at iteration n is calculated by the following,

$$y_i(n) = T(n)\, x_i(n-1)$$

where $x_i(n-1)$ is the ith network unit at iteration $n-1$ and $1 \le i \le N$. In our work we used the following T composite transformation:

(i) A random rotation between $[0, 2\pi)$

(ii) A dilation (increase in eccentricity) comprising of the exponent of a dilation factor which is random between $[0, \log(8))$

(iii) Translations in the vertical, horizontal and radial (away from centre) directions random between $[0, f)$, where f is associated with the required foveal percentage of the resultant retina.

Any input stimuli $y_i(n)$ which lie outside the bounds of the retina were culled before the network weights $x_i(n-1)$ were stimulated to calculate $x_i(n)$.

The network was initialised with a random weight configuration and iterated with the above transformations and the following learning rule:

$$x_j(n) = x_j(n-1) + \alpha(n) \sum_{i \in \Lambda_j(n)} \left(y_i(n) - x_j(n-1) \right)$$

where,

$$\Lambda_j(n) = \left\{ \begin{array}{l} i : \left\| y_i(n) - x_j(n-1) \right\| \\ \qquad < \left\| y_i(n) - x_k(n-1) \right\|, k \neq j \end{array} \right\}$$

$\Lambda_j(n)$ contains the indices to the input stimuli $y_i(n)$ to which $x_j(n-1)$ is the closest network vector. $\alpha(n)$ is a learning parameter which controls the stimulation of the network weights. We linearly reduced (annealed) α from 0.1 to 0.001 throughout the self-organisation to increase convergence.

Intuitively, one can visualise each network unit being updated individually by the input stimuli that are closer to that unit than any other unit in the network. This training methodology causes the network to evolve a regular tessellation which is governed by the composite transformations T.

The retinae that evolved using self-similar neural networks tend to have a regular hexagonal packing tessellation or mosaic.

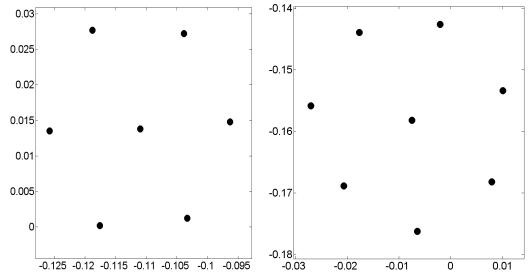

Figure 5: Magnified areas of a self-organised retina that show different packing mosaics.

This is not surprising as a hexagonal tessellation is the approximate pattern in which receptive fields are found in biological retinae (Polyak, 1941). In fact this is a pattern commonly found throughout nature (Morgan, 1999). Receptive fields placed on this tessellation would then be equidistant from their immediate neighbours and Dudgeon and Mersereau (1984) showed that such a hexagonal tessellation is the optimal sampling scheme for a 2D space.

Close inspection of the tessellation (figure 5) shows that at some locations in the retina the tessellation undergoes an intermediate transition phase. In these locations the packing mosaic deviates slightly from a hexagonal cell construction (figure 5, right). These deviations occur relatively frequently in the region between the fovea and periphery of the retina where the retinal tessellation is changing from a uniform to a space-variant packing while maintaining regularity of the node mosaic.

Further details about self-similar neural networks and a discussion of their convergence properties are given by Clippingdale and Wilson (1996).

5 Results of the self-organisation process

Figure 6 contains a retinal mosaic that was created using self-similar neural networks with translations in the horizontal and vertical

directions up to 20% $(f = 0.2)$ of the radius of the retina (refer section 4). As in Clippingdale and Wilson (1996) no radial translations were used to create input stimuli. The tessellation comprises of 4096 network nodes and was self-organised for 250000 iterations. This tessellation was self-organised for an extremely long period in comparison to our other experiments and the high iteration count let the network converge to a remarkably regular mosaic.

The foveal region of the tessellation has a distinctly square shape which was caused by limiting translations to the horizontal and vertical directions. We therefore decided to experiment with different translation approaches to generate a more plausible retina.

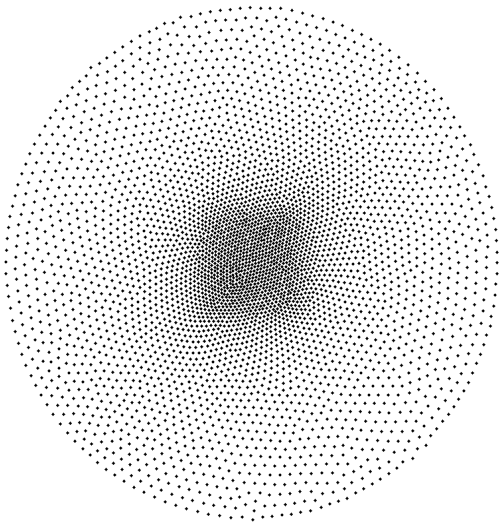

Figure 6: A plot of a self-organised retina with translations made in horizontal and vertical directions.

When we translated a 4096 node network radially instead of in the horizontal and vertical directions we obtained the pattern in figure 7. The radial translation was up to 20% $(f = 0.2)$ of the radius of the retina (section 4) and the network was self-organised for 20000 iterations.

This approach resulted in a mosaic with a higher packing density in the fovea. However the centre of the fovea is approximately uniform and does not reach singularity (figure 8).

We were able to create a tessellation with a large uniform foveal region which was isometric by self-organising a network with a composite of horizontal, vertical and radial translations.

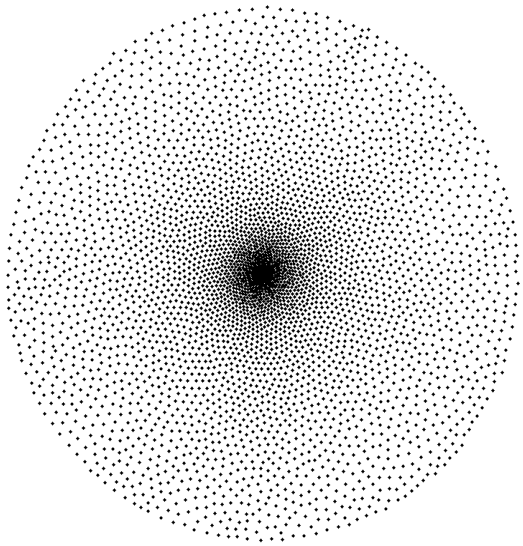

Figure 7: A plot of a self-organised retina with translations made radially.

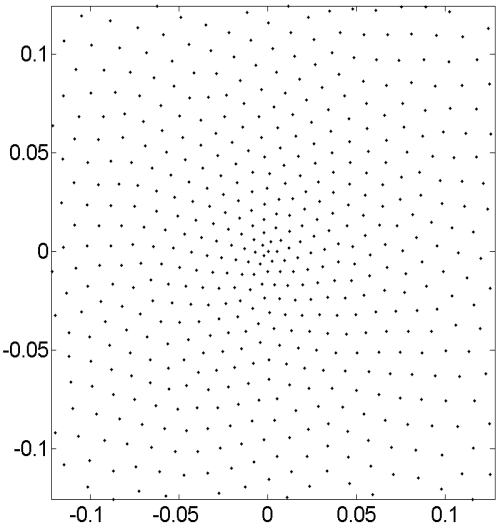

Figure 8: A magnified view of the fovea from the retina illustrated in figure 7 (the radius of the retina is 1 unit).

Figure 9 contains the mosaic that was obtained after self-organising with a radial translation that was up to 20% $(f = 0.2)$ and horizontal and vertical translations which were up to 6.6% $(f = 0.066)$ of the radius of the retina. The network consists of 4096 nodes and was annealed for 20000 iterations.

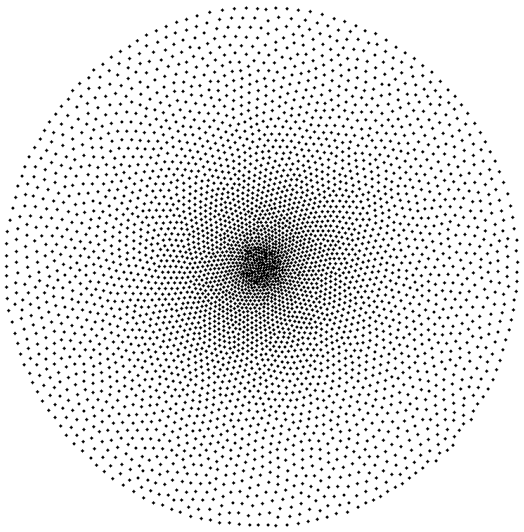

Figure 9: A plot of a self-organised retina with translations made horizontally, vertically and radially.

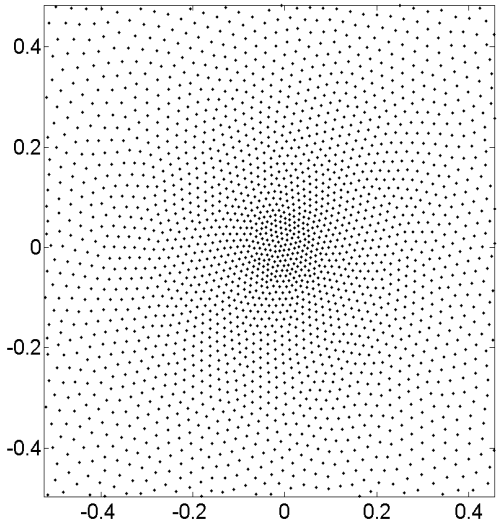

Figure 10: A magnified view of the fovea from the retina illustrated in figure 9 (the radius of the complete retina is 1 unit).

Figure 11 contains a plot of the node density of the retina in Figure 9 with eccentricity. The nodes were binned according to eccentricity into 50 bins and the resulting histogram frequency was calculated. The plot reflects the topography of the retina with a node density reaching a plateau in the fovea.

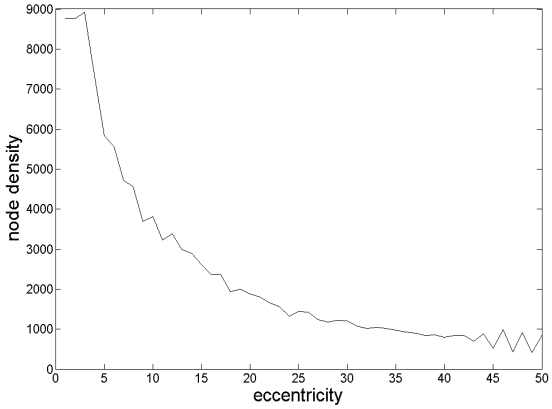

Figure 11: Node density of the retina in figure 9.

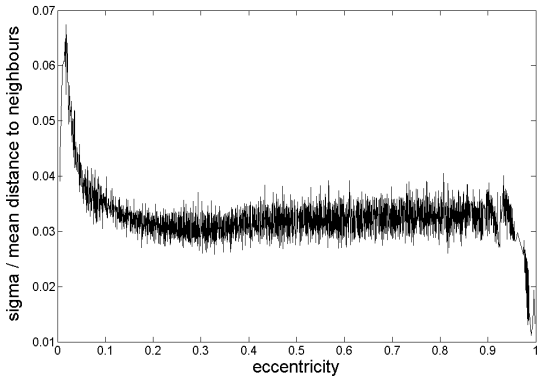

Figure 12: Standard deviation of the distance to a node's immediate neighbours in the retina in figure 9.

The standard deviation (figure 12) of the retina increases in the foveal region and the authors' hypothesise that this is due to the tessellation trying to maintain regularity while being subject to increased stimulatory forces in the fovea.

Further experiments revealed that the node density and variance values of retinae exhibit less variability as one increases the number of iterations during self-organisation.

6 Placing receptive fields on the self-organised tessellation

We constructed retinae out of the self-organised tessellations by placing overlapping receptive fields (or filters) at the calculated node locations. The receptive fields were placed on the calculated node locations with sub-pixel accuracy by altering the coefficients of the filters to reflect the sub-pixel jitter. The size of a receptive field was calculated based on the mean

distance to its nearest neighbouring nodes scaled by a suitable scaling factor (we used 6). This let us create space-variant, overlapping receptive fields. The retina was then scaled to an image such that the mean distance between closest nodes in the fovea was equal to one pixel. Figure 17 is a reconstruction of a cross stimulus that clearly illustrates the space-variant receptive field sizes in the retina. Difference of Gaussian (DoG) receptive fields were used in these experiments to resemble the processing of a simplified retinal ganglion cell layer.

An approximation to a space-variant octave Laplacian pyramid was constructed by sampling an image with retinae constructed from several self-organised tessellations (figure 16). The number of nodes in the tessellations was reduced by a factor of 4 between resolution levels of the octave pyramid.

7 Results of sampling with the retina

All the following results are from retinae that were constructed using self-organised tessellations that were translated horizontally, vertically and radially (figure 9). The reconstruction images are reversals of the sampling operations of the retina on an image. The low dimensional extracted responses from a retinal sampling (for example a 4096x1 dimensional vector in the case of a 4096 node retina) were distributed over the respective receptive fields of the retina to form the reconstruction.

Figure 15 contains the reconstruction of the sampling of the greyscale Lena image by a 4096 node retina with difference of Gaussian receptive fields. Since the receptive fields near the point of fixation are small, these have sampled and reconstructed higher spatial frequencies than in the periphery where there is reduced acuity in the reconstruction. The retina has a restricted field of view and therefore has not sampled the entire Lena image. This was because we scaled the retina to the image as described in section 6.

Figures 13 and 14 contain the reconstructions from retinae with 64 and 256 nodes respectively. We similarly constructed retinae with 16 and 1024 nodes and used all these retinae to construct a superposition of the reconstructions of the Lena image which were sampled by retinae with 4096, 1024, 256, 64 and 16 difference of Gaussian receptive fields (figure 16). The vector that was used for the reconstruction is just 5456x1 (4096+1024+ 256+64+16) yet we still obtained a detailed space-variant representation of the Lena image.

8 Conclusion

We have created a plausible artificial retina with a regular receptive field tessellation and continuity between foveal and peripheral regions. While the tessellation was locally pseudo-random it maintained a sampling density continuum at a macroscopic level, though at the price of sacrificing (analytic) geometric uniformity.

Despite the pseudo-random nature of the tessellation, it is still possible to conduct useful computation on this visual representation. To demonstrate this, space-variant difference of Gaussian receptive fields were placed on the generated retinal tessellation to simulate 5 layers, i.e. 5 separate spatial frequency bandpass channels, of simplified ganglion cells. We then successfully used the resulting retina to sample image information and reconstruct a space variant version of the original input.

9 Future Work

The visual information extracted using this retinal front end will need to undergo further processing in subsequent stages of a visual pathway. However performing image processing operations on visual information extracted using an irregular sampling regimen is not trivial. Wallace et. al. have proposed connectivity graphs for reasoning with space variant images (Wallace et. al., 1995). Since the retina we used was not based on an analytical transform such as the log(z) type mappings in Wallace et. al. (1995), we recently adopted Delaunay triangulation (Barber et. al. 1996) to define a '*topological cortical space*' which we were able to further transform by applying image filtering operations.

To perform a convolution with any arbitrary filter we calculated the coefficients of the filters at topological cortical space node locations. We have already been able to filter oriented image features from the responses extracted from the retina by placing Gabor receptive fields on the topological cortical space and hope to demonstrate the uniform coding of visual information across the entire visual field when using our representation.

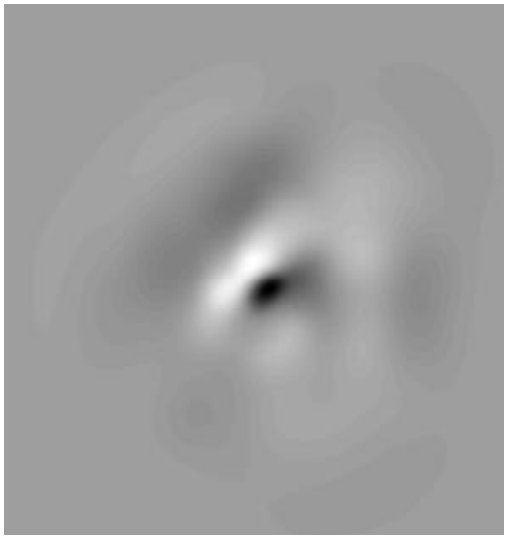

Figure 13: Reconstruction of the sampling of the greyscale Lena image using a 64 node retina with difference of Gaussian receptive fields.

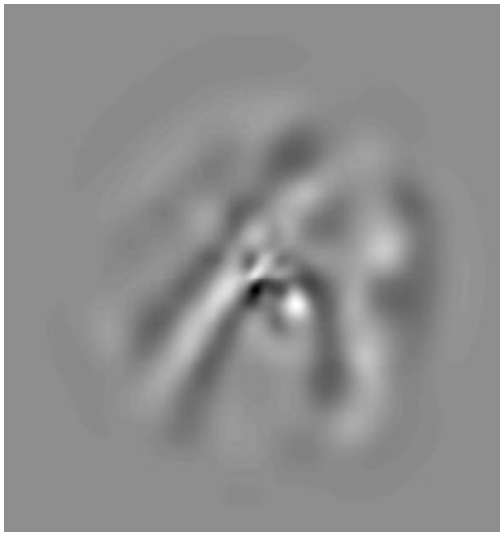

Figure 14: Reconstruction of the sampling of the greyscale Lena image using a 256 node retina with difference of Gaussian receptive fields.

Figure 15: Reconstruction of the sampling of the greyscale Lena image using a 4096 node retina with difference of Gaussian receptive fields.

Figure 16: Reconstruction of the sampling of the greyscale Lena image by the superposition of the image pyramid individual resolution layers

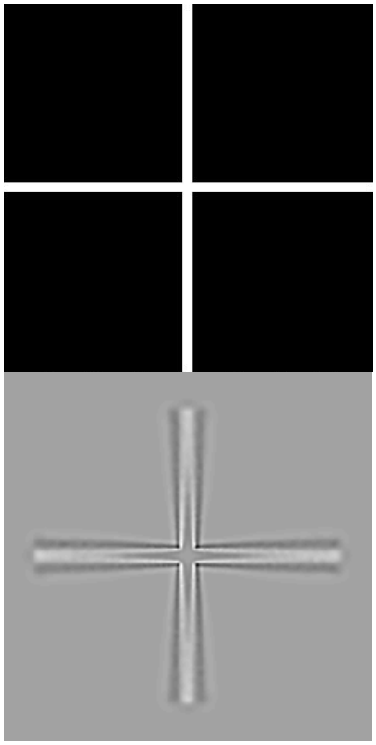

Figure 17: Input cross stimulus (top) and the reconstruction of the cross stimulus using a 4096 node retina with difference of Gaussian receptive fields (bottom).

Acknowledgements

The authors wish to acknowledge the support of the Department of Computing Science, University of Glasgow and the UK Imaging Faraday Partnership.

References

Barber, C. B., Dobkin, D. P. and Huhdanpaa H. T., The Quickhull Algorithm for Convex Hulls, *ACM Transactions on Mathematical Software*, Vol. 22, No. 4, Dec. 1996

Bolduc, M. and Levine, M. A review of biologically motivated space-variant data reduction models for robot vision. *Computer Vision and Image Understanding*, 69. 1998

Burt, P. J. and Adelson, E. H. The Laplacian pyramid as a compact image code, *IEEE Transactions on Communications*, pp. 532--540, April 1983

Clippingdale, S. and Wilson, R. Self-similar Neural Networks Based on a Kohonen Learning Rule. *Neural Networks*. Volume 9, Issue 5, pp. 747-763. 1996

Dudgeon, D. E. and Mersereau, R. M. Multidimensional Digital Signal Processing. *Prentice-Hall, Inc.,* Englewood-Cliffs, NJ, 1st edition. 1984

Hubel D. H. and Wiesel T. N. Brain mechanisms of vision. *Scientific American*. 241:150-162, 1979.

Kohonen, T. Self-organization and associative memory. *Berlin: Springer-Verlag.* 1984

Marr, D. Vision. *W. H. Freeman and C*o. 1982

Morgan, F. The hexagonal honeycomb conjecture. *Transactions of the American Mathematical Society* 351(May):1753. 1999

Polyak, S. L. The Retina. *University of Chicago Press*, Chicago. 1941

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

Schwartz, E. L. Computational Anatomy and functional architecture of the striate cortex, *Vision Research*, vol 20, pp. 645-669. 1980

Smeraldi, F., Makarov, A. and Bigun, J. Saccadic search with Gabor features applied to eye detection. Technical Report EPFL-DI-98/256, *Swiss Federal Institute of technology*, CH-1015 Lausanne, Jan 1998.

Thorpe, S. and Gautrais, J. Rapid Visual Processing using Spike Asynchrony. NIPS, Jordan M. editor, MIT press, pp 901-907. 1997

von der Malsburg, C. The correlation theory of brain function. Internal Report 81-2, *MPI Biophysical Chemistry*, Göttingen, Germany. 1981

Wallace, R., Ong, P., Bederson, B. and Schwartz, E. Space variant image processing. *Internationa Journal of Computer Vision*, 13(1):71--90, 1995.

Wilson, S. W. On the retino-cortical mapping. International *Journal of Man-Machine Studies*, 18(4):361-389. 1983