A low level vision hierarchy based on an irregularly sampled retina

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

Abstract

Biological vision systems process signals extracted by a retina that reduces the dimensionality of visual information for processing by higher cortical areas. To date researchers in machine vision have not reported computing a biologically inspired artificial retina that can sample visual information without over-sampling the central foveal region or creating discontinuities in the retinal tessellation. We have implemented a space-variant retina that has a uniform receptive field density in the central foveal region and becomes increasingly sparse in the surrounding periphery.

The retina contains isotropic receptive fields and the responses from the retina were processed by machinery motivated by biological cortical circuitry. We introduce a cortical graph that enables us to compute space-variant cortical filters that can sample irregular image information generated by the retina and thereby extract edge information and process colour opponent information. In this paper we briefly describe the self-organising methodology we used to generate retinal tessellations and detail the construction of visual information processing layers in our architecture.

1 Introduction

We are developing a generic vision front-end for attention and recognition which is based on a space-variant self-organised retina. The biologically inspired artificial retina extracts a visual information stream from the image which is then operated on by a hierarchy of units which are motivated by the processing in the human visual pathway.

In this paper we will refer to biological computational units as *cells* while machine computational units will be referred to as *filters* or *kernels*. The term *receptive field* will be used with both biological and machine computational units and will refer to the area in the field of view where stimulation results in a response in the computational unit.

The representation of visual information in a system which processes information from a space-variant retina is quite different from the representation used in conventional vision systems. Most current systems give equal processing emphasis to the whole field of view of the camera or image frame and work with visual information which can be stored in a uniform data structure. For example, greyscale information extracted by a conventional CCD imager in a digital camera can be stored in a rectilinear two dimensional array structure. Image processing operations for analysis and feature extraction can be easily applied to this array. Convolution operations are simple to implement by scanning a mask or kernel over the array and performing the necessary multiply and accumulate operations. Similarly the visual information can be easily sub-sampled to reduce its resolution. Rotation and other translation operations are also trivial. This simply is because the local connectivity between adjacent nodes of information in the array is uniform. Pixels have equidistant neighbours above, below and to their left and right (except on the border of the array).

However, space-variant imaging systems do not process the all visual information presented to the system equally. The central or foveal region of the retina has a very high acuity. The image is finely sampled by filters in this retinal region. As we increase eccentricity and move away from the central area of the retina, the acuity of the retina gradually reduces to the periphery where the image is only coarsely sampled. The output from this space-variant sampling needs to be stored in a plausible data structure for higher level operations in subsequent stages of the processing pathway of the system.

Researchers [1, 2, 3, 4] have tried to find an analytical *retino-cortical transform* that can map locations in the field of view to a continuous *cortical image*, thereby creating a data structure that can store the extracted visual information. This is because the change in ganglion cell density with eccentricity in a primate retinae resembles the density needed for analytical transforms such as the complex-log transform [1, 2]. However the actual retinal *tessellations* or locations of retinal receptive fields that are needed to generate a continuous cortical image using these retino-cortical transforms are inadequate, exhibiting singularities and over-sampling the fovea or having discontinuities and distortions in the sampling mosaic. No analytic approach or geometric mapping that can describe the gradual change in topography of the retina between a uniform fovea and space-variant periphery has been reported in the computer vision literature.

Therefore, we used a self-organisation methodology [5] to generate a retinal tessellation that while foregoing geometric regularity of the retinal mosaic had a continuity in sampling density. The structure of the retina (Figure 1) [6] locally resembles a pseudo-regular hexagonal lattice with sight deviations in the hexagonal topology in some locations while maintaining a sampling density continuum at a macroscopic level. The retina has a uniform foveal region which seamlessly coalesces into a space-variant periphery and the tessellation does not have a singularity in the fovea. In Figure 1 each point in the tessellation is a location for the placement of the centre of a receptive field.

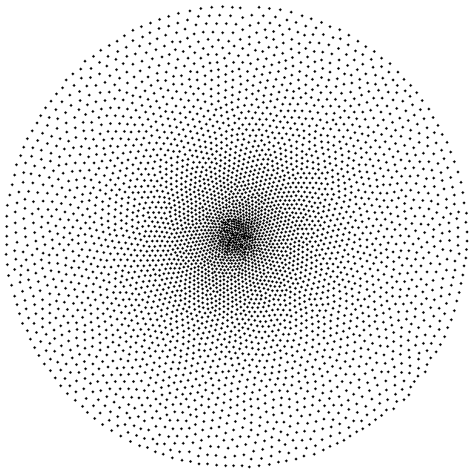

Figure 1: Self-organised retinal tessellation.

A space-variant retina that can sample an image was created by placing filters on the displayed retinal tessellation with receptive field size varying with the local node density of the tessellation. We placed simple *difference of Gaussian* filters[7] on the tessellation with sub-pixel accuracy. Difference of Gaussian filters were used because these isotropic bandpass filters resemble the receptive fields of retinal ganglion cells. The space-variant filter responses of the retina were extracted by multiplying the underlying image pixels with the (pre-computed) filter coefficients. The retina displayed in Figure 1 has 4096 nodes and therefore filters placed on this retina would result in a sampling output of a 4096 by 1 vector. Because an explicit analytic mapping from the retina to a retinotopic cortical image data structure that could be used to store and manipulate extracted image information is not available for our self-organised retina, we had to devise a way of performing filtering operations on the visual data which is in effect a one dimensional vector.

In the human visual pathway retinal ganglion cells initially perform filtering with isotropic centre-surround receptive fields. Following this, nerve afferents carry visual information away from the retina and the achromatic information in the parvocellular pathway [8] (which carries information related to form, colour and texture) is processed by simple cells in the lower visual cortex. These simple cells [9] have been found to have anisotropic receptive fields. The receptive fields are elongated and are thought to be used to extract edge information from the isotropic responses emitted from the retina. The simple cells have oriented receptive fields at different orientations and scales and perform a great deal of processing on the visual information from the retina. It is hypothesised that anisotropic filtering of isotropic responses has evolved in nature because of computational efficiency. This is because the space-variant filtering computed by the isotropic layer dramatically reduces the dimensionality of the visual information so that it can be efficiently operated on by a multitude of oriented anisotropic filters.

It is thought that chromatic information in the parvocellular pathway is processed by double opponent cells [10]. These are circularly symmetric centre-surround cells found in the “blob” regions of the primary visual cortex and are believed to help provide colour constancy to human vision.

The main contribution of this paper is the approach for implementing anisotropic sampling of the achromatic responses of isotropic filters placed on the irregular, pseudo-random retinal tessellation. We will detail the construction of a space-variant visual processing hierarchy based on a retinal sampling with a pseudo-random receptive field tessellation.

Similarly a multi-scale hierarchy using retinae with 4096, 1024, 256, 64 and 16 receptive fields, creating a space-variant pseudo-random approximation to an octave pyramid was implemented.

2 Related Work

Laplacian pyramids [11] have been traditionally used in image processing to enhance salient image features. These can detect activity in an image at different spatial scales. Researchers frequently approximate the Laplacian operator by a difference of Gaussian kernel [12].

Greenspan et. al. [13] constructed an oriented Laplacian pyramid by forming a Filter-Subtract-Decimate Laplacian pyramid and modulating each level of the pyramid with oriented sine waves. They used this structure for rotation invariant texture classification.

The data structures that were used for research discussed in this section have thus for been conventional rectilinear arrays. Wallace et. al. [14] considered image processing using space-variant structures. They used *connectivity graphs* to encode relations between nodes, where graph nodes represent sensor pixels and graph edges represent adjacency relations between pixels. This work dealt with retinal tessellations and sensors based on

analytical retino-cortical transforms [1, 2]. They formed cortical image data structures to store the extracted visual information. Wallace et. al. performed image transformations, pyramid operations and connected components analysis on the space-variant cortical images. They also performed simple edge detection by subtracting the pixel value of adjacent nodes from the pixel value of a node.

Smeraldi and Bigun [15] have developed a facial landmark detection and face authentication system based on low-level features extracted using Gabor filters placed on a retina-like sampling grid. They used SVM classifiers to detect facial landmarks. The search for facial landmarks was conducted by centring their retina on the sampling point that resulted the in a local maximum of SVM output. This appears to be the most complete attempt where an active space-variant retina has been used for a vision task that is represented in the literature to date. However the retina Smeraldi and Bigun used contained just 50 receptive fields. They did not develop a vision hierarchy and the steering of anisotropic (Gabor) filters and other complex filters on the retina itself is inefficient. In this paper we detail the construction of a vision hierarchy that begins with elementary band-pass filtering in a multi-scale detailed retina and extends to orientated edge and colour detecting filters in higher levels of the processing pathway.

3 The self-organised retina

The pseudo-random retinal tessellation that we adopted was obtained by a self-organisation technique [5] that is based on stimulating the network weights by a stimulatory input that is derived by applying a composite transformation to the network weights themselves. When generating retinae, network weights will represent the two dimensional x and y coordinates of the retinal receptive fields. Therefore, for a network of N units, each characterised by a 2 dimensional weight vector $x_i(n)$, the input stimulus $y_i(n)$ at iteration n is calculated by the following,

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

where $x_i\,(n - 1)$ is the i th network unit at iteration n - 1 and $1 \le i \le N$. In our work we used the T composite transformation (equation 1) which comprises of a random rotation between 0 and 2π, a dilation (increase in eccentricity) comprising of the exponent of a dilation factor which is random between 0 and log(8) and translations in the vertical, horizontal and radial (away from centre) directions random between 0 and f, where f is associated with the required foveal percentage of the resultant retina.

The network was initialised with a random weight configuration and iterated with the described composite transformation T and the following learning rule to find the updated weight vector $x_j(n)$:

$$x_j(n) = x_j(n-1) + \alpha(n) \sum_{i \in \Lambda_j(n)} \left(y_i(n) - x_j(n-1) \right) \tag{2}$$

where,

$$\Lambda_j(n) = \left\{ i : \left\| y_i(n) - x_j(n-1) \right\| < \left\| y_i(n) - x_k(n-1) \right\|, k \neq j \right\} \tag{3}$$

$\Lambda_j(n)$ 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. Space limitation restrict our discussion of the self-organisation further, but the reader is referred to Balasuriya and Siebert [6] for further details regarding generating retinae and to Clippingdale and Wilson [5] for details regarding the self-organisation methodology.

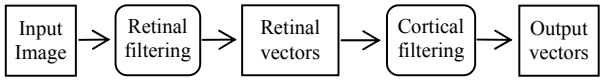

Figure 2 : Diagrammatic representation of the implemented low level vision hierarchy

3.1 Receptive fields on the retina

Overlapping receptive fields were placed on the self-organis retina to retinally sample an image. The receptive fields were placed with sub-pixel accuracy by varying the filter coefficients to reflect the sub-pixel jitter.

Receptive field sizes had to change with the eccentricity of the retina to generate a space-variant sampling. Filters in the periphery had to be much larger than those in the fovea. Since an analytic transform for the self-organised retina that gave explicit receptive field locations was unavailable, a strategy to define receptive field sizes on the irregularly tessellated retina was formulated. Retinal receptive field size was based on local node (receptive field centre) density and the following metric was used to determine the receptive field diameter d_i of retinal node i :

$$d_i = \frac{s}{k} \sum_{j=2}^{k} A_{i,j} \tag{4}$$

where $\boldsymbol{A}_{i,j}$ is the sorted Euclidean distance matrix of the retinal tessellation, k is the neighbourhood size for determining node density and s is a scaling constant. Because the receptive field size varies with what is essentially local node density, the receptive fields will be space-variant with retinal eccentricity.

Since the circularly symmetric Gaussian distribution g(x,y) with two variables and standard deviation σ is

$$g(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (5)$$

the difference of Gaussian filter was calculated by

$$dog(x,y) = \frac{1}{2\pi\sigma_s^2} e^{-(x^2+y^2)/2\sigma_s^2} - \frac{1}{2\pi\sigma_l^2} e^{-(x^2+y^2)/2\sigma_l^2} \quad (6)$$

where σ_s and σ_l are the standard deviations of the two Gaussian distributions that form the difference of Gaussians kernel. The following ratio between standard deviations was used

$$\frac{\sigma_l}{\sigma_s} = 1.6 \quad (7)$$

to make the difference of Gaussians kernel approximate the Laplacian which in turn resembles the receptive fields of biological retinal ganglion cells [12].

Since a cortical image to display the output of the retina does not exist we visualised the sampling and processing of the system by reversing the processing operations to generate a *reconstruction* of the extracted responses in the image domain. These image domain responses of machine units may be considered analogous to the receptive fields of biological neurons found by neuroscientists like Hubel and Wiesel.

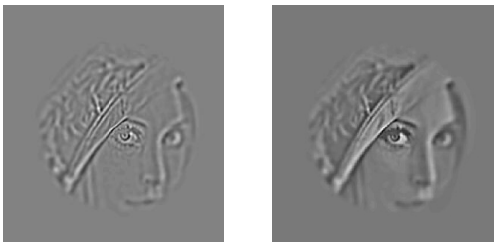

Figure 3: (Left) Reconstruction using a 4096 node retina (Right) Superposition of the reconstructed responses after sampling using a retina pyramid

Figure 3 contains the reconstruction of retinally sampling the grayscale Lena image using standard difference of Gaussian retinal receptive fields. The image on the right was constructed by sampling the Lena image with retinae with 4096, 1024, 256, 64 and 16 difference of Gaussian receptive fields and then reversing the process and summing the reconstructed images. Close examination of the images will reveal the space-variant nature of the retinal sampling. The reader is referred to Balasuriya and Siebert [6] for further details.

In the human retina, chromatic information is encoded into red-green and blue-yellow colour opponent channels. This is done by type I and type II colour opponent retinal ganglion cells [16]. The processing of type II cells was approximated in machine vision by taking the difference between the responses of Gaussian filters which sampled the red and green channels separately. Chromatic filters were placed on the retina with overlapping receptive fields as we discussed earlier.

The following figure illustrates the reconstruction of the responses of our type II filters on a 4096 node retina. Space-variant colour opponent receptive fields have extracted chromatic contrast information from the standard mandrill image. Our implementation of type I filters resulted in similar responses.

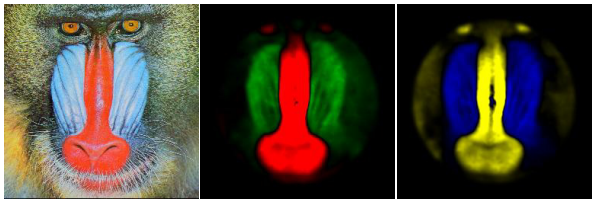

Figure 4: Original mandrill image with reconstructed responses from red-green and blue yellow type II colour opponent retinal filters respectively.

4 Processing an irregular tessellation

The retinal output that we have described needs to be processed by filters in higher levels of the visual hierarchy. These filters will extract edge information and analyse colour information from the responses of the self-organised retina.

Since the output of the retina is essentially a one dimensional vector, applying filtering operations on this data structure is not trivial. While the extracted feature vector is one dimensional, each location on the vector has a spatial semantic relationship with a corresponding location on the retinal tessellation (Figure 1). The problem of applying a filtering operation on the vector was addressed by calculating the support region and coefficients of the required filter at the location on the retinal tessellation that was associated with a position on the vector.

We used the Quickhull algorithm [17] to perform Delaunay triangulation on coordinates of the generated retinal receptive field centres (retinal tessellation). This enabled us to define a *cortical graph* that would help us reason with the extracted visual information. Filter support regions for cortical filters using constant kernel sizes on the cortical graph results in space-variant cortical filter support regions on the image. The unit of distance was an *edge* between two nodes (associated with retinal receptive field responses) in the Delaunay triangulated tessellation (cortical graph).

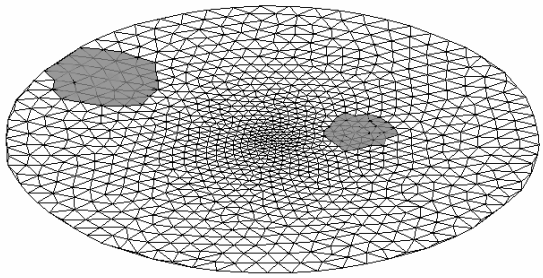

Figure 5 : The Delaunay triangulated cortical graph with two cortical filters highlighted to indicate space-variant filter support regions in the image domain

The radius of the cortical filters highlighted in Figure 5 is a constant three edges on the cortical graph, but these have varying support regions on the image because of the space-variant tessellation of the cortical graph. The following was defined to calculate the neighbourhood $N_k(v_c)$ of a cortical filter with radius k edges on the cortical graph and centred on node v_c.

$$v_i \in N_k(v_c) \text{ when } dist(v_i, v_c) < k \quad (8)$$

where $dist(v_i,v_c)$ is the graph distance or length of the graph geodesic (shortest path) along the cortical graph from node v_i to node v_c. Coefficients need to be calculated at nodes v_i for the cortical filter. While cortical filter neighbourhoods were defined on the cortical graph, the coefficients of the filters were calculated based on the normalised Euclidean displacement of the nodes in image space. This was more accurate than computing the coefficients of cortical filters based on the displacement of nodes on the cortical graph. Even in biology, ontogenesis will cause cortical filters to adapt to ideal receptive fields in the field of view and will not be solely affected by local inter-cortical distances between units.

4.1 Anisotropic Orientated cortical filter layer

A Gabor filter $h(x,y)$ can be described as a sinusoidal plane (at a certain frequency and orientation) modulated by a Gaussian envelope. Gabor filter coefficients at the discrete locations of the cortical graph within a cortical filter's spatial support need to be calculated. For a cortical Gabor filter of size k centred at $v_c(x_0, y_0)$ on the cortical graph the valid filter coefficients will lie on $v_i(x,y)$ in neighbourhood $N_k(v_c)$ (equation 8) where N is the set of neighbourhood points around (and including) point $v_c(x_0, y_0)$. Therefore, the Gabor filter coefficients for the cortical filter at $v_c(x_0, y_0)$ are given by

$$h(x_o, y_o) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_o)^2+(y-y_o)^2}{2\sigma^2}} e^{-j2\pi(U(x-x_o)+V(y-y_o))} \quad (9)$$

where σ is the standard deviation of the Gaussian envelope and U and V are the horizontal and vertical central frequencies of the Gabor filter.

Because of the irregular positions of the nodes on the cortical graph, coefficients may be biased towards one of the two cortical filter subfields of the Gabor filter. For example there may be more cortical nodes on the positive subfield of the filter. This would result in a biased filter response and the cortical filter would even give a response to a uniform non-zero input. Therefore the two cortical filter subfields were balanced by normalising the coefficients on the positive and negative subfields.

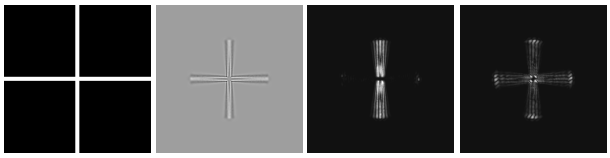

Figure 6 : Reconstruction after filtering of a cross stimulus (left) using difference of Gaussian (centre-left) and Gabor cortical filters oriented vertically (centre-right) and at 45º (right).

4.2 Double Opponent cortical filter layer

For a double opponent cortical filter of size k centred at $v_c(x_0, y_0)$ on the cortical graph the valid filter coefficients will lie on $v_i(x, y)$ in neighbourhood $N_k(v_c)$ (equation 8) where N is the set of neighbourhood points around (and including) point $v_c(x_0, y_0)$. We may define a filtering operation $\otimes$ for the double opponent cortical cell $dop(x_0,y_0)$ on cortical graph node $v_c(x_0, y_0)$ as the following:

$$dop(x_o, y_o) = dog((x - x_o), (y - y_o)) \otimes I(x, y) \quad (10)$$

where dog is a difference of Gaussians filter (equation 6) and $I(x,y)$ are the retinal responses from the Type II colour opponent retinal filters. The following must also be defined for the double opponent cortical filter:

$$\begin{aligned} I(x, y) &\in I_1(x, y),\ dog(x - x_o, y - y_o) < 0 \\ I(x, y) &\in I_2(x, y),\ dog(x - x_o, y - y_o) \geq 0 \end{aligned} \quad (11)$$

Different inputs are sampled in the difference of Gaussian's positive and negative subfields. For example, for a double opponent filter sampling the output of the red-green opponent filters, I_1 would be responses from the type II Red^+Green^- and I_2 would be responses from the type II Red^-Green^+ colour opponent retinal filters[10]. As with the Gabor cortical filters, the filter subfields were normalised.

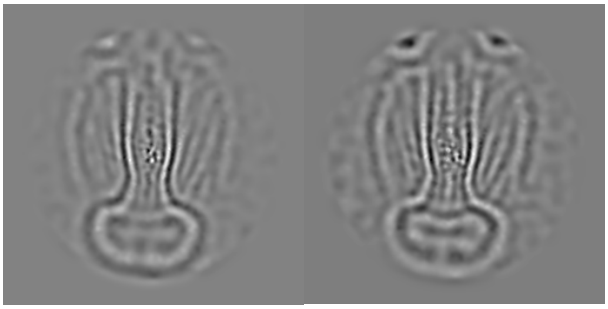

Figure 7 : Responses of red-green (left) and blue-yellow (right) double opponent cortical filters displayed using difference of Gaussian filters.

We hypothesise that the output of the double-opponent cells should be analysed for edge information as there are distinct edge-like responses in this output. Little guidance is available in the psychophysics and neuroscience literature regarding the processing of responses from double opponent cells by higher cortical areas.

5 Conclusion

We have shown that a biologically motivated low level vision hierarchy can be constructed that processed the responses from a self-organised space-variant irregularly sampled retina. A cortical graph was introduced which enabled us to define cortical filters that had receptive fields with space-variant spatial support regions. The coefficients for cortical filters were calculated for nodes within the spatial support of the filter in the cortical graph. The actual filter coefficients were calculated based on displacement on the image or field of view. The resulting vision hierarchy was able to extract edge and colour information using these cortical filters and could be extended to higher layers of processing to extract complex features using the same methodology.

Our current work focuses on implementing an object recognition engine that will process the output vectors of the vision hierarchy and drive a saccadic attention mechanism that will generate salient points for fixation by the retina.

Acknowledgements

The authors gratefully acknowledge the support of the University of Glasgow and the UK Imaging Faraday Partnership.

References

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

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

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

[4] H. Gomes, *Model Learning in Iconic Vision*. PhD Thesis, University of Edinburgh. 2002.

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

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

[7] D. Marr, *Vision*, W. H. Freeman and Co, 1982.

[8] M.S. Livingstone and D.H. Hubel, Segregation of form, color, movement, and depth: Anatomy, physiology, and perception, *Science*, 240: pp. 740-749, 1988.

[9] D.H. Hubel and T.N. Wiesel, Receptive fields of single neurons in the cat's striate cortex, *Journal of Physiology*, 148: pp. 574–591, 1959.

[10] D.H. Hubel, *Eye, Brain and Vision*, Scientific American Library, 1987.

[11] P.J. Burt and E.H. Adelson, The Laplacian Pyramid as a Compact Image Code, *IEEE Transactions on Communications*, 31(4): pp. 532-540, 1983.

[12] D. Marr and E. Hildreth, Theory of edge detection, *Proceedings of the Royal Society of London*, B(207): pp. 187-217, 1980.

[13] H. Greenspan, S. Belongie, P. Perona, R. Goodman, S. Rakshit, and C.H. Anderson, Overcomplete steerable pyramid filters and rotation invariance, in *CVPR*, 1994.

[14] R.S. Wallace, P.W. Ong, B.B. Bederson, and E.L. Schwartz, Space-Variant Image-Processing, *International Journal of Computer Vision*, 13(1): pp. 71-90, 1994.

[15] F. Smeraldi and J. Bigun, Retinal vision applied to facial features detection and face authentication, *Pattern Recognition Letters*, 23: pp. 463 - 475, 2002.

[16] T.N. Wiesel and D.H. Hubel, Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey, *Journal of Neurophysiology*, 29: pp. 1115-1156, 1966.

[17] C.B. Barber, D.P. Dobkin, and H. Huhdanpaa, The quickhull algorithm for convex hulls, *ACM Transactions on Mathematical Software*, 22(4): pp. 469 - 483, 1996.