Computational Simulation of Depth Perception in the Mammalian Visual System

Jesse S. Jin

The School of Computer Science & Engineering
University of New South Wales, NSW 2033, Australia
E-mail: jesse@cse.unsw.edu.au

Abstract

This paper presents a computational model for stereopsis. Laplacian of Gaussian filters are used to simulate ganglion cells and LGN cells and zero-crossings extracted provide spatial features in the visual scene. A set of one-octave Gabor filters is used to extract orientation information, which cover 0 to 60 cycles/degree interval in the human visual system. A Gaussian sphere model is used to map a 3D space onto two 2D image planes, which combines monocular cues with binocular cues in stereo matching. The determinant of the Jacobian of the mapping is derived and matching is performed using zero-crossings associated with their orientation information. The possibility of transferring the knowledge such as the probability of occurrence of visual scenes to the matching process from the mapping is discussed. Relaxation labelling is used as a co-operative process, which simulates binocular fusion and rivalry in the human visual process.

Introduction

Human vision, or the visual system of any vertebrate consists of three major sections, as shown in Figure 1: the photoreceptors in the eyes which capture light and generate messages about that light; the visual pathways (including the lateral geniculate nucleus or LGN) which transmit those messages from the eye; and the visual cortex which interprets the messages in various ways. Stereopsis is a major source of depth perception in the mammalian visual system. Matching stereoscopic views is an important step in stereo calculation. Different methods could be used to recover (relative) depth information from stereo, and its particular choice depends on *features* and *the stereo model* used in the matching process.

It is worth mentioning several significant achievements in understanding the human stereopsis: the response of the receptive fields including ganglion cells, LGN cells (Hartline, 1938), simple cells, complex cells and hyper complex cells (Hubel & Wiesel, 1962), spatial frequency and contrast of visual gratings (Campbell & Robson, 1968), multi-channel sensitivity (Wilson & Bergen, 1979), ransom-dot stereogram (Julesz, 1960), and fusion and rivalry of binocular cells (Blake & Camisa, 1979). Computational vision research began when AI research diversified in the early 70's, but it was not until Marr's (1982) work in the mid-70's that computational vision research began to make extensive use of findings from biological systems. The most significant achievements in Computational Vision include Laplacian of Gaussian (LoG) filtering (Marr & Hildreth, 1980), Gabor filtering (Wilson, 1983; Daugman, 1980), binocular fusion using neural network (Grossberg, 1987), and stereo models (Grimson, 1981; Trivedi, 1985).

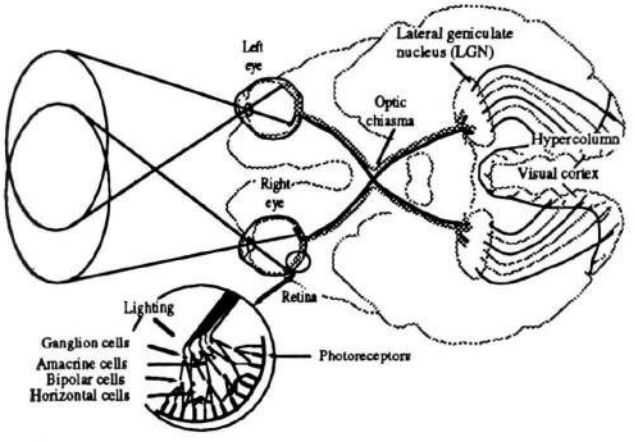


Figure 1: The layout of a mammalian visual system.

The famous random-dot stereograms invented by Julesz have been used to show convincingly that the calculation of stereo disparity (in humans) is not based on monocularly recognisable forms such as a familiar face. Another intriguing aspect of binocular vision which has long been observed is binocular rivalry (Wheatstone, 1838), the alternating periods of dominance and suppression occasioned by stimulation of corresponding retinal areas with dissimilar monocular stimuli. Although there has been much empirical study of this phenomenon since then, only a few major theoretical developments have been made in stereo matching concerning binocular rivalry.

We developed a stereo model to simulate visual processing in the human visual system. Sections 2 describes how receptive fields are simulated using LoG filters and Section 3 discusses extracting orientation information using Gabor filters. A mapping from the three-dimensional space to two stereoscopic views is derived for stereo matching in Section 4. The paper concludes in Section 5 with a discussion on expanding the model.

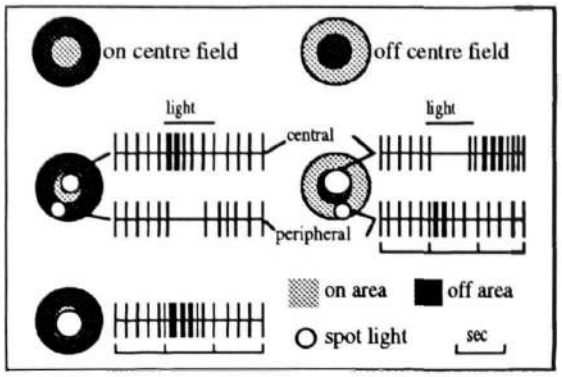


Figure 2: Receptive fields of ganglion cells and their responses to different stimuli.

Receptive Fields and LoG Filtering

The concept of the receptive field is important in neurophysiology. Hartline (1938) found that receptive field organisation exists in the frog retina and in the optic pathways of the cat. Many visual receptive fields have circular organisations which are described as on-centre/off-surround fields, or the opposite, off-centre/on-surround fields, as shown in Figure 2.

The size of the receptive field decides the response of a ganglion cell to the frequency of a stimulus. Campbell and Robson (1968) discovered that the human visual system contains a number of different mechanisms selectively tuned to respond to different bands of light (or spatial frequencies) and that these mechanisms operate in parallel in the processing of spatial information. The unit employed to express spatial frequency is the number of cycles that fall within one degree of visual angle (each cycle is one sinusoidal period). Wilson and Bergen (1979), studied the human visual system in the range 0.25-16.0 cycles/degree and discovered four sensitive peaks, N, S, T and U, which occurred at about 19.35, 9.68, 5.13 and 2.86 cycles/degree, respectively. Later Marr et al. (1980) proposed from the psychophysical data on two-point and line acuity that the smaller foveal channel in human vision must have an excitatory centre with a diameter of around 1.33′, i.e. 45 cycles/degree.

On the basis of Campbell and Robson's work, Marr and Poggio (1979) first used a difference of two Gaussian filters to detect zero-crossings but later Marr and Hildreth (1980) suggested a Laplacian of a Gaussian function (LoG), as shown in Figure 3 (a). Using this basic function, they simulated the multiple channels in the mammalian visual system, as shown in Figure 3 (b).

LoG function is defined as:

$$\nabla^2 G(x, y) = \frac{1}{2\pi\sigma^4} \left(2 - \frac{x^2+y^2}{\sigma^2}\right) \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

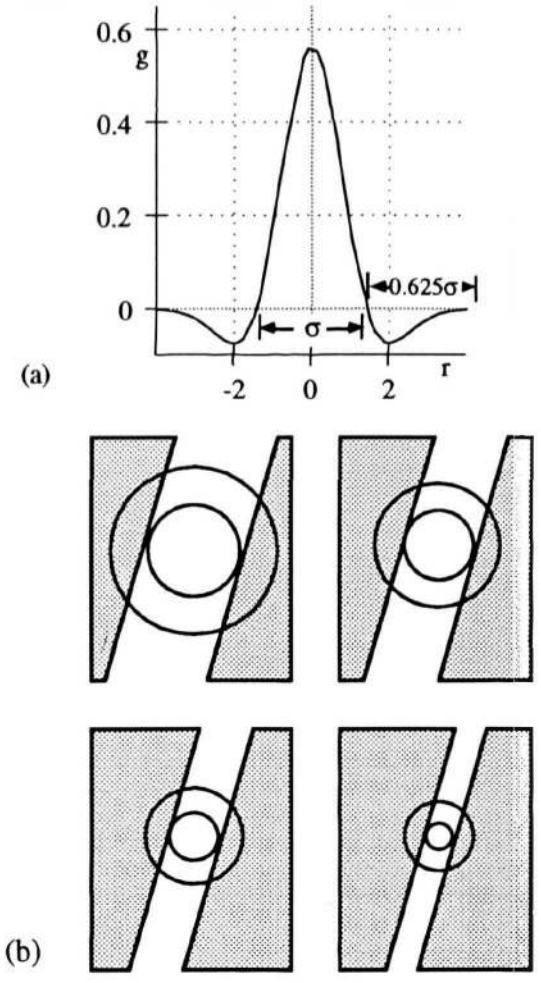


Figure 3: LoG function (a) and multi-channel detection (b).

Different σ values detect intensity changes at different scales. In an image with a viewing angle of 9° and resolution of 512×512 pixels, we propose five channels with σ values of 3, 5, 9, 17 and 35 pixels, respectively. The frequency bandwidth of the five channels is shown in Figure 4 (a). The central excitatory region of each channel is at about 1.58′, 2.64′, 4.75′, 8.96′ and 18.46′, and the centre of each channel is at about 37.93, 22.76, 12.64, 6.69 and 3.25 cycles/degree respectively. For comparison, Wilson and Bergen's (1979) results are shown in Figure 4 (b) and the channel with the smallest visual angle comes from Marr's results. Both axes are displayed on a logarithmic scale.

The zero-crossings extracted from five channels are overlaid on the original image shown in Figure 5.

Orientation Selectivity and Gabor Filtering

Although zero-crossings give a good localisation of intensity changes in an image, they are not the only features computed in early vision (Torre & Poggio, 1986). Worse,

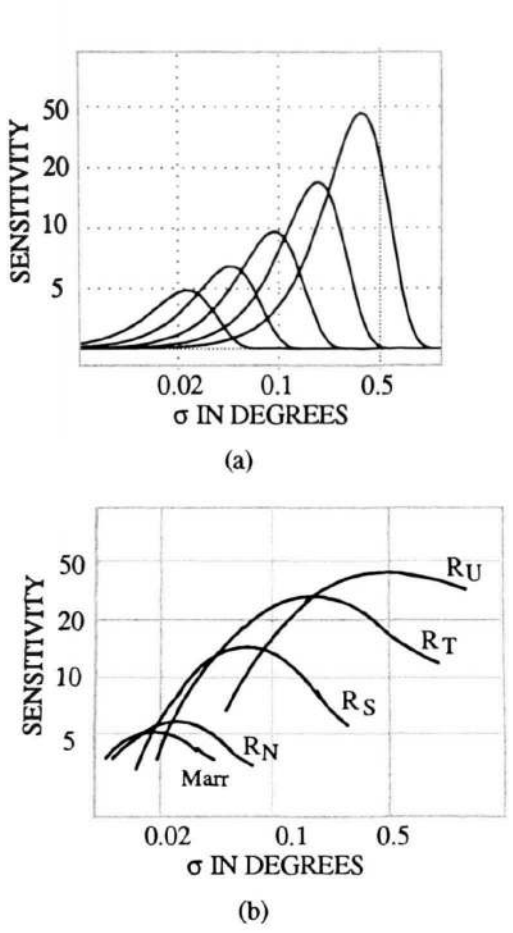


Figure 4: Frequency bandwidth of the multiple channels.

Daugman (1988) had found that some simple information processing operations which are apparent in pattern perception of human vision are impossible in a representation of zero-crossings. Neurophysiological studies have proposed that cells in the visual cortex fire in response to phase, frequency and orientation. Hubel and Wiesel (1962) succeeded in recording the electrical responses of living cells in the visual cortex of the cat and the monkey to various patterns of stimulation. They discovered that the receptive fields in the cat's visual cortex, unlike the simple, circularly organised receptive fields found previously in the retina and lateral geniculate body, are thinner and more elongated in shape. They respond to the presence of contours having a particular orientation. Figure 6 sums up responses of neurons to different light patterns. Marcelja (1980), Daugman (1980), and Kulikowski et al. (1982), among others, have suggested the use of a Gabor function to model this part of visual processing.

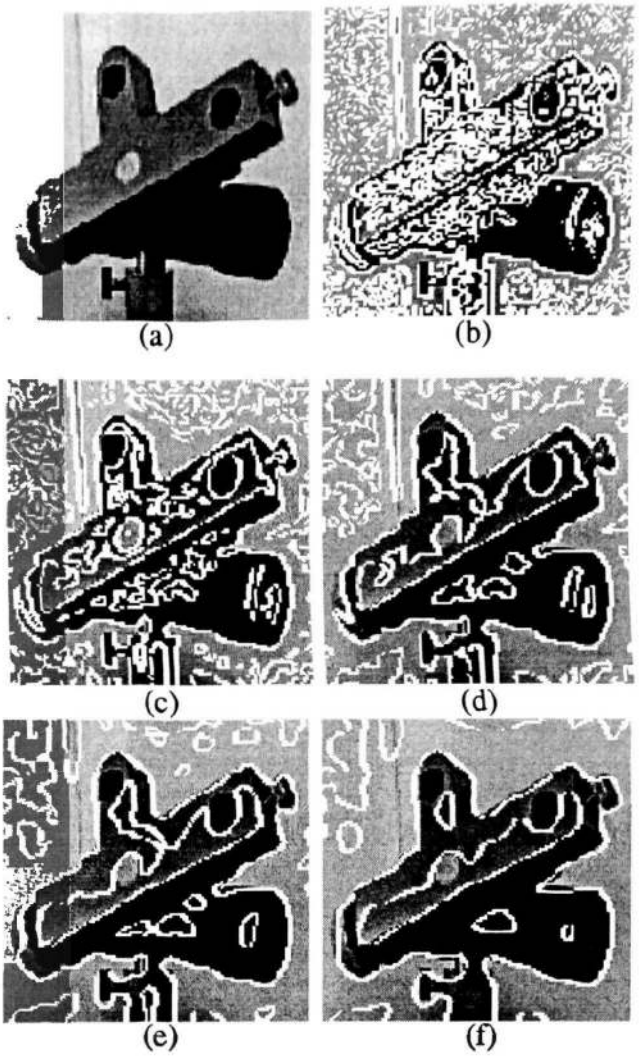


Figure 5: A bracket (a) and its zero-crossings from five channels (b)-(f).

The general form of the 2D Gabor function is given by:

$$g(x, y) = \exp\{-\pi[(x-x_0)^2a^2+(y-y_0)^2b^2]\}\exp\{-2\pi i[u_0(x-x_0)+v_0(y-y_0)]\}$$

with a Fourier transform:

$$G(u, v) = \exp\{-\frac{1}{\pi}[\frac{(u-u_0)^2}{a^2}+\frac{(v-v_0)^2}{b^2}]\}\exp\{-2\pi i[x_0(u-u_0)+y_0(v-v_0)]\}$$

We developed a set of eight one-octave Gabor filters with centroid frequencies located at 0.25, 0.5, 1, 2, 4, 8, 16 and 32 cycles/degree respectively, as shown in Figure 7.

Figure 8 (a)-(e) give filtering results of two stereoscopic views (see Figure 10a) in different orientations, and (f) shows features in a group.

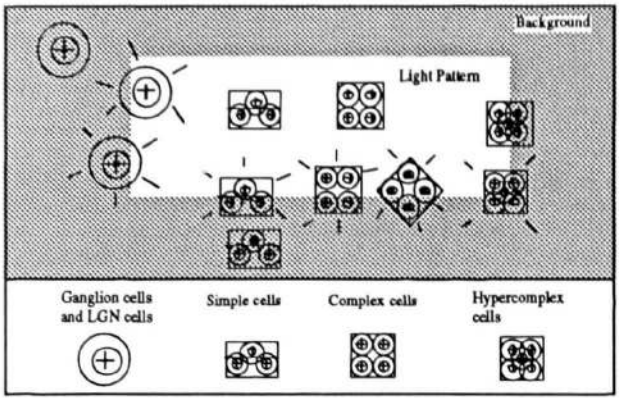


Figure 6: Responses of neurons to a pattern.

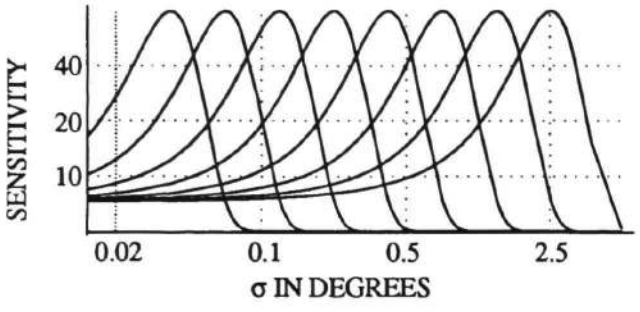


Figure 7: Sensitivity of multi-channel Gabor filters.

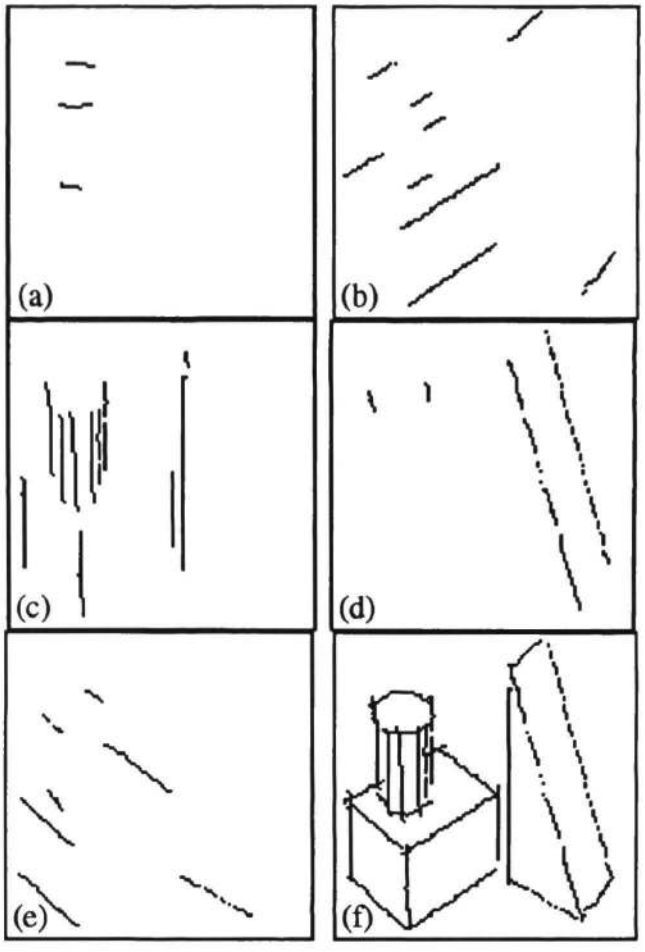


Figure 8: Feature extraction using Gabor filtering.

Matching Stereo Views and Cooperative Process

Julesz (1971) gave another random-dot stereogram in which one of the images is expanded by 15%. Stereopsis can still be easily obtained which suggests that spatial features are not the sole source for matching. Some other information, and particularly binocular arrangement, is important for the eyes to perform stereo matching. To understand the binocular arrangement, we have to know hypercolumns. Frisby (1980) noted that the visual cortex appears to be composed of columns of cells, with each column consisting of a stack of cells all preferring the same orientation. It takes roughly eighteen to twenty neighbouring columns to cover a complete range of stimulus orientations. This aggregation of adjacent columns is collectively known as a hypercolumn. It has been found the binocular specialisation of receptive fields. Such fields are not necessarily in exactly corresponding points in the two retinas. Neurons whose performance fits them for depth perception require a binocular stimulus either in front of, behind, or in the plane of fixation (Kuffler & Nicholls, 1984). Fusion is not the all activity of binocular neurons. Kaufman (1964) raised the idea that rivalry suppression underlies ordinary binocular vision. A class of cooperative mechanisms exists in human stereopsis (Marr & Poggio, 1976).

To simulate stereo matching and cooperative processing in the human visual system we developed a stereo model using zero-crossings and their orientations. However, one problem in stereo matching using both features is that they each require a different coordinate system. Consequently, the final disparities obtained depend critically upon the scale used in the measurement. This problem is overcome by using probabilities.

Let the vision space S be: $X \times Y \times Z$ with X, Y and $Z \subseteq R$. Consider an edge of an object passing through a point (x, y, z). If we represent this edge as an oriented vector in 3D space, it has an angle θ with the x axis in x-y plane and an angle φ with the z axis. By using these two angles, the edge can also be represented as a point on the surface of a unit sphere, whose origin is (x, y, z). This is known as the Gaussian sphere (Arnold & Binford, 1980), and any point located on its surface is defined in terms of its spherical coordinates θ and φ (see Figure 9). The Gaussian sphere thus defines a mapping $(\Delta x, \Delta y, \Delta z) \rightarrow (\theta, \varphi)$.

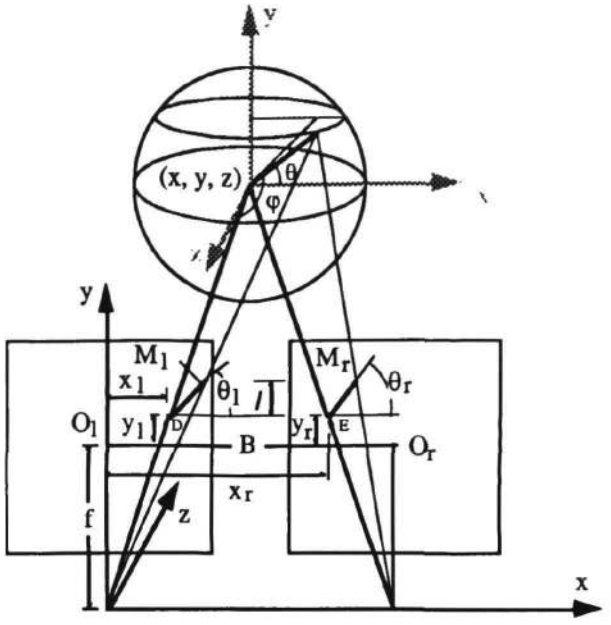


Figure 9 : The Gaussian sphere model for matching.

Given a corresponding pair of edges, one in each image, as shown in Figure 9, we are interested in how their angles are related and how we can use this relationship to guide our matching process. Although the angles θ_l and θ_r could be of any values, they are usually of fairly similar values. This is partly due to a moderate or a small offset of the baseline. The matching process is to find the corresponding points in the left and right image and the more clues to guide the search the better it would be. We know that a continuous function Q exists for mapping the points on the Gaussian sphere to the image angles (θ_l, θ_r), i.e. $\theta\times\varphi \rightarrow \theta_l\times\theta_r$, . More importantly, however, there exists also an inverse function P which maps points in the space $\theta_l\times\theta_r$ to points on the Gaussian sphere, $\theta\times\varphi$. From the probability theorem, the probability distribution of (θ_l, θ_r) equals the probability distribution of (θ, φ) multiplying with the Jacobian determinant of the mapping P. If we assume all edges of objects are randomly and uniformly distributed in the (θ, φ) domain, the probability distribution ψ of (θ_l, θ_r) will be $\psi(\theta_l, \theta_r) = \frac{1}{A}|Jp|$, where A is the area of the definition domain Ω of (θ, φ). When the visual distance z is far enough comparing with baseline B, i.e. $B/z \ll 1$, we have the determinant of the Jacobian matrix defined as:

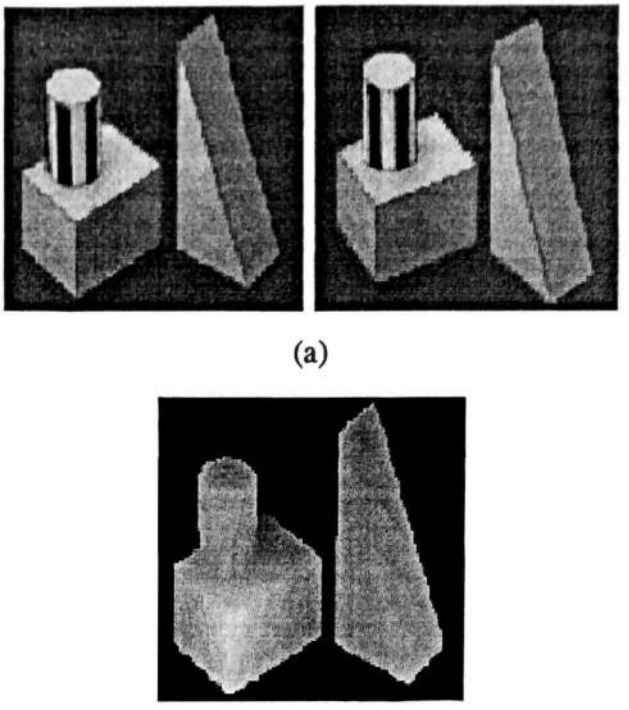
$$|Jp| \approx \frac{y_l\,[(x_r - x_l) - B\cos^2\theta_l]}{(x_l^2 + y_l^2 + 1)\sqrt{x_l^2 + y_l^2}\,\sin^2(\theta_l - \theta_r)}$$

The detailed deduction can be found in (Jin, 1992). This distribution gives a correlation function for θ_l and θ_r. It is noteworthy that the mapping P is not a bijective mapping. It is not defined at (0, 0), as the circle $z = 0$ of points on the sphere for which $\theta = 0$ all map to (0, 0). The mapping is not invertible at that point, which is why we use P to represent the mapping $\theta_l\times\theta_r \rightarrow \theta\times\varphi$ rather than Q^{-1}. This fact tallies with the effect in human vision. When people view a horizontal wire, they often lose their depth perception. This is because the uniform texture on the wire wipes out the size perception so that the stereo matching depends solely on orientation, but the zero orientations in both eyes fail to stimulate binocular neurons to cause fusion. A cooperative process is needed to reflect the fusion and rivalry process in the human visual cortex. We use relaxation labelling, which can be represented as

$$p_i^{(k+1)}(\theta_l) = \frac{p_i^{(k)}(\theta_l)[1+q_i^{(k)}(\theta_l)]}{\sum_{\theta_l} p_i^{(k)}(\theta_l)[1+q_i^{(k)}(\theta_l)]}$$

where $q_i^{(k)}(\theta_l) = \sum_j \sum_{\theta_r} r_{ij}(\theta_l\,\theta_r)P_j(\theta_r)$, $r_{ij}(\theta_l\,\theta_r) = |Jp|$, and $P_j(\theta_r)$ are constants equal to one of the number of lines in the right view.

The model has been used successfully in matching stereoscopic views, shown in Figure 10 (a), and depth is recovered and shown in intensity in Figure 10 (b).



(a)

(b)

Figure 10: Stereoscopic views (a) and recovered depth from stereo calculation (b).

Discussion and Conclusion

The significance of our model is that it defines a relation between the visual world and the two stereoscopic views. The mapping as defined allows us to manipulate the model in various ways and reflect several characteristics of stereo vision in humans. First, any a priori knowledge about the world, either from our knowledge of the visual scenes or from the features extracted from each stereoscopic view, can be applied in the mapping. Second, we can adjust the focus of the view point either to improve the success rate of the stereo matching or increase accuracy of stereo calculation. The distribution of the determinant of the Jacobian varies with $x_l^2+y_l^2$ (i.e. concentrically). Close to the centre, we have a steep distribution along $\theta_l = \theta_r$ which gives more weight for matching the stereo than that for calculating stereo disparity, and vice versa.

The current model has difficulty in dealing with complicated visual scenes. The relaxation labelling process takes a long time to converge and may even fail to converge when the number of similar features is large. The use of different kinds of information could help but another solution is to decompose the (global) computation into many local computations using locally activated regions of neurons. Our suggestion is a network using the Radial Basis Functions (Jin et al., 1993).

References

Arnold, R D & Binford, T O (1980). Geometric constraints in stereo vision. *Proc. SPIE: Image Processing for Missile Guidance* 238, 281-292.

Blake, R & Camisa, J C (1979). On the inhibitory nature of binocular rivalry suppression. *J. Experimental Psychology: Human Perception and Performance* 5, 315-323.

Campbell, F W & Robson, J G (1968). Application of Fourier analysis to the visibility of gratings. *J. Physiology* 197, 551-556.

Daugman, J G (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research* 20, 847-856.

Daugman, J G (1988). Pattern and motion vision without Laplacian zero crossings. *J. Opt. Soc. Am. A* 5, 1142-1148.

DeValois, R L; Albrecht, D G & Thorell, L G (1982). Spatial-frequency selectivity of cells in Macaque visual-cortex. *Vision Research* 22, 545-559.

Frisby, J P (1980). *Seeing*. Oxford: Oxford University Press.

Grimson, W E L (1981). *From Images to Surfaces*. Cambridge: MIT Press.

Grossberg, S (1987). Cortical dynamics of three-dimensional form, color and brightness perception:I Binocular theory. *Perception and Psychophysics* 41, 87-158.

Hartline, H K (1938) The response of single optic nerve fibres of the vertebrate eye to illumination of the retina. *American Journal of Physiology* 121, 400-415.

Hubel, D H & Wiesel, T N (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *J. Physiology* 160, 106-154.

Jin, J S (1992). Depth Acquisition and Surface Reconstruction in Three-dimensional Computer Vision. Ph.D. Thesis, University of Otago, New Zealand.

Jin, J S; Yeap, W K & Cox, B G (1993). A neurocomputing model based on radial basis functions for stereo matching. Int. Conf. on Artificial Neural Network & Expert System, IEEE Press, p92-95.

Julesz, B (1960). Binocular depth perception of computer-generated patterns. *The Bell Syst. Tech. J.* 39, 1125-1162.

Julesz, B (1971). *Foundation of cyclopean perception*. Chicago: University of Chicago Press.

Kaufman, L (1964). Suppression and fusion in viewing complex stereograms. *American Journal of Psychology* 77, 193-205.

Kuffler, S W & Nicholls, J G (1984). *From Neuron to Brain: A Cellular Approach to the Function of the Nervous System*. 2nd Edition, MA: Sinauer Associates.

Kulikowski, J; Marcelja, S & Bishop, P (1982). Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biological Cybernetics* 43, 187-198.

Marcelja, S (1980). Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am.* 70, 1297-1300.

Marr, D (1982). *Vision*. San Francisco: Freeman.

Marr, D & Hildreth, E C (1980). Theory of edge detection. *Proc. of the Royal Society of London* B207, 187-217.

Marr, D & Poggio, T (1976). Cooperative computation of stereo disparity. *Science* 194, 283-287.

Marr, D & Poggio, T (1979). A theory of human stereo vision. *Proc. of the Royal Society of London* B204, 301-328.

Marr, D; Poggio, T & Hildreth, E (1980). Smallest channel in early human vision. *J. Opt. Soc. Am.* 70, 868-870.

Torre, V & Poggio, T A (1986). On edge detection. *IEEE Trans. on PAMI* 8, 147-163.

Trivedi, H P (1985). A computation theory of stereo vision. In: *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, CA, 277-282.

Wheatstone, C (1838). Contributions to the physiology of vision: 1. On some remarkable and hitherto unobserved phenomena of binocular vision. *Proc. of the Royal Society of London* B18, 371-394.

Wilson, H R & Bergen, J R (1979). A four mechanism model for spatial vision. *Vision Research* 19, 19-32.

Wilson, H R (1983). Psychophysical evidence for spatial channels. In: *Physical and Biological Processing of Images*, edited by O J Braddick & A C Sleigh, NY: Springer-Verlag, 88-99.