A Thesis Submitted for the Degree of PhD at the University of Warwick

Phase Relationships in Stereoscopic Computation

by

Keith Langley

Department of Computer Science

Thesis submitted to the University of Warwick
in partial fulfilment of the degree of
Doctor of Philosophy
March 1990

To Christopher and Yelena:

May your life be full and happy.

Acknowledgments

I am grateful to many people in the presentation of this work.
To them all I owe my thanks:

My supervisor Dr. M.H.E.Larcombe.
L.J.Clewlow: for many days of discussion and fun at Warwick.
My wife and children for their undeserved patience.
Professor J.M.Brady and Dr. B.Rogers who provided the opportunity
to complete, and made sure that I did.

A special thankyou to T.J.Atherton without whose help,
I would never have started this work.
A final thanks to the SERC who supported me initially as a Research Student
at the University of Warwick and later as a Research Assistant at the
University of Oxford.

Phase Relationships in Stereoscopic Computation

by

Keith Langley

SUMMARY

We apply the notion that phase differences can be used to interpret disparity between a pair of stereoscopic images. Indeed, phase relationships can also be used to obtain orientation and probabilistic measures from both edges and corners, as well as the directional instantaneous frequency of an image field. The method of phase differences is shown to be equivalent to a Newton-Raphson root finding iteration through the resolutions of band-pass filtering. The method does, however, suffer from stability problems, and in particular stationary phase.

The stability problems associated with this technique are implicitly derived from the mechanism used to interpet disparity, which in general requires an assumption of linear phase and the local instantaneous frequency. We present two techniques. Firstly, we use the centre frequency of the applied band-pass filter to interpret disparity. This interpretation, however, suffers heavily from phase error and requires considerable damping prior to convergence. Secondly, we use the derivative of phase to obtain the instantaneous frequency from an image, which is then used to improve the disparity estimate. The second measure is implicitly sensitive to regions that exhibit stationary phase. We prove that stationary phase is a form of aliasing. To maintain stability with this technique, it is essential to smooth the disparity signal at each resolution of filtering.

These ideas are extended into 2-D where it is possible to extract both vertical and horizontal disparities. Unfortunately, extension into 2-D also introduces a similar form of the motion aperture problem. The best image regions to disambiguate both horizontal and vertical disparities lie in the presence of corners. Fortunately, we introduce a measure for identifying orthogonal image signals based upon the same filters that we use to interpret disparity. We find that in the presence of dominant edge energy, there is an error in horizontal disparity interpretation that varies as a cosine function. This error can be reduced by iteration or resolving the horizontal component of the disparity signal.

These ideas are also applied towards the computation of deformation, which is related to the magnitude and direction of surface slant. This is a natural application to the ideas presented in this thesis.

Contents

List of Figures

List of Tables

Chapter 1

CONTEXT OF THE RESEARCH

1.1 Fundamental Issues

Let us consider the way with which man negotiates with the environment. Typically, navigation requires many sensing devices which are coordinated in a homogeneous manner. Vital to our navigational capabilities lie the optical passive sensors from which we glean our local and distant environment in some meaningful way. In general, we can regard visual processing as providing three main functions that are essential for navigation:

- Object recognition.
- Motion Analysis.
- Range finding.

The distinct advantage of visual processing lies with its passive nature. We might consider sonar as an active sensor for depth measurement. Such a sensor has indeed found wide uses from Polaroid cameras to submarine warfare. Yet there are many instances under which active sensing fails. In particular, at non frontoparallel planes. Under these conditions, the sonar beam is specularly reflected at an obtuse angle to the observer. Only minor quantities of sound are reflected back to the observer, while the main bulk of the signal is transmitted into space. Should such a signal return to the receiver, then the travelled path would be arbitrary and difficult to interpret. Vision does not have these problems. Naturally one needs light. But light waves come in many guises and can traverse across an abundance of media. There is only one drawback. We cannot yet merge visual processing into the natural environment as we might like. One approach is to emulate the type of processing that natural organisms utilise.

This introduces the concept of *stereopsis* as a biological solution to passive depth perception. Accordingly, this work, will draw heavily on the classical theory of Signal Analysis, and attempt a merger with what we currently understand by biological vision systems. One could not possibly be so bold as to claim that the developed theories of this work are truly biological. Nonetheless, a considerable proportion of the findings in this work bear striking similarities to the observed perceptual effects in our own optical system.

1.2 Stereoscopic depth perception

Have you ever wondered why with two eyes we rarely see double? Why a predator has two eyes placed at the front of the face, while his prey has lateral-viewing gaze? Why a camouflaged insect appear invisible in his natural terrain when viewed monocularly, only to leap out of the background environment when viewed binocularly. These issues form the evolutionary motivation for stereopsis.

It may be succinctly abbreviated as precision in depth discrimination. Additionally, stereopsis increases the the signal to noise ratio by comparison to monocular viewing. Bishop[5] found that the binocular luminance threshold is at least $\sqrt{2}$ lower than monocular viewing. Because of the visual acuity of humans, we can in theory discriminate objects with respect to depth up to 500m using stereoscopic processing.

The eyes are positioned about 7 cm apart with approximately horizontal separation relative to the ground during normal posture. The two independent views of the immediate space changes the retinal location of similar image features relative to an identical displaced reference in each retina. These disparities possess a functional relationship between objects and their distances from the observer. Intrinsic to the problem, is the recovery of these retinal displacements to yield depth discrimination. Psychophysically, the merging of two disparate images is referred to as fusion. The process of binocular stereoscopic vision is termed stereopsis. The purpose of this thesis is to recover stereopsis by artificial means.

1.3 Computer Vision

The primary interest of Computer Scientists and Engineers in visual processing is to control an autonomous system such as an AGV (Automated Guided Vehicle) or a Robotic manipulator. Vision serves to enhance the flexibility of Automation in the same manner that it aids natural animals. Recently, the access to fast massively parallel computers enhances the possibilities of vision as a realistic goal. In view of the massive architecture, and neural inter-connections of the mammalian visual cortex, it is not surprising that vision has been slow to develop as a unified theory with the previous limitations of computational hardware. The situation is equally frustrating because of the ease with which we use our visual skills at almost every instance.

1.4 Binocular Space

It is worth reviewing the space of the problem that is to be solved. Lunneberg[62] provided a generalised theory of binocular space based upon the angular disparity of objects in space. His work, was a detailed analysis concluding that binocular space was based upon hyperbolic geometry. His theory was able to account for many psychophysical observations. In particular:

- The psychophysical deviation from the Vieth-Muller circle.
- The inability of human observers to distinguish rectangular rooms from rooms with hyperbolic curvature.

The Vieth-Muller circle can be seen in figure 1.1. Point objects that subtend an equal angle to the retina should by Euclidian space, be perceived to be equi-distant. In practice, this is not observed. Hillebrand[79] showed that the Vieth-Muller percept only occurs at a unique distance. Psychophysically, the horopter (the locus of world coordinate points that project zero binocular disparity onto a stereo pair of sensors) is convex at far distances, and concave close to the observer (fig. 1.2). The second finding relates to the ambiguity of stereopsis and has subsequently been extended by several authors[61]. The importance in representing disparity space in hyperbolic geometry is to my knowledge unexplored, probably because the correspondence problem has yet to be adequately solved. The second point is equally as important, and probably questions the use of stereopsis as a single tool for an autonomous system.

1.5 Disparity gradient limit

Burt and Julesz[14] introduced the concept of the disparity gradient into vision community. However, Koenderink and Van Doorn[49] had previously discussed the same topic in terms of differential invariants of stereoscopic image transformations. Burt and Julesz found that disparities alone were not sufficient to dictate the perception of fusion within a stimulus. In particular, they found that the disparity gradient, (defined as the difference between the disparities of neighbouring objects divided by their Euclidian separation) could not exceed a critical value and still retain binocular fusion. Since this applied to features with small

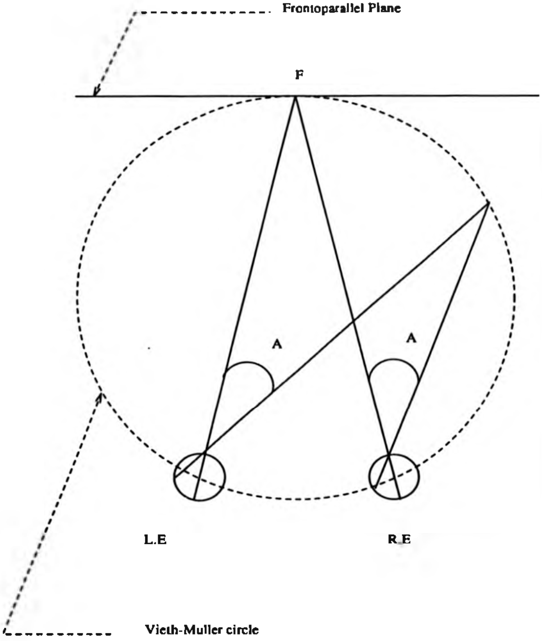


Figure 1.1: *Vieth-Muller circle, showing theoretical points that subtend equal visual angle to an observer fixating at a point F in the visual world.*

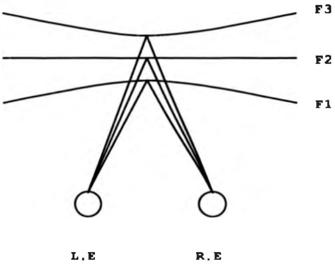


Figure 1.2: *Psychophysically observed deviation from the Vieth-Muller circle observed by Hillebrand, and predicted by Lunneberg in Hyperbolic space.*

disparity differences, the classical notions of Panum's fusional areas[79] and his limiting case were re-interpreted in the context of disparity gradient.

Burt and Julesz suggested that there could exist a critical angular difference between line segments, which viewed stereoscopically prevented fusion. This was seen as a direct consequence of orientationally selective filtering. Such a mechanism will be intrinsic to the later theories developed in this report. Trivedi and Lloyd[94] showed that a disparity gradient of less than |2| preserves the topology of matched image features. Thus, a disparity gradient greater than |2| implies that corresponding points do not lie on the same surface. An example of a stimulus that violates the disparity gradient limit occurs with points on a surface normally invisible monocularly, but still observable because a particular surface is transparent. The disparity gradient limit is therefore a statement concerning the surface structure of a scene such that surfaces are assumed opaque.

1.6 Primary and Secondary cues for stereopsis

There are several hypotheses that can be combined for models of human stereopsis. Ultimately, these models are based upon both monocular and binocular cues. Primary cues in human stereopsis is defined as the mechanism used by the visual sensors to simultaneously provide local depth information that is independent on monocular vision. This stage is often referred to by *low-level vision* because the process appears to function prior to the high cortical levels, of cognitive influence. Primary cues therefore operate oretinal disparities. Complementary to this notion it is proposed that a higher level of cognitive depth perception operates to matches features independently and subsequently merge monocular sensations. Gibson[28] proposed that secondary cues such as:

- Object occlusion or overlay.
- Texture density.
- Relative size.
- Brightness and shading.

that are inferred monocularly can influence depth perception, and in some instances dominate the primary cues for stereopsis. There exist important differences between the two mechanisms. First, primary stereopsis can function effectively in the absence of monocular cues. We know this from the psychophysical fusion of the random dot stereogram (RDS)[44] where there are no visible monocular features but there are binocular disparities. Since the perception of depth is still apparent, it is supposed the low level vision uses primary cues for stereopsis which must rely on the *signals* present in a stereoscopic pair. In contrast, it is likely that secondary cues rely on the globality or consistency of image in the form of monocular cues. This would hold the advantage of ensuring that correct correspondences have indeed been established from the low level process and further disambiguating false matching sensations.

1.6.1 Primary and secondary cues in machine vision

Ultimately, we would like to consider utilising both primary and secondary cues to obtain stereoscopic depth perception. However, within machine vision the primary cues for stereoscopic computation are particularly attractive because it is an intrinsically autonomous process. In contrast, secondary cues rely heavily on learned behavioural patterns from past experiences and an influence from pattern recognition. We are, however, still short of these secondary perceptual goals.

We therefore argue that machine vision should be predominantly influenced by a primary mechanism for computing stereoscopic disparities. Thus, a computer algorithm whose goal is to emulate a low-level vision process in its implementation should be:

- Autonomous.
- Process disparities where available that is independent of the structure of the scene.
- Able to construct a depth map from the field of view of two cameras.
- Robust in its practical implementation (e.g not rely heavily on the epipolar constraint [1]).
- Processed in parallel (speed) providing parallelism is economic.
- Understood in terms of its weaknesses and strengths.

The fundamental goals of this thesis are to indeed implement such a system by artificial means.

1.6.2 Basic Principles

The theoretical motivation for the approach presented in this work, is succinctly summarised by the title of Granlund's[31] early publication:"*In search of a general picture processing operator*". This work has subsequently been extended by two particular authors: R.G.Wilson[98, 100] and T.J.Atherton[2] who have been guiding influences in the development of this work. Granlund, idealised that there existed a generalised representation of an image function that may be linearly combined to extract the information that we might require from an image function. Within this approach, there exist important basic principles:

- Contrast independence.
- Scale Invariance.
- Linearity.

That the human visual system functions under contrast independency was discussed theoretically by Wilson and Knuttson[99] as inference in the human visual system, and indeed

---

[1] Here epipolar constraint refers to parallel camera geometry such that disparities are assumed to lie along horizontal image raster lines

forms the basis for much of the feature based analysis within image processing. The essential point identifies that correspondence or decisions regarding image data should not be influenced by mean intensity changes. An obvious example lies with the variation in background illumination that typically occurs under artificial lighting conditions. The second principle is perhaps easier to comprehend in the context of pattern recognition. Thus, the operations required to recognise an object should be invariant to the solid angle projected onto the image sensor. This statement is in fact equivalent to the requirement that the operations applied at any given resolution of filtering should all be equivalent. The final principle is also associated with the method of solution, but in any case states that the types of operations, and decisions obtained from image interpretation are based upon linear approximations. Within the context of this approach, it is not important that other algorithms may yield more accurate disparity information. The importance of this work is derived from its implicit generality to include these goals of image processing in a single representation.

1.7 Underlying premises used

The principle that we will use to attempt to gain retinal disparities, will be based upon the differences in phase between a pair of *analytic signals* formed by the Hilbert transform. We will refer to this technique as the *method of phase differences*[73, 90, 100, 42]. Granlund and Larcombe (personal communication) have also independently conceived the notion. There are a number of fundamental constraints that we implicitly introduce by applying this technique. These assumptions follow directly from Marr[67] in the form of:

- *Compatibility*: Image signals can only be matched with zero-phase differences, and both similar orientations and spatial frequencies.
- *Uniqueness*: Correspondence involves one-to-one matching.
- *Continuity*: The disparity field varies smoothly almost everywhere.

These are the basic principles that are involved with the method of phase differences. The presence of measurement noise adds the restriction that phase differences should be minimised. Because we use phase differences to measure disparity, we are not constrained to the monotonicity constraint [2] and can measure different signs of disparity without difficulty. At one point in this work, we do define a minimum disparity of zero with parallel camera geometry, which does not appear unreasonable.

1.7.1 Advantages of the phase representation

There are a number of advantages that a phase descriptor through bandpass filtering holds within the context of signal processing. The list presented is by no means exhaustive:

- The division between odd and even filter pairs removes the filter envelope from analysis, leaving a descriptor based purely on signal symmetry.
- Reduced number of bits required to encode a signal[80].
- The instantaneous frequency can be extracted via the derivative of phase.
- Reconstruction is possible to within a constant that is considerably improved in comparison to the magnitude only reconstruction[92].
- Improved restoration of noise blurred images (blind deconvolution).
- The auto-correlation of a phase only function is always an impulse[80].
- Stability through Scale-Space filtering[41].

1.8 Current Problems in Low-level Machine Vision

Without referring to specific algorithms, it is worthwhile reviewing the major problems within stereo correspondence algorithms:

- Large disparities.
- False matching.
- Discontinuities in the disparity field.
- Illumination.

---

[2] The monotonicity constraint refers to parallel epipolar camera geometry such that disparities are constrained in sign

- Sensor noise.

Large disparity differences are partially dealt with by coarse to fine resolutions of filtering. This method implicitly contains restrictions on the image signals. In the presence of noise, image detail with fine spatial extent are difficult to fuse because of aliasing. However, near objects subtend a larger visual angle on the image sensor, and even though disparities also increase proportionally, the detection of close features is enhanced because of the spatial representation on the retina relative to other features. False matching is particularly problematic for edge based approaches, since edges themselves are difficult to distinguish. Within the contxt of this work, we apply spatial frequency, and orientation constraints to reduce the possibility of false matching. Discontinuities in the disparity field involve marked image intensity differences between image pairs typified by occluded features and surfaces. An approach which is based upon linerity will be unlikely to char, acterise such regions. In the same manner as sensor noise, and illumination differences we deal with these artifacts by smoothing, thus attempting to reduce the accuracy of results by these factors.

Chapter 2

BACKGROUND TO MACHINE STEREO VISION

2.1 Background to Machine Correspondence algorithms

The evolution of microprocessor technology has heralded several new possibilities for research. In particular, it has enabled large quantities of numerical data to be evaluated with increased speed and resolution. Attempts to replicate human visual functions has been an area of considerable interest for over twenty years. With the prospects of image interpretation moving into the real-time environment, and the immense industrial and military applications of this technology, it is not surprising that the area of image understanding has become a highly competitive environment.

Interest in the perception of the three dimensional world can be traced back to the early work of both Helmholtz and Wheatstone[35, 96] in the late 19th century. However, replicating these mammalian functions can be traced only relatively recently with some early work on road tracking vehicles in the 1960's. There are several approaches to the correspondence problem which may be neatly divided into the categories of *area based* and *feature based* matching, with newer approaches that combine both techniques into a single representation.

2.2 Area based methods

Two-dimensional windowing operators are used to measure the similarity in structure between two local areas in a number of images. Typically, cross-correlation would be the normal method applied to these techniques. To provide some measure of illumination invariance to the matching criterion, normalised cross-correlation has some notable advantages in that compensation to mean changes in illumination can be permitted[34]. Area based correspondence has been applied successfully to the analysis of rolling terrain, but it does degrade rapidly when the image is not smoothly varying or continuous. Inherent to the problem, is deciding the shape and size of correlation windows in the general case. In addition, the uncertainty problem plays a predominant role in these techniques. In brief, the uncertainty lies in large window areas, that are essential for statistical significance with an associated loss in resolution and accuracy in the image domain. Levine et al, and Mori[59, 76] have addressed this problem, and vary their correlation window sizes with the local surface intensity variance. They make the assumption that high local variance implies high local texture and therefore requires smaller correlation windows. In contrast, low variance suggests surface uniformity and the need for larger correlation windows. Gennery[26] criticises area based correspondence because the process is essentially serially driven, with finer correspondence based upon the larger window operations. In the case where false matches occur, it is clear that these methods can become unstable which will lead to large errors in the estimation of disparity. In particular, a complex image will contain many 'local minima' of potential false matches. It is therefore essential for an algorithm to detect a false match. In brief, area based methods assume a one-to-one matching constraint, but can break down severely at depth discontinuities. Psychophysics, has shown us that human subjects deal with surface discontinuities by binocular rivalry. Machine Vision has to date limited scope for such phenomena.

2.2.1 Gimel'farb, Marchenko and Rybak

Gimel'farb[30] proposed a stereo correspondence algorithm using a linear line by line epipolar constraint. They also used known disparity constraints to limit the correspondence search. They optimised a cost function based upon dynamic programming and normalised cross-correlation. To reduce the computation time, they suggested using previously matched lines to guide the matching constraint on the current epipolar lines of interest. They also performed a crude coarse to fine matching strategy, which has subsequently been used by many additional authors. Presentation of their results was restricted to single line segments and therefore it is not possible to evaluate their work.

2.2.2 Levine and O'Handley Algorithm

Levine and O'Handley[59] describe a proposal for the Mars Rover Vehicle's autonomous navigation. They used intensity cross-correlation techniques, and collinear epipolar imaging to limit the correspondence search. The matching constraint was achieved via intensity cross-correlation , with an adaptive window based upon the statistical variance of the neighbourhood of interest. Thus the algorithm proposed a crude coarse to fine matching constraint. Since their algorithm was intended for natural terrain, they proposed to limit their search constraints based upon the upper, and lower lines of their stereo pair of images. They presupposed that the scene always extended towards the horizon on the upper lines of the images. Therefore disparity would increase from the top to bottom of the camera images. On the basis of this matching constraint, they set local maximum and minimum disparity ranges. They assigned "tie points" to areas with large variance values, and proceeded to iterate their coarse to fine matching process. More expensive correlation was then applied to refine their disparity estimates. Cross-correlation was also limited to image segments of similar variances. Their algorithm assumed that occluded features and edge reversals were not present in the image pictures. The results presented were smooth disparity maps, with the false matches removed.

2.2.3 Hannah Algorithm

Hannah[34] placed considerable emphasis on techniques to increase the efficiency of area-based correlators. Her work suggests improvements by:

- Correlating over a sample of the image and refine the match based upon a correlation over the whole of the image. Points with maximum correlation coefficients therefore define matches.
- Correlating over reduced resolutions and then refine the matches at increased resolutions of examination.
- Extracting statistical image characteristics as a guide for limiting correlation (Levine and Handley) window sizes.
- Using the known camera geometry to limit the search (epipolar analysis).

Hannah implemented her algorithm by expanding correspondences from matched pairs into larger regions. She assumed surface continuity, and by the use of the autocorrelation function, she assessed the quality of matches.

2.2.4 Moravec's Algorithm

Moravec[75] aimed his research towards autonomous vehicle guidance by visual sensing. His technique reduced the computational load, by analysing regions with high local variance which he used to provide cues for motion calibration and obstacle avoidance. There were 3 main contributions to the vision community from his work. Moravec's *interest operator* was a filtering technique primarily aimed at locating corners and other distinguishable image features of high local variance. These regions of interest were then processed by a *binary correlator*. The binary correlator functioned in a coarse to fine strategy, where each feature found by the interest operator was represented at different sampling resolutions. The algorithm begins from the coarsest scale of filtering. The pair of images are then cross-correlated, and the largest cross-correlation coefficient is taken as the correct match. The window from the next highest resolution of sampling, is then used to improve on the estimate from the previous resolution of filtering. Here the region which is considered at the next level of cross-correlation is defined by the window size of the previous resolution. The algorithm terminates at the finest scale of filtering. Moravec claims a 10% error rate with this method. The final part of his algorithm :*slider stereo* involved moving the camera along a horizontal track, and taking 9 equally spaced , horizontally separated images from the same vehicle position relative to the scene. Correlations from the 36 possible permutations of stereo pairs provided a series of depth estimates which were represented assuming a Gaussian distribution. By weighting each depth estimate by the correlation coefficient from the binary correlator, the weighted mean was taken as the correct depth estimate. For Navigation purposes, Moravec also incorporated a predictor/corrector driven search based upon the known movements of the vehicle, and the previous depth estimates taken from the cross-correlation procedure.

2.2.5 Gennery Algorithm

Gennery[26] proposed a system specific to autonomous navigation for vehicles. The algorithm used cross-correlation to position points in space. The system incorporated the Moravec *interest operator* and *binary correlator*. These techniques were used to estimate a ground plane in the scene. It was then assumed that most image features would lie above the ground plane. This assumption was applied to estimate the relative orientations of the camera sensing devices. The assumptions made by Gennery permitted the epipolar constraint to be applied and analysis was then restricted to a 1-D search. Having gained estimates of the scene noise characteristics (variance, gain and bias between the image data) he defined a correlation measure to provide sub-pixel positioning of corresponding windows and accuracies of correspondences. The algorithm progressed from left to right in raster fashion, and local depth estimates were used to predict future match sites. In the case of ambiguous matches, the epipolar constraint was also used to provide further suggestions for correspondence based on the parallel camera epipolar constraint. Because of the last assumption the algorithm began the correspondence search from zero disparity and continued until either a match was found, or a previously matched window located. Within the vicinity of a selected area, the maximisation of a local correlation coefficient was set to determine the best match. The novel feature of Gennery's work, lay with the notion of fitting ellipses around detected obstacles, which he contended were an appropriate shape representation for obstacle avoidance and scene matching.

2.3 Feature Based Analysis

We criticised area based methods in view of the poor definition of window sizes, and also because the methods assume that the photometric properties of a scene are invariant to image position. This is of course not the case. Feature based analysis avoids much of the problem by working on the premise that a local measure on the intensity function is representative of the physical change in the underlying scene. By the physical change in the underlying scene, we refer to depth discontinuities, textural differences, luminance changes, shadows etc. Feature analysis implicitly uses the semantics of intensity variations to infer the physical change in 2-D image data, which are then used to obtain the 3-D structure of a scene. Probably the most widely known feature based algorithm is that of Marr and Poggio:

2.3.1 Marr and Poggio Algorithm

The approach of Marr and Poggio[67], was primarily based on the Psychophysical and Neurophysiological data that was then available. Their work may therefore be seen the first serious attempt to merge the fields of Image Processing and Psychophysics. In its raw form, they implemented an algorithm primarily based upon the *uniqueness* and *continuity* constraints, and implemented a matching algorithm based upon the zero-crossings obtained by convolving an image with the $\nabla^2 G$ operator. The algorithm was implemented on real image data by Grimson[32] as follows:

- Four stereo planes are assigned with zero-crossing values and orientations corresponding to 4 different spatial frequency tuned band-pass filters.
- Set the initial vergence values for the eyes.
- Match zero-crossings with similar gradients and and orientation preferences. Matches were assigned as positive, negative and zero disparity relative to the vergence angle.
- Mark ambiguous or "no matches" as such.

- Check unmatched points in regions and where unmatched points exceed 30 %: delete all matches.

- On the basis of low frequency filter matches, make various positive or negative vergence movements to bring unmatched (high spatial frequency) elements into correspondence, and iterate on the matching process.

After the matching with zero-crossings has been achieved, a subsequent process interpolates a smooth surface to the sparse depth map to obtain a richer depth description. The results published include random dot stereograms composed of 4x4 elements with disparity ranges from 2 to 6 dots width. They also present some real data from the Mars Viking Vehicle. This particular algorithm has been criticised by several authors. Baker[3] and Frisby[24], both believed that coarse to fine matching is an inadequate model for human stereopsis. However, it is proven in Psychophysics that depth cues are dependent on both the image spatial frequency elements present, and retinal eccentricity[85]. Retinal eccentricity provides indirect evidence. It is well accepted that cortical cells that receive inputs from the retinal periphery are predominantly tuned to lower spatial frequencies than from the fovea. Experimentally, it is also accepted that Panum's fusional areas also increase with retinal eccentricity. These factors are all consistent with the expected effects of spatial-frequency analysis and large retinal disparities (which cannot occur simultaneously in the fovea). Therefore, coarse to fine matching cannot be dismissed. Baker also criticised the absence of monocular stimulus in the Marr and Poggio algorithm. However, the theory developed by Marr and Poggio was essentially a low-level vision hypothesis. Thus while Baker's criticisms may be correct, monocular cues were not an issue considered by the Marr and Poggio algorithm. Baker further criticises the Marr and Poggio algorithm on the basis that the implementation does not define the precise nature of the vergence mechanism but relies on chance for large disparities. The biological vision system also functions with saccades which could serve the same purpose. Mayhew and Frisby[72] also raised the objection that the human visual system appears to use the local energy peaks as well as the zero-crossings from bandpass filtering. Such a criticism is consistent with the approach considered in this thesis.

There are two criticisms that might be added to this work. First it is established that a considerable portion of cortical processing relies heavily on orientationally selective filtering. Indeed Hubel[38], points out that all the binocularly tuned cells he studied exhibited orientation preferences. The $\nabla^2 G$ operator is not orientationally selective, and does not respond to image corners. In addition, cells in the LGN and striate cortex appear to conform with both even and odd symmetric profiles with orientation preferences[18]. No effort is made to comply with these findings. Therefore, the stereoscopic aperture problem[54] cannot be solved by Marr's ideas. We suggest that Marr and Poggio's work appears too early in the hierarchy of mammalian visual processing. Marr and Ullman[68] did, however, provide a theory for orientation selection based upon the time derivative of the response to zero-crossings in addition to logical operations to constrain velocity. This theory, however, is not strictly applicable to stereopsis as an instantaneous evaluation.

2.3.2 Baker Algorithm

Baker[3] as an extension to Henderson[37], describes an edge based algorithm for correspondence. Using epipolar geometry and the continuity constraint, he applies these factors to limit matches and remove the edge reversal problem that can occur with converged camera sensors. While recognising the limitation of sparse disparity maps in obtaining a rich description of the scene, Baker additionally matches intensity values defined between edge segments to obtain an improved depth map of the viewed scene. Baker thereby attempts to draw on the advantages of both area based and edge based correspondence. To optimise the edge matching strategy, he employs a modified version of a Viterbi algorithm which is a recursive optimal solution to the problem of estimating the state sequence of a discrete time finite state Markov process. The process is Markov because the probability of correspondence was defined as dependent only on the present state and time of search. This aspect of the algorithm is implemented by dynamic programming.

Baker also proposes a probabilistic weighting based upon orientation differences between matched segments, but it is not clear from his work, exactly how edge orientation is obtained. Fundamental to his approach lies the computational savings achieved by image subsampling, which is also used to reduce the noise between stereo pairs of images to less than one intensity measurement. The edges obtained from the reduced resolutions, are then tracked to the highest resolution image (subject to the noise constraint) at which point, the processes is repeated and constrained by the coarse to fine resolutions of matching.

In short, Baker's work combines the best features from the available algorithms at the time. Results are presented for synthetic and for natural terrain data. In the case of the natural terrain data, the final depth surfaces required median filtering which implies that the algorithm still contains local instability in depth estimation. This is not surprising with natural image data.

The results presented by Baker are clear, impressive and require no manual intervention, which had been a previous limitation to many of the stereoscopic algorithms previous to his work.

Baker's work may be criticised on two fundamental issues. First, he implemented the Sobel operator which is not optimal for edge detection[16]. Second, Baker did not have a mechanism to deal with the spatial migration and bifurcation of edge based data through Scale-Space (e.g [17]). This behaviour of "edges" through Scale-Space introduces a further Scale-Space correspondence problem.

2.3.3 Porrill et al Algorithm

Porrill et al[84] have recently proposed an algorithm based upon edge extraction by applying the Canny edge operator. This algorithm was seen as an extension to the PMF algorithm[83] which has been successfully applied to many stereo image pairs. Their main contribution to knowledge applies the disparity gradient limit to constrain possible matches.

The PMF algorithm was recently updated and renamed SMM . This new algorithm may be summarised as follows;

- A feature is chosen that is in focus
- The S closest features greater than a length L are identified.
- Potential matches are identified.
- Consistent matches with similar geometrical relationships are established (disparity gradient hypothesis).
- Clusters of potential matches are grouped together with similar disparity measurements. The transformation of image data is recorded.
- Inconsistencies with the matching are searched and isolated from the matching process.
- The goodness of match is ranked on the basis of the numbers and length of all members in a cluster.

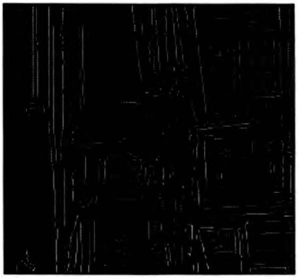


Figure 2.1: *The PMF algorithm applied to our test stereo pair presented in figure B.1. Intensity of edges proportional to disparity.*

- A sparse disparity map is then created on the basis of the inverse disparity transformation.

Results from the sparse disparity map are unclear. No error measurements are presented. Images represented "block worlds" and were relatively featureless. Their method may survive in the restricted environment of a controlled industrial workplace, but would pose numerous difficulties in a general application. This point is stressed in figure 2.1 where the algorithm has been tested to our own raster test sequence presented in figure B.1. The algorithm has managed to match a considerable number of edges, however, it has failed to match the lamp feature in the central region of the image because of the large disparity difference. Typically, the main areas area of concern would be false matching and ambiguity owing to a high density of edge information. Consequently, the uncertainty problem imposes a difficult constraint on the algorithm. Additionally, it is not clear how a depth map reconstruction would function in the event that a high proportion of surface edges were not matched. We would anticipate, that either a coarse to fine matching strategy or further intensity based correlation between edges would considerably improve their algorithm.

2.3.4 Bolles et al Algorithm

Recently, Bolles et al[9] proposed a solution to many of the stereoscopic and motion correspondence problems by forming a spatio-temporal solid. By taking up to 100 pictures from a single camera in rapid succession, they form a "solid" which represents the 3-D scene of the world. In essence, they take the epipolar-plane analysis to the limit. By knowing the camera motion characteristics, they analyse their data along horizontal rasters. Of particular interest is the location occluded boundaries between subjects, and the ability to build a 3-D map of "free space". Fundamental to their approach lies the fact that edges can be tracked by their small displacements between image pictures. Occluded boundaries appear as a locus of crossed edges along their time slice of an epipolar plane, which occurs because of motion parallax. Their method resolves considerable ambiguity in image understanding. However, by restricting their use to epipolar geometry, they cannot interpret image orientation. Therefore, their method will be highly suspect when faced with the Motion Aperture problem in the General case.

2.3.5 Terzopoulos, Witkin and Kass Algorithm

As an alternative to traditional feature and area based methods, Terzopoulos et al[93] proposed a paradigm for both shape and correspondence based on dynamic energy constraints. They proposed to model image features as deformable elastic bodies which are subject to external forces owing to the constraint being imposed on the image data. Their approach seeks to define a minimum energy constraint based upon the deformation required to map a motion, or stereoscopic sequence of images into correspondence. They also track deformations through coarse to fine resolutions of filtering, to reduce the possibilities of their algorithm becoming trapped in local minima. The minimisation is taken over the spatial derivative of the energy function. Since the amplitude and derivative of a spatial convolution is also closely related to the spatial frequencies detected by a filter (within a constant), their methods hold a similarity to the work presented in this thesis, although via different roots. Their advantage over the methods presented in this thesis lies with the stability of an energy response over a phase calculation which is particularly true over low-energy regions of a complex image function. Unfortunately, their algorithm cannot claim to deal with contrast independence which is one of the advantages of the methods proposed in this thesis.

They also claim that their method can compute 3-D object models directly without requiring a $2\frac{1}{2}$ -D sketch, as originally proposed by Marr[66]. Their methods present impressive results from a simple motion sequence. However, their stereoscopic image data has been largely restricted to simple data, with small disparity differences.

2.3.6 Brint and Brady Algorithm

Recent extensions to the feature based analysis lies with the work of Brint and Brady[12]. They form and use the premise, that regions of high image curvature occur infrequently within natural image data, and therefore can be used to reduce the correspondence problem. Although their work does not make claim to solve the correspondence problem they do outline methods from which corresponding curves from stereoscopic data can be matched. Ironically, their published work applies the Canny edge detector to their natural image data, which in itself fails to isolate regions with high curvature (i.e corners). They are at present, rectifying this flaw in their processing by the application of improved edge detecting algorithms (personal communication). Fundamental to their approach lies the representation of an edge contour by an elastic string, in which *knots* are located at regular intervals which serves to minimise a cost function based upon the measured curvature and a defined energy required to deform one curve into a stereoscopic pair for correspondence. Stereoscopic matching is then obtained by identifying corresponding *knots* from the stereoscopic pair (in their case trinocular), subject to a minimisation of the deformation between stereoscopic edge contours, and the disparity gradient hypothesis. Their work is still under development, and therefore it is difficult to fully appraise their notions. We can, however, suggest that they could improve their algorithm by obtaining an orientation field based upon coarse to fine strategies of directional filtering. The result of this type of processing would not be restricted to edges, but would enable meaningful orientational descriptions throughout the image providing a signal exists within the receptive field of the filters. Such an approach would be less noise sensitive and also readily implemented on a coarse to fine mechanism.

2.4 A critical overview

We have discussed several algorithms that have been applied towards the computation of disparities. In chapter 3, we will discuss the role of uncertainty in Image Processing[25, 98, 99] which forms its basis in the simultaneous representation of a signal in both space, and frequency. From the words of Marr[65]: The role of vision is to find out *what* is *where*. Unfortunately, both edge and area based techniques suffer from similar forms of the uncertainty principle[25]. With edge based stereoscopic algorithms, the presence of edge information alone yields considerable ambiguity since edges themselves are difficult to distinguish and can also occur spuriously within an image owing to the presence of noise, and the complexity of natural image data. In terms of the uncertainty principle we know *where* but not *what*. Recent attempts (e.g [12]) have attempted to relieve the ambiguity by classifying edges by orientation and curvature, with the premise that curvature and orientation differences between image functions are sufficient not only to reduce the correspondence problem but also yield unambiguous matches. However, the sparse disparity map that results from these approaches also requires further interpretation, which relies heavily on the assumption that correct matches have been established.

In contrast, area based correspondence algorithms usually apply normalised cross-correlation techniques to obtain measures of similarity between image functions. These techniques suffer from poor definition of window sizes that are used during the cross-correlation procedure. Cross-correlation mechanisms are also unable to define a *what* to the signal. We form the premise that at the lowest level of vision processing, *what* refers to both the local spectral properties (instantaneous frequency) and orientation of our image data and *where* is obtained through the resolutions of band-pass filtering. Thus we serve to combine both the benefits from feature based, and area based algorithms, in a single concurrent representation that may be subsequently applied for both motion analysis and pattern recognition.

2.5 Area and Feature based analysis

More recent techniques are now being proposed that combine the approaches of area and feature based analysis. Although these techniques can be viewed as general extensions to the Marr and Poggio algorithm reviewed earlier, the techniques retain the advantage of permitting a greater density of information to consider. This is achieved by the creation of an *analytic form* of the image signal by forming its Hilbert Transform. Within these techniques, edge and bar (lines) features are represented by the phase of the analytic signal. Within the phase circle, edges correspond to odd symmetry and the line refers to even symmtery. Transitions between edges and bars are merely represented by a local measure of symmetry.

2.5.1 Miller Algorithm

Miller[73] suggested a method of stereoscopic correspondence based upon a hardware implemented version of a phase locked loop (fig. 2.2). His methods, are conceptually the same as proposed in this thesis, however, the method is not able to utilise the true dimensionality of Image Processing, since the analysis was restricted to the single dimension. His techniques are therefore prone to labelling false matching from orientation differences that cannot be resolved in 1-D.

The basic technique employed by Miller is shown in figure 2.2. He proposed that pixels recorded from the the right CCD array (SR(t)) are clocked out in raster lines at a fixed rate using an analogue shift register. The image data in the second stereo image is similarly clocked out but at a rate determined by the disparity buffer. The two signals are then fed into the phase detector which emits an amplitude proportional signal phase difference (SP(t)). The phase signal is low-pass filtered and then added to the disparity buffer and the process repeated. Miller based his system for phase detection on simulation of the Motorola MC14568 digital phase detector. His method was also limited to binary threshold image data, and by his own admission, took no account of edge reversal or occluded boundaries. The limitation to binary thresholded data serves to enforce a constant envelope present in the information bearing signal. The application of limiters in the demodualtion of frequency modulated signals in well known in Telecommunication theory[91]. However, with image data it is difficult to see why Miller chose to threshold image data at such an early stage. Miller did not give a rigid description of the technique that he employed although it is clear that his system incorporated many of the ideas that have subsequently been applied in this thesis. The Phase-locked loop technique, produced a smooth disparity map. Therefore, occluded boundaries which may be considered with frequency characteristics similar to broadband noise (assuming that they are of small spatial extent) were effectively removed by smoothing. A similar effect can be observed by the methods proposed in this thesis, but restricting analysis to single channel of Gabor filter tuned to low spatial frequency. Miller does, however, suggest that improvements could be made by applying the phase-locked loop technique in parallel. This aspect will be investigated later.

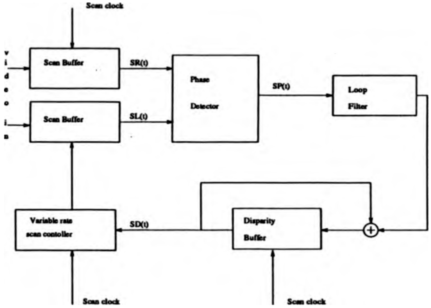


Figure 2.2: *Block diagram of the hardware system proposed by Miller. Image raster lines SL(t) and SR(t) are fed in at video rate with an output voltage (SP(t)) proportional to the phase difference. The phase differences are used to control a disparity buffer which introduces a time delay proportional to the estimate of disparity, which is used to alter the reading of the second stereo image.*

2.5.2 Sanger Algorithm

Sanger[90] was the first to apply the method of phase differences towards disparity computation. He applied Gabor filters and interpreted disparity using the center frequency of the filter. Interestingly, the largest disparity that Sanger considered was a seven pixel shift in a 512x512 image. However, Sanger did show that the technique of phase differences could operate over relative disparities as small as 1/32 pixel.

His basic method employed a coarse to fine strategy of filtering, followed by a simple weighted estimate:

$$d = \frac{\sum_{\omega} \Delta X_{\omega} r_1(\omega)}{\sum_{\omega} r_1(\omega)} \tag{2.1}$$

where d refers to the mean disparity, $r_1(\omega)$ refers to an energy weighting at each resolution, and ΔX_{ω} refers to the disparity estimated through each resolution of bandpass filtering. This was the first scheme considered during the progress of this thesis. However, it was felt that the technique relied too heavily on the coarse resolutions of filtering and therefore did not

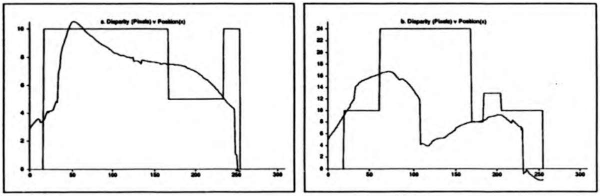


Figure 2.3: *Implementation of Sanger's algorithm to (a) figure B.3,* $r = 0.88$, $\bar{x} = -0.96$, $\sigma_n = 0.19$ *and (b) figure B.2,* $r = 0.65$, $\bar{x} = -6.3$ $\sigma_n = 0.67$. *Gabor filters at the coarsest scale were* $7.8125\text{x}10^{-3}$ *cpp up to 0.25 cpp sampled at 0.5 octave intervals in the frequency domain. The measured edge based disparity values are shown superimposed.*

fully utilise bandpass filters at higher spatial scales of resolution. This is particularly true in the case of large local shifts in the image. The technique is therefore extremely susceptable to phase wrap-around at coarse resolutions of filtering, which will result in large errors in disparity measurement. We present results of the technique in figure 2.3, which should be compared to figures B.6a and B.6b.

Sanger also proposed a disparity gradient limit based upon local differences in the pre-envelope obtained by convolution between quadrature filters and the stereo images. However, it is not clear how such a notion was implemented from his paper.

While Sanger dealt with a number of important issues, he did not consider the two-dimensional extension of the technique to obtain both vertical and horizontal disparities. He also did not realise that the spatial phase gradient can be interpreted as an instantaneous frequency and hence provide a natural mechanism for describing a disparity gradient limit. The local phase gradient may also be applied to indicate regions that are not analytic by the technique, and also improve disparity measurement.

Sanger presented results to random dot stereograms, and natural image data. Results were impressive and a suitable acknowledgement should be registered.

2.5.3 Wilson and Knutsson Algorithm

Wilson and Knutsson[100] have recently proposed an algorithm based upon phase differences. Their algorithm, also used Gabor filters, however, they limited their theoretical analysis to white noise and random dot stereograms. Since the Fourier transform of white noise is constant, they assumed that the spatial gradient of quadrature phase may be represented by the center frequency of the quadrature Gabor filter. They did not, however, consider the variance of phase behaviour, nor did they consider the error in disparity interpretation from their assumption.

Some of their ideas will be used in this thesis. In particular, they proposed using Willsky's[97] error measure on a circle:

$$\frac{1}{2}(1 + \cos\theta_d)$$

to reduce the possibilities of phase instability when phase differences approach the aliasing limit of ± π radians. This damping term is, however, especially required when the interpretation of disparity is taken from the center frequency of the filter. In the general case, phase extracted from a dominant spectral component will behave in harmony with the signal frequency and not the filter. Unfortunately, we will show that the magnitude of the disparity error is proportional to the phase difference measured from a pair of images, and the error in image frequency interpretation. Therefore, we would also expect severe errors in disparity measurement as phase differences approach phase wrap around. The application of Willsky's error measure is then appropriate under these assumptions. Their algorithm also propsed applying Gabor filters in the form of a Phase-locked loop, however, they also suggested that several iterations of the phase differencing technique might be employed at each scale. We present two 1-d sequences based on their technique in figure 2.4. It should be stressed that each resolution was iterated 3 times to obtain convergence. To be fair we point out, that successive iterations at a given resolution based upon local disparity estimates contributed significantly to the instability. In addition, they also used filter pairs with 2-D spatial extent, which we would expect to decrease the noise sensitivity considerably. In any case, the Willsky error measure provides significant improvements in stability, at the expense of resolution loss at each bandpass resolution.

The method proposed by Wilson and Knutsson for obtaining the argument of phase was based upon the sums and differences of squared responses from filter pairs. However, by

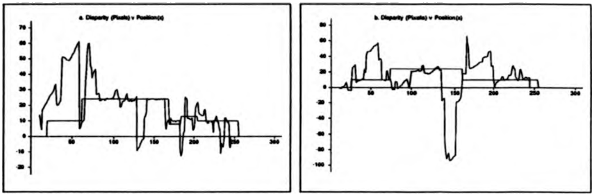


Figure 2.4: *Example of Wilson and Knutsson algorithm applied to: (a) figure B.2, r= 0.46;* $\bar{x} = 3.37$; $\sigma_n = 1.14$ *and (b) figure B.4., r = 0.04;* $\bar{x} = -1.99$; $\sigma_n = 1.97$.

expanding their expression we see that the equivalent form may be represented by:

$$\tan \phi_d = \frac{< \mathbf{z}_l, \mathbf{z}_{r\perp} >}{< \mathbf{z}_l, \mathbf{z}_r >} \qquad (2.2)$$

where $\mathbf{z}_l$ and $\mathbf{z}_r$ represent the convolution between quadrature filter pairs in the corresponding images. This form, although considered was not used in this thesis because it was felt potentially unstable (personal communication from M.H.Larcombe:1987). This can becomes apparent in the presence of phase singularities[41], where a quadrature pair of filters under convolution simultaneously return a zero amplitude response. The above equation then reduces to an indeterminate form. However, our techninque for disparity interpretation assumes a definition of zero phase is the absence of an energy response from quadrature filter pairs.

2.5.4 Jepson and Jenkin Algorithm

Jepson and Jenkin[42] have also proposed an algorithm based upon phase differences. Their earlier work also extensively applied phase analysis to stereoscopic computation which they expressed as a complex solution to non-linear differential equations. They have only recently appreciated the significance of phase subtraction. Their work follows a similar pattern to the work presented in this thesis, but they have not fully understood the significance of phase as the integral of spatial frequency. While recognising that it is essential to take an average spatial frequency to interpret disparity, they write " *where* w_l *and* w_r *is nearly the peak* *band-pass frequency of the Gabor pair* " which is unclear and imprecise.

Jepson and Fleet also applied Gabor filters with a spatial frequency domain bandwidth of 0.8 octaves. The Gaussian envelope employed by Jepson and Fleet was unnecessarily broad in the spatial domain, by the criterion proposed in this thesis (we apply Gabor filters at 1.3 octaves which borders the quadrature behaviour for Gabor filters). We have also observed that filters with excessive spatial extent are also unstable in their representation of phase contours in Scale-Space in comparison to filters with larger bandwidths. This could then account for the instability present in their algorithm. They also proposed a disparity gradient limit, however, their actual expression merely restricts the local disparity gradient to within the upper and lower cut-off frequency of the applied band-pass signal, which bears no relationship to the actual disparity gradient hypothesis used by Pollard et al[83]. From their paper, it is clear that they have not fully understood the method of phase differences although they did appreciate the biological plausibility and power with this form of algorithm. They also restricted their analysis to 1-D while at the same time inferring that 2-D extensions were straight forward. This is not true.

Results were presented to several random dot stereograms, and natural image data. The results were not impressive, and the phase locking technique that they implemented was highly unstable even after local averaging of disparity values.

2.6 Goals of this research

The goals of this research may be considered as an extension and improvement of the techniques that aim to combine the benefits from area and feature based stereo matching. This technique we will refer to as the *method of phase differences*.

While we have reviewed several existing techniques for the method of phase differences, we will:

- Consider the improvement of existing one dimensional techniques.
- Extend the one dimensional technique into two dimensions.
- Consider an application of the technique towards the computation of deformation.

Chapter 7 also considers some experimental work towards the detection of edges and corners with directionally selective filtering. While this section can be considered as a review of previous work (e.g [48]), we include some new ideas on corner detection. While only some of these ideas are applied in this thesis, we contend that local symbolic descriptions of image data[99] may be useful to verify that correct stereo matching has occured by a higher level of processing.

Chapter 3

FUNDAMENTAL THEOREMS

3.1 Introduction

Prior to developing a complete theory for 2-Dimensional stereoscopic correspondence, it is useful from both an experimental and understanding viewpoint to study the problem in the single dimension. Indeed, several algorithms rely heavily on the epipolar constraint in 1-D for correspondence[3]. However, several authors (e.g [60, 29]) have indicated the role of vertical disparity in stereoscopic computation. The biological application of this source of information remains an open question. However, the presence of any vertical disparity clearly indicates that stereoscopic matching in keeping with motion analysis, cannot be restricted to a single dimension in image understanding.

The fundamental hypothesis that forms the basis for the work presented in this thesis was first proposed by Gabor in 1946[25].

3.2 Gabor's Uncertainty Principle

The problem that Gabor addressed, was the simultaneous representation of a signal in both its own space, and frequency in the context of information transmission. He also proposed, that the principle could be applied to any linear system. Since stereopsis primarily involves the correspondence between image pairs at any single instance, it is useful to discuss the conceptual properties of stereopsis in signal space as a wavelength or spatial frequency in cycles per pixel (cpp). It is worth looking at Gabor's work in some depth, since his uncertainty principle is fundamental to Image Processing[98, 99]. Gabor was first interested in representing a real signal $s(x)$ in a different form comprising two rotating vectors in quadrature. Since image data is recorded by an intensity measurement from a camera sensor then this last statement bears significance. Gabor was interested in representing a real signal of the form:

$$s(x) \;=\; a\cos\omega x + b\sin\omega x \tag{3.1}$$

by the complex form:

$$\psi(x) \;=\; s(x) + j\sigma(x) \;=\; (a - jb)e^{j\omega x} \tag{3.2}$$

The function $\sigma(x) = a\sin\omega x - b\cos\omega x$ is formed from $s(x)$ which:

"represents the signal in quadrature to $s(x)$ which added to it, transforms the oscillating into a rotating vector"

In more simple terms Gabor expressed this transformation by the instruction " *Suppress the amplitudes belonging negative frequencies and multiply the amplitudes of positive frequencies by two*".

The uncertainty principle is of paramount importance in the context of signal processing. Wilson and Granlund[98] provide a well defined insight into the problem. They argue: "*one of the key tasks facing a "vision" system is to move from a pointwise image description: "the pixel at location* X *has magnitude m" to a description in terms of elementary events: "an event of class* C *is located at position* X ". They argued that it was the relative distribution of energy among the classes of events that should be used. From this standpoint they further argued that relative classes of events could not have arbitrary wide bandwidths since there would be no descrimination of events while an arbitraily large spatial extent would not permit spatial localisation. Thus there exists a physical restriction in the simulataneous representation of classes of events and spatial position. This restriction was formalised by Gabor:

3.2.1 Classical definition of Uncertainty

The measure of duration that Gabor used to formulate his Uncertainty hypothesis was based upon a minimisation of the second moment of a signal represented in both space and frequency:

$$d^2 = \frac{1}{E}\int_{-\infty}^{\infty} x^2|f(x)|^2\,dx \tag{3.3}$$

$$D^2 = \frac{1}{2\pi E}\int_{-\infty}^{\infty} \omega^2|F(\omega)|^2\,d\omega \tag{3.4}$$

By Parseval's formulae:

$$E = \int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} |F(\omega)|^2\,d\omega \tag{3.5}$$

He showed that providing $\sqrt{x}f(x) \rightarrow 0$ as $|x| \rightarrow \infty$ then:

$$Dd \geq \frac{1}{2} \tag{3.6}$$

This equation became an equality when:

$$f(x) = A\exp[-\alpha x^2] \tag{3.7}$$

or more generally:

$$\Psi(x,\omega) = \exp[-\alpha^2(x-x_o)^2]\; cis(\omega x+\phi) \tag{3.8}$$

Gabor, proposed to call the function of minimum area/bandwidth product an *elementary signal*. Gabor then proposed that an arbitrary signal could be expanded in terms of the elementary signals. To summarise briefly, Gabor's uncertainty principle states that a signal cannot be arbitrarily described in both space and frequency. This result, which is a consequence of the Fourier scaling theorem $f(ax) \rightarrow \frac{1}{|a|}F(\frac{\omega}{a})$ states that if $f(x)$ and $F(\omega)$ form a Fourier transform pair, then they cannot both be of short duration. Gabor's work, was primarily related to the to the Hilbert Transform. The properties of the Hilbert Transform are fundamental to the work in this thesis. We will therefore, review the definition and properties of the Hilbert Transform as a foundation for the remainder of this thesis.

3.2.2 Hilbert Transforms

Fundamental to theory of signal processing, and bandpass filtering lies the concepts associated with the *Hilbert Transform*. These functions are frequently used in signal analysis problems to create a complex signal $z_h(x)$, which is referred to as the *analytic signal* or *pre-envelope*. This function holds the important property of having its Fourier Spectra "one-sided", with the remainder of the information of the signal incorporated as a phase. The principle has formed the basis for much of the signal processing within RADAR[22]. We will review the important properties that we intend to consider in this thesis.

3.2.3 Fundamental Theorems

These follow from Papoulis[81] and Franks[22], and are useful properties of Hilbert Transform pairs $(h(x,\omega) + j\hat{h}(x,\omega))$. We define:

$$\Psi(x,\omega) * I(x) = h(x,\omega) + j\hat{h}(x,\omega) \tag{3.9}$$

as the convolution between a complex bandpass filter $(\Psi(x,\omega))$ and real image intensity function $(I(x))$. Thus $h(x,\omega)$ and $\hat{h}(x,\omega)$ are also bandpass functions.

Definition

We consider a real function $(h(x,\omega))$:

$$h(x,\omega) \xrightarrow{FT} H(\omega)$$

we form:

$$Z_h(\omega) = 2H(\omega)U(\omega) \qquad U(\omega) = \begin{cases} 1 & \omega > 0 \\ 0 & \omega < 0 \end{cases}$$

with its inverse transform:

$$z_h(x,\omega) = \frac{1}{\pi}\int_0^\infty H(\omega)\exp[j\omega x]d\omega$$

$z_h(x,\omega)$ is complex with real part $h(x,\omega)$ because:

$$h(x,\omega) = \frac{1}{\pi}\mathcal{R}e\int_o^\infty H(\omega)\exp[j\omega x]d\omega \tag{3.10}$$

The imaginary part of $z_h(x,\omega)$ denoted by $\hat{h}(x,\omega)$ is complex, and is known as the Hilbert transform of $h(x,\omega)$.

The frequency domain definition of Hilbert Transform pairs is perhaps the most pleasing to consider:

$$\hat{H}(\omega) \;=\; -j\text{sgn}\omega\; H(\omega)$$

and

$$H(\omega) \;=\; j\text{sgn}\omega\hat{H}(\omega)$$

from which the following properties are based:

Orthogonality

$$|h(x,\omega)| = |\hat{h}(x,\omega)| \tag{3.11}$$

$$< h(x,\omega), h(x,\omega) >= 0 \tag{3.12}$$

Differentiation

$$h^{(n)}(x,\omega) \overset{FT}{\rightarrow} (j\omega)^n H(\omega) \tag{3.13}$$

$$\frac{d[\Psi(x,\omega) * I(x)]}{dx} = \frac{d[\Psi(x,\omega)]}{dx} * I(x) \tag{3.14}$$

3.2.4 Modulus-Argument form

Pre-envelope

$$E(x,\omega) \;=\; \sqrt{h(x,\omega)^2 + \hat{h}(x,\omega)^2} \tag{3.15}$$

Phase

$$\Phi(x,\omega) = \tan^{-1}\frac{\hat{h}(x,\omega)}{h(x,\omega)} \tag{3.16}$$

Instantaneous Frequency

$$\begin{aligned} f_i(x) = & \tfrac{d\Phi(x,\omega)}{dx} = \mathcal{I}m[\tfrac{d}{dx}\ln z_h(x,\omega)] \\ & = \tfrac{1}{2\pi}\tfrac{[\dot{\hat{h}}(x,\omega)h(x,\omega)-\hat{h}(x,\omega)\hat{h}(x,\omega)]}{E^2(x,\omega)} \end{aligned} \tag{3.17}$$

with the condition that the four quadrature filters are independent if:

$$W(x,\omega) = \det\begin{vmatrix} h(x,\omega) & \hat{h}(x,\omega) \\ \hat{h}(x,\omega) & \dot{\hat{h}}(x,\omega) \end{vmatrix} \neq 0 \tag{3.18}$$

where $W(x,\omega)$ is the Wronskian of $E(x,\omega)$ evaluated at x.

Pre-envelope weighted Instantaneous Frequency

$$f_{iav}(x,\omega) = \frac{\sum_{-s}^{s}\det|W(s-x,\omega)|}{\sum_{-s}^{s}E(s-x,\omega)} \tag{3.19}$$

Correlation

$$\cos\theta = \frac{<\mathbf{h_1},\mathbf{h_2}>}{|\mathbf{h_1}|^{\frac{1}{2}}|\mathbf{h_2}|^{\frac{1}{2}}} \tag{3.20}$$

where $\mathbf{h_1}$ and $\mathbf{h_2}$ are Hilbert Transform pairs which we have in this case represented by a vector.

3.3 Biological evidence

Marcelja[64] was the first to recognise the similarity between the receptive field profiles of simple cells in the mammalian striate cortex, and the even or odd elementary signals of Gabor (fig. 3.1). The interest of these particular functions in the context of Image Processing is two fold[31]. First, theoretical analysis suggests that these functions are ideally suited for obtaining the best compromise between the position of a signal in space and for identifying the frequency component of the signal. Secondly, the Gabor function conveniently provides a mechanism for describing a spatial filter in spatial frequency, phase and envelope. Many experimentalists have confirmed Marcelja's observation that there is at least "first approximation" to the Gabor function and receptive field profile of simple cells.

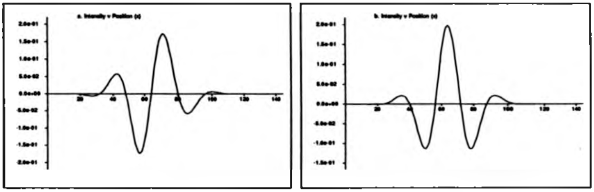


Figure 3.1: *Receptive field properties of (a) Imaginary part (b) Real Gabor function.* $\sigma_x = \frac{2}{\omega}$

Field and Tolhurst[20], however, expressed some doubt to the notions that simple cells generally fit into even and odd symmetric catagories. However, this is not an important property. It suffices that the phase difference between filter pairs is in quadrature. This we will discuss in chapter 5. Daugman[18] showed that the error between the receptive field profiles of orientated complex cells and Gabor functions was less than would be expected from experimental error. However, the parameters that Daugman used to describe the Gabor function in his comparison, showed no appreciable sidelobes and we can conjecture that the parameters that Daugman used may not have described the Gabor function under linear phase conditions. In this case, the imaginary part of the Gabor function and its Hilbert transform becomes very similar to the Hilbert pair of the 1st derivative of the Gaussian distribution. This can be seen by the polynomial expansion of the imaginary part of the Gabor ($\Psi(x,\omega,\sigma)$) function:

$$\Psi(x,\omega,\sigma) = k\exp[-\frac{x^2}{4\sigma^2}]\sin\omega x = k\exp[-\frac{x^2}{4\sigma^2}][\omega x + \frac{(\omega x)^2}{2!} + \ldots.\,] \qquad (3.21)$$

where ignoring second and higher terms because $\frac{1}{\omega} >> \sigma$:

$$\Psi(x,\omega,\sigma) \approx k\exp[-\frac{x^2}{4\sigma^2}]\omega x \qquad (3.22)$$

This leaves some uncertainty regarding the appropriateness of the Gabor function as a true biological model.

3.4 Summary

We have discussed briefly the role of the Uncertainty Principle in Image Processing, and the properties of Hilbert Transform Pairs. We have also discussed the presence of these operators in Biological systems. We will now formulate the principles that we have discussed into a mechanism for the measurement of disparity.

Chapter 4

DISPARITY MEASUREMENT

4.1 Introduction

The essence of the approach, is based primarily upon the the Fourier shift theorem:

$$f(x-d) \rightarrow F(\omega)\exp[jd\omega] \tag{4.1}$$

Where d represents the displacement in one signal domain relative to the other, and is therefore representative as a disparity. However, because we are adding a complex signal in our representation of an image by convolution with a quadrature pair of filters, a proper and more general representation of the equations that require a solution are:

$$z_l(x,\omega) = E_l(x,\omega)\exp[j\phi_l(x,\omega)] \tag{4.2}$$

and

$$z_r(x,\omega) = E_r(x,\omega)\exp[j\phi_r(x,\omega)] \tag{4.3}$$

Where $z(x,\omega)$ implies the spatial convolution between the quadrature filter $\Psi(x,\omega)$ and a 1-D image signal $I(x)$ i.e ($z(x,\omega) = \Psi(x,\omega) * I(x)$). Since $\Psi(x,\omega)$ is complex, then $z(x,\omega)$ is also complex and we may represent the outcome of the convolution in modulus-Argument form, where $E(x,\omega)$ represents the pre-envelope, and $\phi(x,\omega)$ the argument of the left (l) and right (r) image respectively. We define a phase difference ($\phi_d(x,\omega)$) equation by:

$$\phi_d(x,\omega) = \phi_r(x,\omega) - \phi_l(x,\omega) \tag{4.4}$$

Where $\phi_d(x,\omega)$ is the measure used to obtain disparities between two signals separated by spatial position.

The notion that phase might be used to calculate disparity is not new. Larcombe [58] and Wilson[100] have previously used phase differences to estimate image disparities. Indeed, as early as 1967, Lange [51] wrote:

" *if it is possible to allot a mean frequency, with defined phase position to a narrow spectrum of a fluctuating process, then it is also possible to allocate a certain reciprocal phase position to two fluctuating processes of the same frequency band.* " His method, an extension of Montgomery [74] was based upon the cross-correlation of two narrow band filtered random processes $x_1(t)$ and $x_2(t)$:

$$\phi_{12} = \lim_{T\rightarrow\infty}\frac{1}{T}\int_{-T}^{T} x_1(t)x_2(t+\tau)d\tau \tag{4.5}$$

For which the maximum correlation for a value τ which he called τ_{opt} was equated to the time delay difference of the correlated time functions. Since the two signals were band-passed filtered then:

$$\Delta\phi \;=\; 2\pi f_o \tau_{opt} \tag{4.6}$$

which therefore retrieves the phase difference. He failed to appreciate, however, that a quadrature pair of filters can directly compute the phase difference. In the proposed algorithm, phase differences are used to estimate the displacement of a pair of similar signals in the signal space, a subtle difference from the work of Lange.

4.2 Evidence from Psychophysics

There is considerable evidence from neurophysiological data that bipolar cells in the visual cortex are both orientational and phase sensitive (e.g [63]). Thus it is possible that the visual system is sensitive to phase differences which can be used to measure disparity. Such a mechanism would be a low-level process. Mayhew and Frisby[71] argued against this notion.

In their experiment, they used a 2-D circularly symmetric band pass filter and introduced a disparity difference greater than the Nyquist limit of the filtered image. Since they observed no difference in depth perception, they concluded that the human visual system did not utilise spatial frequency analysis to obtain depth perception. A 2-D circularly symmetric filter, however, will not remove all the low frequency signal components since in the 2-D Fourier plane $F(u, v)$, a spatial frequency $|\omega_o|$ can be resolved into:

$$\omega_o \;=\; (u^2 \;+\; v^2)^{\frac{1}{2}} \tag{4.7}$$

and ω_o can still comprise of low spatial frequency elements along the horizontal axis which is the primary orientation that we would expect to extract disparity information. The results of Mayhew and Frisby do not exclude, but may be reinterpreted to support the mechanisms of the type proposed. While it is possible, that *feature based stereopsis* is an integral part of human depth perception, such a mechanism would not be a low-level vision process under our definition.

4.3 Phase from Gabor filters

First let us consider two sinusoidal signals, with equal frequency but a constant phase difference. This may be perceived as a rotation of the original signal by an angle that contains the phase difference. By obtaining an estimate of both the spatial-frequency and the phase difference between the signals it is possible to compute the displacement of the second signal.

Consider a pair of quadrature Gabor filters centered at a pre-determined spatial frequency and similarly positioned in a pair of images. Consider the convolution of this operator with a sinusoid grating. Interpreting the convolution in the spatial-frequency domain, the result will be an impulse with a corresponding magnitude related to the displacement of the sine grating from the center frequency of the Gabor function. In the shifted image, there will be a response that incorporates the phase displacement of the signal, and an associated energy response which is not necessarily linear with respect to the first image. To extract the phase difference between the two image functions, it is necessary to solve four convolutions[52]:

$$I_s(x,u_o,u_g,\sigma) = \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp[\frac{-(x-X)^2}{4\sigma^2}] \cos 2\pi u_g(x-X)\, I(X)\, dX \qquad (4.8)$$

here x refers to image position and u_g refers to the center frequency of the Gabor filter.

$$I_a(x,u_o,u_g,\sigma) = \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp[\frac{-(x-X)^2}{4\sigma^2}] \sin 2\pi u_g(x-X)\, I(X)\, dX \qquad (4.9)$$

$$I_{s\phi}(x,u_o,u_g,\sigma) = \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp[\frac{-(x-X)^2}{4\sigma^2}] \cos 2\pi u_g(x-X)\, I(X+d)\, dX \qquad (4.10)$$

$$I_{a\phi}(x,u_o,u_g,\sigma) = \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp[\frac{-(x-X)^2}{4\sigma^2}] \sin 2\pi u_g(x-X)\, I(X+d)\, dX \qquad (4.11)$$

Let $I(X) = \cos 2\pi u_o X$ and $I(x+d) = \cos 2\pi u_o(X+d) = \cos 2\pi u_o x + \phi$ represent our 1-D pair of image functions. Here the disparity introduced is d pixels with the phase difference ϕ. Expanding equations (4.10) and (4.11) , we find:

$$I_{s\phi}(x,u_o,u_g,\sigma) = \cos\phi\ I_s(x,u_o,u_g,\sigma) - \sin\phi\ I_1(x,u_o,u_g,\sigma) \qquad (4.12)$$

where $I_1(x, u_o, u_g, \sigma)$ is:

$$I_1(x, u_o, u_g, \sigma) = \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp[\frac{-(x-X)^2}{4\sigma^2}] \cos 2\pi u_g(x - X) \sin 2\pi u_o X \, dX \quad (4.13)$$

and

$$I_{a\phi}(x, u_o, u_g, \sigma) = \sin\phi \, I_2(x, u_o, u_g, \sigma) - \cos\phi(x, u_o, u_g, \sigma) \, I_a(x, u_o, u_g, \sigma) \quad (4.14)$$

where:

$$I_2(x, u_o, u_g, \sigma) = \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp[\frac{-(x-X)^2}{4\sigma^2}] \sin 2\pi u_g(x - X) \sin 2\pi u_o X \, dX \quad (4.15)$$

By taking the Fourier transform of I_s and observing that

$$F(x) * G(x) \overset{\mathrm{FT}}{\rightarrow} F(u)G(u)$$

we arrive at the following equation:

$$F(u) = (8\pi\sigma^2)^{\frac{1}{2}} \left(\exp[-4\pi^2\sigma^2(u - u_g)^2] + \exp[-4\pi^2\sigma^2(u + u_g)^2]\right) \frac{1}{2}(\delta(u - u_o) + \delta(u + u_o)) \quad (4.16)$$

The inverse Fourier transform of the above equation effectively provides a solution to the integral in (4.8):

$$I_s(x, u_o, u_g, \sigma) = (8\pi\sigma^2)^{\frac{1}{2}} \left(\exp[-4\pi^2\sigma^2(u_o - u_g)^2] + \exp[-4\pi^2\sigma^2(u_o + u_g)^2]\right) \cos 2\pi u_o x \quad (4.17)$$

By applying the same technique to equations (4.9),(4.13) and (4.15) then we arrive at the following expressions:

$$I_a(x, u_o, u_g, \sigma) = (8\pi\sigma^2)^{\frac{1}{2}} \left(\exp[-4\pi^2\sigma^2(u_o - u_g)^2] - \exp[-4\pi^2\sigma^2(u_o + u_g)^2]\right) \sin 2\pi u_o x \quad (4.18)$$

$$I_1(x, u_o, u_g, \sigma) = (8\pi\sigma^2)^{\frac{1}{2}} \left(\exp[-4\pi^2\sigma^2(u_o - u_g)^2] + \exp[-4\pi^2\sigma^2(u_o + u_g)^2]\right) \sin 2\pi u_o x \quad (4.19)$$

$$I_2(x,u_o,u_g,\sigma) = (8\pi\sigma^2)^{\frac{1}{4}}\left(-\exp[-4\pi^2\sigma^2(u_o-u_g)^2] + \exp[-4\pi^2\sigma^2(u_o+u_g)^2]\right)\cos 2\pi u_o x \tag{4.20}$$

By forming the simple ratio $\frac{I_1(x,u_o,u_g,\sigma)}{I_a(x,u_o,u_g,\sigma)}$ and simplifying then:

$$\frac{I_1(x,u_o,u_g,\sigma)}{I_a(x,u_o,u_g,\sigma)} = \frac{\exp[8\pi^2\sigma^2u_ou_g] + \exp[-8\pi^2\sigma^2u_ou_g]}{\exp[8\pi^2\sigma^2u_ou_g] - \exp[-8\pi^2\sigma^2u_ou_g]} \tag{4.21}$$

i.e

$$I_1(x,u_o,u_g,\sigma) = \frac{I_a(x,u_o,u_g,\sigma)}{\tanh[8\pi^2\sigma^2u_ou_g]} \tag{4.22}$$

Using a similar method it can also be shown that:

$$I_2(x,u_o,u_g,\sigma) = -I_s(x,u_o,u_g,\sigma)\tanh[8\pi^2\sigma^2u_ou_g] \tag{4.23}$$

We may now substitute the above two equations into equations (4.12) and (4.14) from which the phase may be represented in terms of a normal rotation matrix $R_{\phi\zeta}$ where $\zeta = \tanh[8\pi^2\sigma^2u_ou_g]$ i.e

$$\begin{bmatrix} I_{s\phi}(x,u_o,u_g,\sigma) \\ I_{a\phi}(x,u_o,u_g,\sigma) \end{bmatrix} = \begin{bmatrix} \cos\phi & -\frac{\sin\phi}{\zeta} \\ \sin\phi\zeta & \cos\phi \end{bmatrix} \begin{bmatrix} I_s(x,u_o,u_g,\sigma) \\ I_a(x,u_o,u_g,\sigma) \end{bmatrix} \tag{4.24}$$

Here, we notice that the non-linearity in phase incorporates a shearing of the transformation matrix. Improvements in computational stability may be achieved by removing ζ from the rotation matrix. The concise solution to the equation becomes:

$$\begin{bmatrix} I_{s\phi}(x,u_o,u_g,\sigma) \\ \frac{I_{a\phi}(x,u_o,u_g,\sigma)}{\zeta} \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} I_s(x,u_o,u_g,\sigma) \\ \frac{I_a(x,u_o,u_g,\sigma)}{\zeta} \end{bmatrix} \tag{4.25}$$

and

$$\phi = \tan^{-1}\left(\frac{\frac{I_{a\phi}(x,u_o,u_g,\sigma)}{\zeta}}{I_{s\phi}(x,u_o,u_g,\sigma)}\right) - \tan^{-1}\left(\frac{\frac{I_a(x,u_o,u_g,\sigma)}{\zeta}}{I_s(x,u_o,u_g,\sigma)}\right) \tag{4.26}$$

With the disparity estimate (D_{est}) given by:

$$D_{est} = \frac{\phi}{2\pi u_g} = \frac{u_o}{u_g}d \text{ pixels} \tag{4.27}$$

If we make a further examination of the tanh correction factor then we see that:

$$8\pi^2u_ou_g\sigma^2 = 8.0 \tag{4.28}$$

and

$$u_o = u_g \tag{4.29}$$

then:

$$\tanh[8\pi^2 u_o u_g \sigma^2] \approx 1.00 \tag{4.30}$$

If we consider these conditions in the frequency domain then:

$$4\sigma_u = u_g \tag{4.31}$$

where $\sigma_u = u_o$ represents the standard deviation of the filter in frequency space.

4.3.1 Filter Bandwidth

Consider a Gaussian function $G(u, \sigma_u)$ centered at the origin of the frequency domain:

$$G(u, \sigma_u) = K \exp[-\frac{u^2}{4\sigma_u^2}] \tag{4.32}$$

its half height is found from:

$$0.5 = \exp[-\frac{u^2}{4\sigma_u^2}]$$

i.e

$$u_{\frac{1}{2}} = 1.665\sigma_u = \frac{1.665 u_g}{4}$$

by the Fourier shift theorem:

$$g(x, \sigma_x) \cos u_g x \overset{FT}{\rightarrow} \frac{1}{2\pi}[G(u - u_g, \sigma_u) + G(u + u_g, \sigma_u)]$$

thus the Gabor function is effectively a Gaussian envelope centered at u_g in the frequency domain. The bandwidth in octaves (B_{oct}) is then defined by:

$$B_{oct} = \log_2 \left[\frac{u_o + u_{\frac{1}{2}}}{u_o - u_{\frac{1}{2}}} \right] = 1.28 \tag{4.33}$$

Clearly, if $\zeta \approx 1.0$ then the Gabor function may be considered as a linear phase filter. The bandwidth of a Gabor filter constrained by this limit of linearity is equal to 1.28 Octaves. Several authors [19] who studied cat cortical simple cells, showed that they had a mean bandwidth of 1.2 to 1.47 octaves. Therefore, the visual system may apply Gabor filters at the limits of phase linearity. It is noticeable, that if $\zeta < 1$ then Gabor filters are not in quadrature. This is represented by equations (4.17) and (4.18).

4.4 Instantaneous Frequency

We now consider the estimation of instantaneous frequency using Gabor functions. First, consider the ratio of $I_a(x, u_o, u_g, \sigma)/I_s(x, u_o, u_g, \sigma)$ from equations (4.17) and (4.18), which may be defined as the instantaneous phase of a signal. We may express this as:

$$\phi(x) = \tan^{-1}\left(\frac{I_a(x, u_o, u_g, \sigma)}{I_s(x, u_o, u_g, \sigma)}\right) = \tan^{-1}\left(\frac{\tan[2\pi u_o x]}{\zeta}\right) \tag{4.34}$$

Since the right hand side of the above equation is independent of the center frequency from the Gabor filter (assuming $\zeta = k$). It can easily shown that:

$$\frac{d[\phi(x)]}{dx} = \frac{2\pi u_o}{\frac{\sin^2[2\pi u_o x]}{\zeta} + \zeta \cos^2[2\pi u_o x]} \tag{4.35}$$

which follows from equation (3.17). Providing $\zeta = 1.0$ then the spatial derivative of the phase response may be considered as an estimate of the instantaneous frequency of a signal. Notice that in this case, the condition $\zeta \neq 1$ introduces oscillatory behaviour into the estimation of instantaneous frequency, whose magnitude is determined by ζ.

4.4.1 Local measurement of Instantaneous frequency

From equations (3.17) and (3.14) we observe that the local instantaneous frequency can be numerically calculated using the spatial derivatives of the Gabor function ($G_{rd}(x, u_g, \sigma)$) and $G_{id}(x, u_g, \sigma)$:

$$G_{id}(x, u_g, \sigma) = \frac{-1}{(2\pi\sigma)^{\frac{1}{4}}} \exp[\frac{-x^2}{4\sigma^2}](\frac{2x}{4\sigma^2} \cos 2\pi u_g x + 2\pi u_g \sin 2\pi u_g x\ ) \tag{4.36}$$

and

$$G_{rd}(x, u_g, \sigma) = \frac{1}{(2\pi\sigma)^{\frac{1}{4}}} \exp[\frac{-x^2}{4\sigma^2}](\frac{-2x}{4\sigma^2} \sin 2\pi u_g x + 2\pi u_g \cos 2\pi u_g x\ ) \tag{4.37}$$

by:

$$f_i(x) = \frac{1}{2\pi} \frac{[I_s(x, u_o, u_g, \sigma) I'_a(x, u_o, u_g, \sigma) - I'_s(x, u_o, u_g, \sigma) I_a(x, u_o, u_g, \sigma)}{I_s^2(x, u_o, u_g, \sigma) + I_a^2(x, u_o, u_g, \sigma)} \tag{4.38}$$

where $I'_a(x, u_o, u_g, \sigma) = G_{id}(x, u_g, \sigma) * I(x)$ and $I'_s(x, u_o, u_g, \sigma) = G_{rd}(x, u_g, \sigma) * I(x)$.

4.5 Error in Disparity measurement

An examination of equation (4.27), indicates that the phase shift of a signal in pixel units, is a function of the signal frequency. Therefore, to be exact in terms of the pixel shift, we require the fundamental frequency of the signal under analysis.

One method to find the fundamental signal under analysis would be to apply a range of filters at the same location in the image space, but tuned to different spatial frequencies[81]. The filters which then lie in closest proximity to the fundamental frequency of the signal will provide the greatest accuracy in disparity interpretation and also respond with the largest energy.

In general, there will be an error in the calculated phase shift that will be related to the difference between the Gabor center frequency and the examined image frequencies. This is because we are subsampling in both the spatial and frequency domains. The criterion for subsampling follows from the theoretical work of Gabor and Bastians [25, 4]. We propose subsampling in the spatial frequency domain by $\sqrt{2}nu_s$, where u_s is the coarsest filter applied. By applying the filters in a geometric progression in the manner prescribed, filters are localised in the spatial frequency spectrum at low spatial frequencies and less selective at higher spatial frequencies. The sampling in the frequency domain can be seen in figure 4.1. Here, the Fourier Transform of each filter has been superimposed to indicate the sampling of the frequency domian. It is apparent that filters tuned to the lower spatial frequencies are far more selective than the filters tuned to higher spatial frequencies.

If disparity was interpreted using a local energy peak from a series of filters similarly applied to a point in the image, the greatest error will occur when the image fundamental frequency lies at the mid point between an adjacent pair of filters in the frequency domain. By using the fundamental frequency of the filter to interpret the phase difference we expect that the filter tuned to a slightly lower spatial frequency will over estimate the disparity measurement, while the filter tuned to a higher spatial frequency will accordingly under estimate the image displacement.

An estimation of the maximum error from equation (4.27) indicates that in this case, there will be an upper and lower frequency error of 21% and 15% of the actual disparity this follows from equation (4.46).

However, since the magnitude of error is dependent upon the phase displacement,

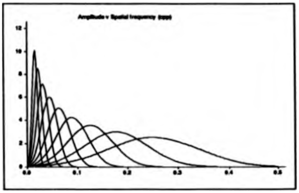


Figure 4.1: *Fourier Spectrum of Gabor filters separated by* $\sqrt{2}nu_s$ *in the frequency domain. Here* u_s = *1.0/64.0 cpp. Each filter was chosen to approximate a quadrature filter from equation (4.31). y-axis shows the filter amplitude against frequency (x-axis).*

by iteration we would expect to converge to a better estimate in a coarse to fine strategy. At high spatial frequencies large shifts in the image are also restricted by aliasing. Hence, low spatial frequency filters can tolerate large disparities with high spatial uncertainty, while high spatial frequency filters must be localised to correctly interpret disparity. The previous argument, suggests that filters should be ordered with low frequency channels providing coarse estimates of disparity, with higher frequency channels giving progressively finer estimates. Such a process may well be the mechanism underlying the control of convergence of objects near to the observer.

It is interesting to note that for loci of zero disparity that lie on the notional horopter[79], zero disparity (in the absence of noise) may be detected by any spatial filter without error.

In the same manner as Wilson and Knutsson[100] we have assumed that the instantaneous frequency of the stereoscopic image pairs are similar. Such a generality cannot possibly be maintained without serious error, particularly when considering the complexity of natural image data. A similar aspect of stereopsis follows from Clark and Lawrence [17] who studied the error involved with different spatial frequency elements in the image domain. They restricted their analysis to zero-crossings of Marr's Laplacian. They came to the conclusion that there was no error in disparity estimation providing spatial frequency components were equal and fused. We have already reported similar findings at zero disparity. Consider the interpretation of disparity from signals with different spatial frequencies. By assigning a phase difference as before we can write:

$$\phi_d(x) = \omega_l(x + d_l) - \omega_r(x + d_r) \quad \phi_d \in [-\pi, \pi] \tag{4.39}$$

It is obvious that in the case where $\omega_l = \omega_r$ then interpretation of disparity ($d = d_l - d_r$) is simple. When this condition is not met, interpretation can be difficult.

Consider a complex stereo signal, with a partially sighted surface in one image A set of phase sensitive filters might respond with a complete phase wrap around cycle caused by this feature. If a process merely responded to phase differences in an exact and precise manner by using equation (4.27) for instance, then tracking the difference could lead gross error. This is because it would be difficult to separate classical diffrequency[7] (diffrequency refers to the local differences in spatial frequency between stereoscopic views of the world, which implies a disparity gradient) from depth discontinuities.

Let us re-define the phase difference at a position (x) as the roots of:

$$\phi_l(x) - \phi_r(x_o) = 0 \tag{4.40}$$

Suppose that x is the exact root that we require. Expanding the above equation as a Taylor series with $x = x_o + d$ as an exact solution we have:

$$\phi_l(x_o + d) - \phi_r(x_o) = 0 \tag{4.41}$$

$$= \phi_l(x_o) - \phi_r(x_o) + d\phi_l'(x_o) + \ldots\ldots$$

let us then make d_1 an approximation to d in which case we have:

$$d_1 = -\frac{-[\phi_l(x_o) - \phi_r(x_o)]}{\phi_l'(x_o)} \tag{4.42}$$

from which we iterate to find the root of our initial equation by:

$$x_n = x_{n-1} - \frac{\phi_l(x_{n-1}) - \phi_r(x_o)}{\phi_l'(x_{n-1})} \tag{4.43}$$

which is in fact Newton-Raphson convergence with d_n as our current disparity estimate. Thus we can can also interpret disparity using the local derivative of phase (instantaneous frequency)[55]. We can immediately state[40] that convergence from this method can only be expected providing the new estimation of the root x_n lies between the previous estimate x_{n-1} and the exact solution x.

It would be expected that the inverse rotational phase transformation from equation (4.42) will also give the same disparity measurement independently of whether we choose to match the left and right image. However, this would only be correct assuming that $\phi_l'(x) = \phi_r'(x)$. We can add the constraint that the derivative of phase or instantaneous frequencies from both image pairs must at least be of the same sign for convergence to be established. One advantage of a Newton-Raphson iteration lies with its convergence properties. If instead of considering sinusoidal signals we were to consider a phase function $Arg[\Psi(x;\omega_g) * N(x)]$ where $N(x)$ is broadband Gaussian noise[10] (note: $2\pi u_g = \omega_g$), then:

$$\phi_l(x,\omega) = \omega_g x + \Phi(x) \tag{4.44}$$

with the constant disparity displacement in the corresponding stereo pair given by:

$$\phi_r(x,\omega) = \omega_g[x - d] + \Phi(x - d)$$

the phase difference at x_o becomes:

$$\phi_d(x_o) = \phi_l(x_o) - \phi_r(x_o - d) \ = \ \omega_g d + \Phi(x_o) - \Phi(x_o - d) \tag{4.45}$$

thus the disparity error (ϵ) in interpreting disparity from the center frequency of the filter becomes:

$$\epsilon = \frac{\Phi(x_o) - \Phi(x_o - d)}{\omega_g} \tag{4.46}$$

Taylor expanding the above equation at x_o we have:

$$\epsilon = \frac{1}{\omega_g}[d\Phi'(x_o) + d^2\Phi''(x_o) + \ \ldots \ ]$$

If we now substitute the local phase gradient by its numerical determination from equation (4.38), then the error in disparity measurement (assuming convergence) becomes:

$$\epsilon = d^2\frac{\Phi''(x_o)}{\Phi'(x_o)} + \ \ldots \tag{4.47}$$

which is indicative of the quadratic convergence of the Newton-Raphson iteration.

By observing that there should exist an equivalent inverse rotation between our vector pairs, we immediately introduce an ambiguity should we wish to use the derivative of phase to interpret disparity. This is because we may use the instantaneous frequency properties from either image as the Newton-Raphson method informs us. By the established theories of Continuum Mechanics, this is equivalent to considering the *Eulerian* and *Langrangian* description of a deformation field. Assuming small deformations we can consider a simple mean ($f_{av}(x)$):

$$f_{av}(x) = \frac{f_l(x) + f_r(x)}{2} \tag{4.48}$$

were $f_l(x)$ and $f_r(x)$ refer to the local instantaneous frequency of both image pairs.

Indeed, Eric Grimson (cited in Marr[65]) pointed out that matching can be achieved from either or both eyes. Interestingly, because of the ambiguity, one might consider the implications of this result with regard to eye dominance in human vision, since we would in principle only require the instantaneous frequency properties from a single image to interpret disparity. To reduce the error in disparity estimation, we propose taking a weighted average between the instantaneous frequency estimates. Our improved interpretive frequency is thus defined as:

$$f_{av}(x) = \frac{E_l(x)f_l(x) + E_r(x)f_r(x)}{E_l(x) + E_r(x)} \tag{4.49}$$

from which we weight the frequency for disparity interpretation to the dominant energy response ($E(x)$) from either image pair. Here $f_l(x)$ and $f_r(x)$, refer to the instantaneous frequencies of the left and right image pair. We also note that it is possible to apply the weighted pre-envelope definition of instantaneous frequency ($\bar{f}_{av}(x)$) from equation (3.19) to reduce noise sensitivity in the estimation of the local phase behaviour:

$$\bar{f}_{av}(x) = \frac{1}{2\pi} \frac{\sum_{-s}^{s} \det|W_r(s-x,\omega)| + \sum_{-s}^{s} \det|W_l(s-x,\omega)|}{\sum_{-s}^{s} E_l(s-x,\omega) + \sum_{-s}^{s} E_l(s-x,\omega)} \tag{4.50}$$

4.6 The Pre-envelope

The pre-envelope response may be calculated as:

$$E^2(x,u_o,u_g,\sigma) = I_s^2(x,u_o,u_g,\sigma) + I_a^2(x,u_o,u_g,\sigma) \tag{4.51}$$

$$= (8\pi\sigma^2)^{\frac{1}{2}} [\exp(-8\pi^2\sigma^2(u_o-u_g)^2)+\exp(-8\pi^2\sigma^2(u_o+u_g)^2)+2\exp(-4\pi^2\sigma^2(u_o^2+u_g^2))\cos 4\pi u_o x] \tag{4.52}$$

An examination of the above equation indicates that the energy response is influenced by three factors:

- The separation of Gabor center frequency and signal frequency.

- The displacement of Gabor and signal frequency from the d.c level.
- The phase of the signal at the sampled point.

Thus, the pre-envelope will have maxima when:

$$\cos 4\pi u_o x = 1 \quad i.e\ x \ = \ 0, \frac{\pi}{2}, \ldots \tag{4.53}$$

and $u_o \ = \ u_g$, this gives us a maximum energy response ($E_{max}(x, u_o, u_g, \sigma)$):

$$E^2_{max}(x, u_o, u_g, \sigma) \ = \ (8\pi\sigma^2)^{\frac{1}{2}}[1 + \exp(-32\pi^2\sigma^2u_g^2) + 2\exp(-16\pi^2\sigma^2u_g^2)] \tag{4.54}$$

Therefore, the pre-envelope maximum is approximately proportional to σ. In this representation low frequency signals will respond with greater energy than higher frequencies. This is a convenient condition, in the respect that an image will in general contain a broad range of instantaneous frequencies. Thus low frequency signals, which can tolerate the largest pixel shifts without aliasing have the greatest probability of maximal response.

We observe that the derivative of the pre-envelope response with respect to displacement is equal to:

$$\frac{dE^2(x, u_o, u_g, \sigma)}{dx} \ = \ (8\pi)^{\frac{3}{2}} u_o\sigma \exp(-4\pi^2\sigma^2(u_o^2 + u_g^2)) \sin[4\pi u_o x] \tag{4.55}$$

Which has a corresponding maximum and minimum dependent on whether $\sin 4\pi u_o x \ = \ 0$. We should also consider the d.c sensitivity of the filter. Indeed, this is a special case which can be particularly troublesome. From equations (4.17) and (4.18) we observe that the imaginary component (i.e $I_o(x, u_o, u_g, \sigma)$) is not sensitive to d.c. However, the real part of the Gabor function holds a d.c bias explained by equation (4.17)[46]:

$$I_{s;u_s=0}(x, u_o, u_g, \sigma) \ = \ 2A(8\pi\sigma^2)^{\frac{1}{4}} \exp[-4\pi^2\sigma^2u_g^2] \tag{4.56}$$

where A represents the amplitude of the d.c bias.

Naturally, this expression must be included in *both* our definitions of phase and energy response as :

$$Arg[\Phi(x) * I(x)] \ = \ \frac{\sinh[8\pi^2\sigma^2u_ou_g]\sin[2\pi u_o x]}{\cosh[8\pi\sigma^2u_ou_g]\cos[2\pi u_o x] \ + \ 2A\exp[-4\pi^2\sigma^2u_g^2]} \tag{4.57}$$

thus an image function with d.c bias will not necessarily respond with linear phase. With regard to the energy response, we should merely note the addition of the extra terms $I^2_{x;u_o=0}$ and:

$$4A\exp[-4\pi^2\sigma^2u_g^2]\cosh[8\pi\sigma^2u_ou_g]\cos[2\pi u_ox]$$

which one would expect to oscillate in phase with the signal under view.

4.6.1 Instantaneous Frequency analysis from the Energy response.

Papoulis[81] suggests that localised instantaneous frequency estimation can be obtained by applying multiple bandpass filters, and estimating a mean spatial frequency based upon the energy response from each filter. Unfortunately, this method also requires a suitable scaling factor to compensate for the energy differences that we might expect through the resolutions of bandpass filtering. Because of the uncertainty regarding a suitable weighting measure, this particular approach was not considered in this thesis.

We now clarify what we interpret by energy. We see from the Gabor function, that the magnitude of the energy response is determined by the phase of the signal under view, and its departure from the preferred frequency of the applied filter. The former term being the dominant parameter that is observed under spatial convolution, i.e:

$$dE = \frac{\partial E}{\partial x}dx + \frac{\partial E}{\partial \omega_o}d\omega_o \tag{4.58}$$

The above equation is rather unusual. Traditionally, spatial derivatives of energy[93, 12] are not separated into these two components. We will later show a method for obtaining this term from the first derivatives of a convolution operator.

4.7 Designing the Quadrature filter

From equation (4.17) we notice that the Gabor function retains some sensitivity at d.c (illuminance sensitivity). By choosing the bandwidth of the Gabor function at 1.28 octaves, we have assumed that the d.c sensitivity is negligible. To enforce this condition, we make the observation that imaginary part (sinusoidal modulation of a Gaussian) is not sensitive to d.c. Therefore, by creating the Hilbert transform pair of the imaginary part we can ensure that the Gabor function is a quadrature filter with no sensitivity at d.c. The loss in quadrature can be seen by comparing equations (4.17) and (4.18) where $|h(x)| \neq |\hat{h}(x)|$ but $< h(x), \hat{h}(x) >= 0$.

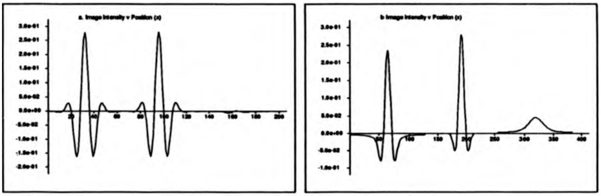


Figure 4.2: *(a) Gabor filter characteristics chosen for linear phase. i.e* $\sigma_x = \frac{2}{\omega_g}$*. Left : Hilbert Transform of imaginary component, Center: Real part of Gabor, Right: difference. (b) Gabor filter characteristics chosen under non-linear phase conditions for* $\sigma_x = \frac{1}{\omega_g}$*. Graphs show amplitude (y-axis) against spatial position (x-axis).*

Clearly in the Gabor representation this first condition is not exactly satisfied because of the energy differences involved with sinusoidal and cosinusoidal modulation. Naturally, when the filter pairs are tuned to the image signal's fundamental form, the Gabor filter pairs may be assumed to exhibit the properties of a Hilbert pair, but this is not generally applicable over all the possible range of spatial frequencies. Gabor did not point this out in his original paper. When the parameters of the Gabor filter are designed to operate towards linear phase, we would expect there to be a small difference between the Hilbert Transform of the imaginary part of the Gabor function, and the real part of the Gabor. When this condition is not met, we expect marked differences as indeed is the case (fig. 4.2). The differences are primarily caused by a d.c component which we notice from the Gaussian distribution of the difference, which we would predict from equations (4.17) and (4.18).

4.8 Compression/Expansion Transformations

Within the 1-D framework we are able to discuss the compression/expansion transformations of a vertically slanted smooth surface (wallplane). In the spatial domain this transformation can be represented by frequency modulation. Let us consider our original sinusoid signal $\sin 2\pi u_0 x$. Let the displacement of a second sinusoidal signal be proportional to the square of the displacement relative to some arbitrary fixed point at $x = 0$. Then:

$$f_l(x) = \sin 2\pi u_o x \tag{4.59}$$

and

$$f_r(x) = \sin 2\pi x(u_o + \beta x) \tag{4.60}$$

Let:

$$u(x) = u_o + 2\beta x \tag{4.61}$$

where $u(x)$ is the instantaneous frequency of the FM signal linear in x.

It suffices to note that a linear FM signal has monotonic increasing phase and that it is possible to assign a phase difference based upon the instantaneous phase of the linear FM signal, and a stationary sinusoidal grating:

$$\phi_d(x) = \phi_l(x) - \phi_r(x) \tag{4.62}$$

General analysis of the above equations is difficult, however, the Fourier transform of equations of the type:

$$f(x) = r(x)\exp[j\beta x^2] \tag{4.63}$$

can be approximated by [81]:

$$F(\omega) = \frac{1}{2}\sqrt{\frac{j\pi}{\beta}}\ \exp[-\frac{j(\omega-\omega_o)^2}{2\beta}]r(\frac{\omega-\omega_o}{2\beta})\ ; \quad \omega > 0 \tag{4.64}$$

The phase can be shown to be parabolic centered at ω_o, providing there is little overlap between $F(\omega_l - \omega_o)$ and $F(-\omega - \omega_o)$. It is therefore clear, that we may assign an appropriate spatial phase difference:

$$\phi_d(x) = \int_{-x_o}^{x} 2\pi u(s) - 2\pi u_o \, ds \tag{4.65}$$

4.9 Methods

To highlight the principles that we have discussed, we will present some experimental results to some simple, and real image data.

4.9.1 Methods of Evaluation

In figures B.2 to B.4 we have produced graphical displays of the grey level pixel values corresponding to a selection of stereo raster lines. These sequences will be used as a control to enable a quantitative evaluation of the methods outlined whenever real image data is used. Accordingly, an edge based interpolation (manual) was employed to provide the idealised disparity response between the pairs of raster lines (fig. B.6). The edge based disparity measurements are shown superimposed on experimental work to provide an indication of the accuracy of results.

With each technique employed, the mean ($\bar{x}$), normalised correlation (r), and standard deviation (σ_n) of disparity error was calculated based upon the measured response and those of the algorithms employed in this thesis. These results are formally presented in tabular form in Appendix A, and also given in the appropriate figure captions.

4.9.2 Disparity measurement

Experiment 1

In the first experiment a uniform shift of 20 pixels was induced into one of a pair of gratings. The estimates of disparity were obtained from a series of filters tuned to different spatial frequencies. Disparity measurements, were based upon the center frequency for each filter.

Experiment 2

We repeat experiment 1, however, we have used *a priori* knowledge of the sine grating as the parameter to interpret disparity from equation (4.27). We also examine the error that is implicitly included within the phase correction factor ζ.

Experiment 3

We examine the error in the interpretation of disparity by applying sinusoidal gratings of different spatial frequencies to both images with constant (pixel) shift, and apply the same Gabor function in both images to estimate disparity. Disparity is interpreted by the center frequency of the filter.

Experiment 4

We apply equation (4.35), towards the estimation of instantaneous frequency for a static sinusoidal grating. We also apply the local estimate of instantaneous frequency to interpret disparity to several images, each comprising of a stationary pair of sinusoidal gratings with a constant pixel shift.

Experiment 5

Here we consider the estimation of disparity applied to raster lines taken from real stereo data. We will only consider the direct interpretation of disparity from a selected number of scales to indicate the error and difficulties that are associated within the technique. We will also examine qualitatively the estimation of instantaneous frequency by both the spatial derivative of phase, and the application of quadrature filters and their derivatives. We also consider thresholding the local phase gradient defined by equation (4.49) to restrict the interpretation of disparity from image regions that are not analytic.

Experiment 6

We examine the pre-envelope defined by equation (4.51) with the purpose of examining the response of a single filter to different spatial frequency stimuli and also the response of different filters through spatial scales that operate on sine gratings tuned to the center frequency of the filter. We will also examine the spatial variation of the pre-envelope to a sinusoidal function with a mean intensity level (d.c component) and also the spatial behaviour of the pre-envelope that is not tuned to the frequency present in the image.

Experiment 7

This experiment refers to the *diffrequency* hypothesis proposed by Blakemore[7]. We consider the phase differences taken from a simple sine grating stimulus with respect to a second sinusoidal signal that is frequency modulated. Here, frequency modulation can be interpreted to infer a smooth change in the local disparity (i.e depth modulation) which can still be interpreted by phase differences.

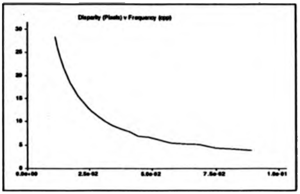


Figure 4.3: *Disparity estimates for different filters applied at the same location to a sinusoidal grating of* $1.5625\text{x}10^{-2}$ *cpp (cycles per pixel). Preferred filter center frequencies varied from* $7.8\text{x}10^{-3}$ *cpp to 0.125 cpp. A Disparity of 20 pixels was introduced between each grating. Graph shows disparity (y-axis) v filter center frequency (x-axis).*

4.10 Results

4.10.1 Disparity Measurements

Experiment 1

By maintaining the image grating constant and varying the center frequency of the filter, the disparity measurements were overestimated for the case of filters tuned to lower spatial frequencies than the grating and underestimated for filters tuned to higher spatial frequencies. The function of the variation of disparity is a hyperbolic curve which follows from equation (4.27) since $D_{est}(\omega_g) = \phi/\omega_g$, where ω_g is the center angular frequency of the filter and the phase difference ϕ is a constant. Notice that the correct disparity of 20 pixels is only available in this case when the center frequency of the filter, and image frequency coincide (fig.4.3).

Experiment 2

We have repeated experiment 1, however we have used the center frequency of the grating to interpret disparity. The actual shift of 20 pixels is now apparent. Disparity measurements taken from equation (4.27) are considerably improved incomparison to experiment 1, with a disparity error of ± 0.5 pixels for filters within 2.5 octaves of the fundamental frequency of the signal. Operating under the non-linear phase region of the filter can be seen comparing figures 4.4a and b. In figure 4.4b the ζ term was also introduced into the phase estimates for

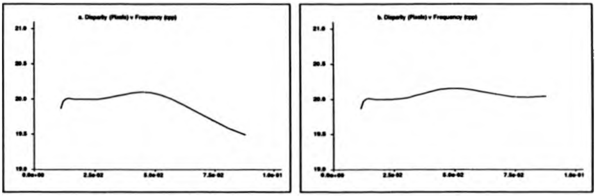


Figure 4.4: *(a)Disparity estimates (y-axis) to the same stimuli pair as fig. 4.3 for different sinusoidal image functions (x-axis). Disparity interpreted from the fundamental form of the signal (b) Improved disparity estimates obtained for the same stimuli as (a) but also estimating the shear in rotation from the correction term* ζ.

measuring disparity. Notice that improvements owing to this term only arise for filters with higher spatial frequency preferences than the image signal. This is because the correction factor is influences by σ_x^2 and u_x for a fixed grating. Since there exists an inverse relationship between the filter's standard deviation and fundamental frequency to the first order, then σ_x^2 forces the filter to respond linearly. This does not apply when the filter is tuned to a higher spatial frequency than it receives. Under these conditions, there will be a rotational correction that is required. This rotation correction can in general, only be found by local instantaneous frequency examination. We note that it is in principle possible to gain considerable accuracy from pure signals and any filter we choose. Real data does not permit such luxuries. In particular, band-pass filters have a weighted preference for their own fundamental frequency. Thus it would be unreasonable to expect a single resolution of filtering to unambiguously measure disparities in a complex image signal.

Experiment 3

In figure 4.5, we have used a single Gabor function to interpret disparity while altering the spatial frequency of the grating stimulus for a constant shift. In this case, the error in disparity measurement is linear which follows from equation (4.27) since $D_{est} = \omega_a k$ where ω_a is the frequency of the grating, and the constant $k = d/\omega_g$ which is the actual disparity

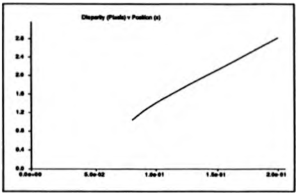


Figure 4.5: *Variation in disparity measurement (y-axis) for gratings of different spatial frequency (x-axis). A constant 2 pixel shift was induced into the stimulus. Center frequency of Gabor at 0.125 cpp.*

divided by the center angular frequency of the Gabor filter.

Experiment 4

Figure 4.6 shows the instantaneous frequency calculated from the phase gradient. For a sinusoidal image tuned to the center frequency of the Gabor filter there was a numerical error of ± 0.2 cpp. For a signal differing by 25% in spatial frequency, the error in instantaneous frequency was ± 0.892 cpp. We can substitute the instantaneous frequency into the disparity estimate, to improve our accuracy in disparity measurement. The potential improvement in disparity estimates by the local instantaneous frequency is presented in figures 4.7a and 4.7b. Unfortunately, should the phase from Gabor filters operate under the non-linear region, we also introduce oscillatory behaviour into the disparity estimate (fig. 4.7c).

We note that in this case, the separation between image and filter signals is greater than we would normally permit. An interesting point to observe in figure 4.7c, is with the center frequency of the filter at $6.25\text{x}10^{-2}$ cpp with the image grating at $1.5\text{x}10^{-2}$ cpp. The disparity was 20 pixels. Although the associated error with this filter was ± 4.0 pixels, the filter measured a disparity difference in excess of its own wavelength. We should also observe that the oscillatory nature of the disparity estimate can be entirely predicted by:

$$\frac{\sin^2[2\pi u_o x]}{\zeta} + \zeta \cos^2[2\pi u_o x]$$

Which is recognised as a cyclic function. It is not possible to correct for phase non-linearity

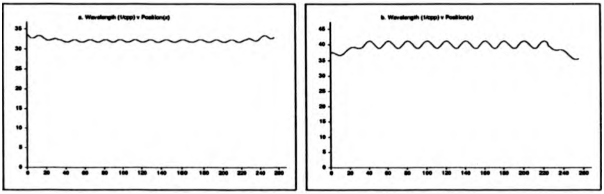


Figure 4.6: *(a) Instantaneous frequency (y-axis) estimates against position (x-axis) for image grating of* $3.125 \text{x} 10^{-2}$ *cpp. Filter tuned to grating. (b) Instantaneous frequency estimates from the same filter pairs in (a) but image signal now 0.025 cpp.*

by this method since it has in itself implied that the correction term was already close to unity.

Experiment 5

We will now consider the application of the theory that we have developed to real image data. Figure 4.8 shows the instantaneous frequency estimates for a real image scan line based upon equations (4.35) and (3.17). Filters were tuned to $1.5 \text{x} 10^{-2}$ and 0.0625 cpp. The instantaneous frequency response shows the interesting property of negative frequency response and stationary phase (stationary phase refers to a zero phase gradient). Since the oscillatory term owing to filter non-linearity is always positive, we can only assume that the change in gradient is caused by negative frequencies. In the words of Gabor[25]: "negative frequencies are suppressed", but we add that they are not removed. The exact nature of the negative phase gradient can be shown by the examination of the quadrature phase as a function of image position. This we present in figure 4.9 for two stereo raster sequences. The disparity interpretations from these phase functions can be seen in figures 4.10c and 4.11b. Notice that the differences in the image intensity functions from figure B.2 has introduced negative frequency and stationary phase behaviour in only one phase function. Thus the phase differencing technique is not analytic because of this local behaviour. The restriction of stationary phase to a single image sequence is owing to the difference in local instantaneous

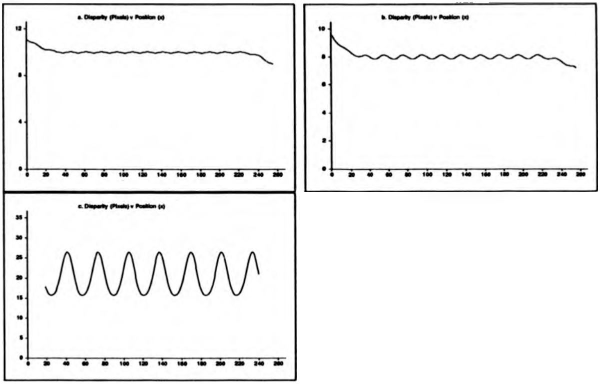


Figure 4.7: *Disparity estimates (y-axis) interpreted from the derivative of phase against position (x-axis). Initial disparity of 10 pixels introduced to a pair of filters centered at* $3.125\text{x}10^{-2}$ *cpp. (a) Grating of* $3.125\text{x}10^{-2}$ *cpp. (b) Grating of 0.025 cpp. (c) Filter tuned to* $6.25\text{x}10^{-2}$ *cpp, grating* $1.5\text{x}10^{-2}$ *cpp and disparity of 20 pixels introduced.*

frequency owing to stereoscopic transformation. We will show in the next chapter that this behaviour is due to interference. The measured disparity functions for these raster pairs is shown in figure B.6.

At the lowest resolution (fig 4.10a), using the center frequency of the filter to interpret disparity has resulted in an under estimation of the disparity. This is because of the error in the interpretation of the local phase gradient which we can see by comparing figures 4.10a and 4.11a with B.6. From figure 4.9 we also observe that two turning points have occurred in the phase advance and that the presence of negative frequency components has been detected. During the transition from positive to negative frequency, the instantaneous frequency must pass through stationary phase i.e d.c. We are therefore trying to apply phase differences to markedly different frequency components of the two images at this point. More seriously, the turning point also yields an ambiguity since there are effectively two points that could

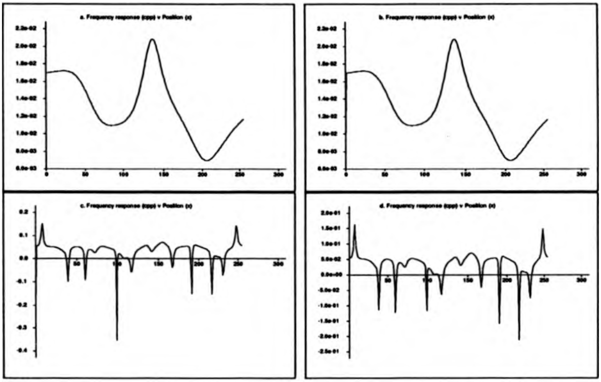


Figure 4.8: *Instantaneous frequency (y-axis) estimates obtained from figure B.2b as a function of position (x-axis). (a) From the spatial derivative of phase (b) From Gabor filters and their first derivatives. Filters were tuned to* $1.5\text{x}10^{-2}$ *cpp. (c) Spatial derivative of phase from Gabor function. (d) Instantaneous frequency obtained from Gabor filters and their derivatives. Center frequency of Gabor function at 0.0625 cpp.*

be potentially matched. This is contrary to the one to one correspondence constraint. Under these conditions, the method of phase differences is not converegent and we expect instability. The phase function obtained by convolution with the Gabor filter will always respond to the local frequency content of the image stimulus. In complex image data, where there exist many local Fourier energies, we would expect the local phase gradient to lie within the upper and lower cut-off bounds of the bandpass filter. Stationary phase and negative frequency is a clear indication that these assumptions are insufficient and arise because of a non-linearity that has been introduced into the system.

To prevent a phase difference to be taken at image regions that exhibit these non-linearities, it is proposed to apply a threshold based upon an upper and lower cut-off of one

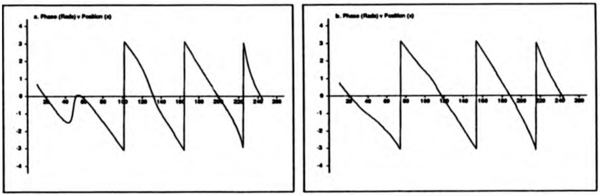


Figure 4.9: *Phase function (y-axis) response against position (x-axis) obtained by convolving with filters tuned to* $1.5\text{x}10^{-2}$ *cpp, from figure B.2. (a) Left image (b) Right image*

octave with respect to the center frequency of the filter. The threshold is based upon the local phase gradient. Some results for disparity interpretation using this technique are shown in figure 4.12 where we have not examined a phase difference within regions of the image which violate this criterion. Thresholds are applied based upon equation (4.49).

Experiment 6

We show that maximal energy from each filter occurs at the spatial frequency of interest (fig. 4.13). Here the image signal remains stationary, and we alter the center frequency of the filter. Notice the rapid exponential decay that is predicted. from equation (4.52). We also notice the skew that is introduced from the energy response. Figure 4.13b shows the weighted preference that the selected Gabor filters hold for d.c signals. Each filter was tuned to the signal frequency of preference. The graph shows the energy response for different gratings with equal contrast. The prediction that the energy response is dependent on the spatial location of the filter is also an important aspect to consider. Because of this, it is difficult to obtain information exclusively from energy derivatives with respect to spatial location (fig. 4.13c). It is important to observe, that the cyclic oscillation of energy response, is twice the fundamental frequency of the input grating and that the magnitude of these oscillations become more pronounced, when the signal and image filter differ. This can be seen in fig. 4.13d where the image signal and filter center frequency differ by an octave. Notice that the oscillations increase relative to the amplitude of the response. In figure 4.13e we

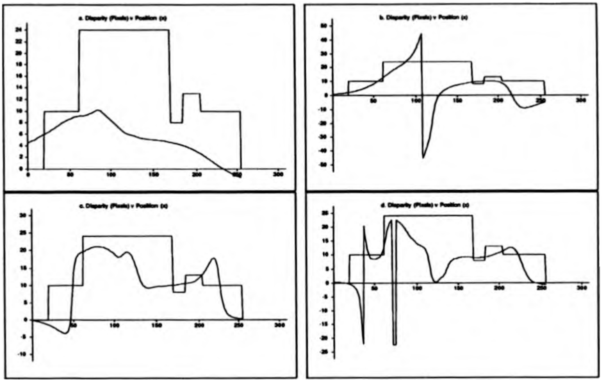


Figure 4.10: *Disparity estimates (y-axis) from phase differences against position (x-axis) for the image raster in B.2. (a) Filters tuned to* $7.8\text{x}10^{-3}$ *cpp. (b)* $1.09\text{x}10^{-2}$ *cpp. (c)* $1.5\text{x}10^{-2}$ *cpp (d)* $2.2\text{x}10^{-2}$ *cpp.*

show the d.c bias of the Gabor filter to a sinusoidal signal with non-zero mean intensity. The energy response now appreciably oscillates at the fundamental frequency of the image signal. These energy oscillations are particularly concerning should we wish to use the pre-envelope as a probabilistic weighting.

Experiment 7

Figure 4.14 shows the phase difference between two sinusoidal signals, one of which is linearly frequency modulated. However, the increase in phase differences has lead to wrap-around. The *diffrequency* paradigm[7] hypothesised that surface slant can be predicted from local differences in horizontal spatial frequency. Blakemore's main driving theory was based upon a frequency modulated grating that was stereoscopically fused relative to a constant grating to provide the perception of surface slant. Because of the frequency modulation, his right

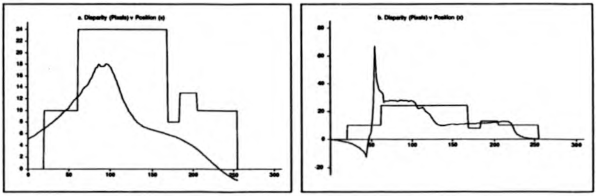


Figure 4.11: *Instantaneous frequency applied to figure B.2 and used to interpret disparity (y-axis) against position (x-axis). Filters tuned to (a)* $7.8\text{x}10^{-3}$ *cpp. (b)* $1.5\text{x}10^{-2}$ *cpp.*

image contained an additional cycle of intensity stimulus. Since stereopsis was possible, and he observed no difference in stimulus owing to the additional cycle, he argued that this mechanism was indeed used to provide cues for surface slant. Our results indicate that the method of phase differences is also consistent with this representation. In figure 4.15 we observe a sinusoidal modulation. Since the derivative of a sinusoidal phase is also an oscillatory function, we know that the first derivative of phase can pass through stationary values. Under these conditions we would expect difficulty in disparity interpretation since the test grating is a static sinusoid. Therefore the local differences in instantaneous frequency must change markedly between the two stimuli. A phase difference between a static (constant phase advance) grating and sinusoidally modulated phase grating would be expected to produce a sinusoidal depth modulation. Results to this stereo pair by applying equation (4.49) indicate the error in disparity interpretation was 40%. However, we have not iterated to obtain better estimates.

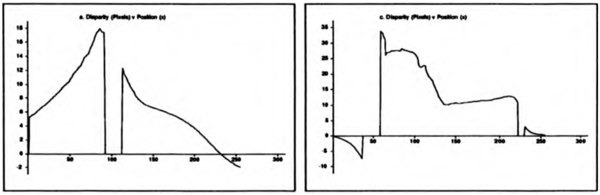


Figure 4.12: *Upper and lower cut-off of 1 octave applied to disparity estimates (y-axis) against position (x-axis) using the derivative of phase to interpret disparity from figure B.2. Filters preferred frequencies were (a)* $7.8\text{x}10^{-3}$ *cpp. (b)* $1.5\text{x}10^{-2}$ *cpp.*

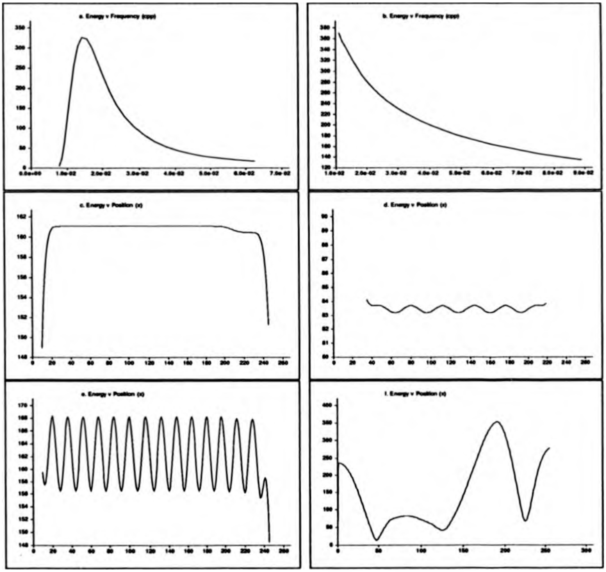


Figure 4.13: *Pre-envelope response (y-axis) against position (x-axis) presented for ideal sinusoidal signals. (a) Energy response for different bandpass filters to a grating of* $1.5\text{x}10^{-2}$ *cpp applied at the same location in the image domain. (b) Change in energy response from quadrature sets of filters applied to gratings with mutual preference. (c) Energy response to a grating of* $6.25\text{x}10^{-2}$ *cpp convolved with a pair of Gabor functions tuned to* $6.25\text{x}10^{-2}$ *cpp. (d) Energy response as (c) but filter preference at* $3.125\text{x}10^{-2}$ *cpp and image grating* $1.5\text{x}10^{-2}$ *cpp. (e) As (c) but the sinusoidal grating has a d.c bias introduced. (f) Energy response from raster line B.2. Filter with center frequency at* $1.5\text{x}10^{-2}$ *cpp.*

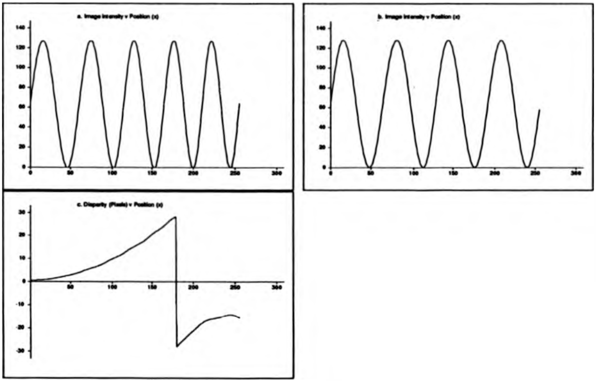


Figure 4.14: *Pair of sinusoidally based stereo rasters. (a) Grating with instantantaneous frequency linear in spatial position ($\beta = 0.001$). (b) Reference grating (c) disparity estimate. Center frequency of filter and reference grating* $1.5\text{x}10^{-2}$ *cpp.*

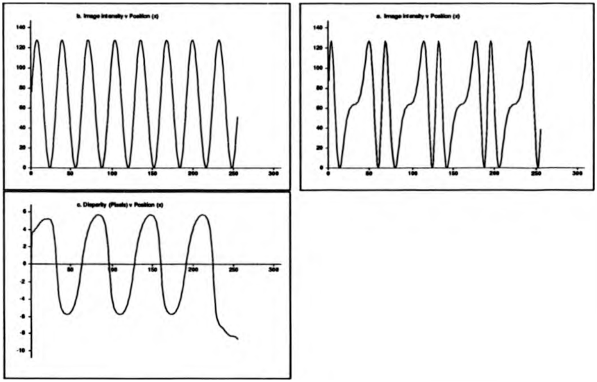


Figure 4.15: *Pair of sinusoidally based stereo rasters with a disparity frequency of 10 pixels peak to peak and* $1.56\text{x}10^{-2}$ *cpp. Carrier frequency was* $3.125\text{x}10^{-2}$ *cpp. Filter center frequency* $3.125\text{x}10^{-2}$ *cpp. (a) Reference (b) Modulated grating. (c) Depth estimate.*

Chapter 5

NOISE AND INTERFERENCE

5.1 Approximations to added Gaussian Noise

The study of noise within phase sensitive systems has previously found motivation with the study of frequency modulated communication theory and Phase-locked loop tracking systems. From this point of view we are fortunate, since many questions and solutions have already been provided. We therefore extract the relevant information, and apply the results to our own needs. We began using a model proposed by Foschini et al[21], which we will modify to our own notation. Here we note that the use of bold type will indicate that the function obtained by $(\Psi(x,\omega) * I(x))$ will be represented as a vector as opposed to the modulus-Argument form. We have:

$$\mathbf{z}_l = \mathbf{R}_\phi \mathbf{z}_r + \mathbf{n} \tag{5.1}$$

where the vectors are given by:

$$\begin{bmatrix} I_{s\phi} \\ I_{c\phi} \end{bmatrix} = \mathbf{z}_l \quad \begin{bmatrix} I_s \\ I_c \end{bmatrix} = \mathbf{z}_r \quad \begin{bmatrix} n_s \\ n_c \end{bmatrix} = \mathbf{n}$$

and $\mathbf{R}_\phi$ is the normal rotation matrix, and $\mathbf{n}$ is a quadrature component of Gaussian noise added to the signal, whose elements n_c and n_s are both independent stationary wideband Gaussian processes with one-sided spectral density (N_o).

5.1.1 Preliminary approximations

We begin by assuming that $n_c(x)$ and $n_s(x)$ can be considered as Hilbert pairs after bandpass filtering and assume that each noise element has similar statistics, with the same power spectrum and autocorrelation function:

$$n(x) = n_c(x)\cos\omega_s x + n_s(x)\sin\omega_s x$$

and

$$E(x) = \sqrt{n_c^2 + n_s^2}\,,\ \Phi(x) = \tan^{-1}[\frac{n_s}{n_c}]$$

which can be re-written as:

$$n(x) = E(x)\cos[\omega_s x + \Phi(x)]$$

which derives the physical interpretation of a slowly varying amplitude, while the phase oscillates with an expected frequency related to the center angular frequency of the bandpass filter (ω_g). As we have already observed with the Gabor function, the spectral components present in the image function are proportional to the change in the envelope through bandpass filtering, which is also the case with a random noise input. It is fairly simple to show that for bandpass noise[10] the probability density for $E(x)$ is described by a Rayleigh distribution:

$$f_E(r) = \frac{r}{\sigma^2}\exp[-\frac{r^2}{2\sigma^2}]\,, \quad r \geq 0$$

and the corresponding phase distribution is uniform:

$$f_\Phi(\phi) = \begin{cases} \frac{1}{2\pi}\,, & \text{if } 0 \leq \phi < 2\pi \\ 0\,, & \text{otherwise} \end{cases}$$

It is important to note, that the processes $E(x)$ and $\Phi(x)$ are not statistically independent from each other, but the statistical samples of phase and envelope at any one instance are indeed independent.

5.1.2 Added noise

For noise that can be considered additive, it is possible to draw straight forward conclusions regarding the behaviour of a phase sensitive process. We have[90]:

$$\Psi(x,\omega,\sigma) * (I_r(x) + N_r(x)) = \mathbf{z}_r(x,\omega,\sigma) + \mathbf{n}_r(x,\omega,\sigma) \tag{5.2}$$

where $N_r(x)$ refers to an added noise term. The affects of additive noise may be described by a circle of uncertainty (fig. 5.1). The largest angular displacement (phase error) owing to the noise term (ϕ_{max}) will clearly be when :

$$\sin^{-1}\frac{|\mathbf{n}_r(x,\omega,\sigma)|}{|\mathbf{z}_r(x,\omega,\sigma)|} = \phi_{max}$$

or more precisely, when the noise and signal are orthogonal in the complex plane. If we assume that the noise can be represented with respect to one image only then if $\mathbf{z}_r(x,\omega,\sigma) >> \mathbf{n}_r(x,\omega,\sigma)$ we have:

$$\phi_{max} \approx \frac{|\mathbf{n}_r(x,\omega,\sigma)|}{|\mathbf{z}_r(x,\omega,\sigma)|} \tag{5.3}$$

which implies that the loss in accuracy with respect to noise increases linearly in the presence of a dominant image signal. Naturally should $|\mathbf{z}_r(x,\omega,\sigma)| < |\mathbf{n}_r(x,\omega,\sigma)|$ then $\phi_{max} \in [-\pi,\pi]$.

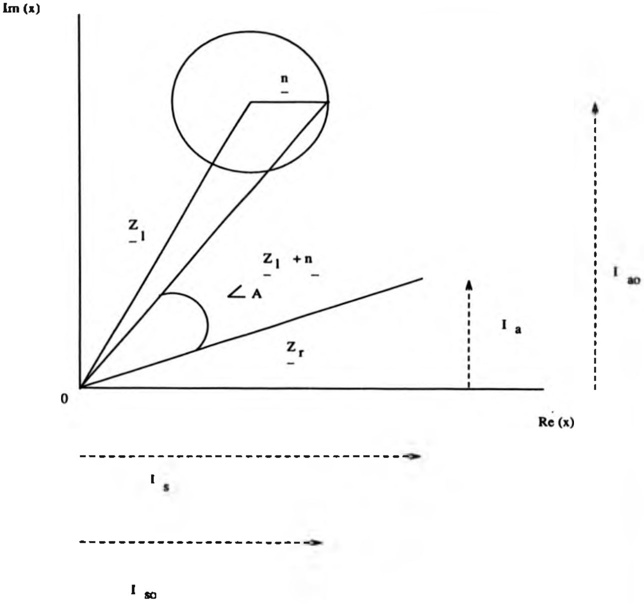


Figure 5.1: *The consequence of added noise into the circle of phase differences showing the uncertainty of disparity interpretation formed from a circle with radius equal to the magnitude of the noise, but with unknown phase.*

5.1.3 Mechanisms for reducing noise sensitivity

A least squares estimate for the phase angle ϕ is not generally possible to obtain from a single measurement. However, for small ϕ using equation (5.8):

$$\mathbf{n}^T\mathbf{n} \approx [(\mathbf{z}_l - \mathbf{z}_r) - \phi\mathbf{z}_{r\perp}]^T[(\mathbf{z}_l - \mathbf{z}_r) - \phi\mathbf{z}_{r\perp}] \tag{5.4}$$

we can now take the derivative of the above equation with respect to ϕ and set to zero:

$$\frac{d\mathbf{n}^T\mathbf{n}}{d\phi} \approx 2\phi\mathbf{z}_{r\perp}^T\mathbf{z}_{r\perp} - (\mathbf{z}_l - \mathbf{z}_r)^T\mathbf{z}_{r\perp} - \mathbf{z}_{r\perp}^T(\mathbf{z}_l - \mathbf{z}_r) \tag{5.5}$$

and obtain:

$$\phi \approx \frac{<\mathbf{z}_{r\perp},\mathbf{z}_l>}{|<\mathbf{z}_{r\perp},\mathbf{z}_{r\perp}>|} \simeq \frac{<\mathbf{z}_{r\perp},\mathbf{z}_l>}{|<\mathbf{z}_{r\perp},\mathbf{z}_{r\perp}>|^{\frac{1}{2}}|<\mathbf{z}_{l\perp},\mathbf{z}_{l\perp}>|^{\frac{1}{2}}} \tag{5.6}$$

This function then represents a least squares measure of phase differences for small ϕ. Within the 1-D framework, the only alternative we can propose to obtain a true least squares measure would be to add filter coefficients that are not in strict quadrature. Thus we might consider adding 4 sets of filters who phase angle of separation is $\frac{\pi}{4}$ radians. This topic has been dealt with by Lange[51] who quotes an improved signal to noise ratio of 17% using this method.

5.1.4 Steerable property of quadrature filters

The steerable properties of 2-D Gaussian derivative operators has been discussed in the orientation domain by Freeman and Adelson[23]. We make the observation that it is also possible to interpolate a quadrature filter with arbitrary phase from any pair of quadrature filters with orthogonal phase differences. This follows from the orthogonality property of quadrature phase and the linearity of convolution[10]. Consider the complex modulated quadrature Gabor filter $\Psi(x;\omega,\sigma)$ convolved with an image function $(I(x))$:

$$z(x;\omega,\sigma) = \Psi(x;\omega + \phi,\sigma) * I(x) \tag{5.7}$$

where ϕ refers to the phase of the modulating sin and cosine part relative to the envelope of the filter. Referring to the Gaussian envelope of the Gabor function by $G_{ab}(x;\sigma)$ we have:

$$\Psi(x;\omega + \phi,\sigma) = G_{ab}(x;\sigma)\exp[\jmath\omega x + \phi]$$

expanding equation (5.7) we have for the real part:

$$\mathcal{R}e\Psi(x;\omega + \phi,\sigma) * I(x) = \cos\phi\ G_{ab}(x;\sigma)\cos[\omega x] * I(x) - \sin\phi\ G_{ab}(x;\sigma)\sin[\omega x] * I(x)$$

and the complex part:

$$\mathcal{I}m\Psi(x;\omega+\phi,\sigma) * I(x) = \sin\phi\ G_{ab}(x;\sigma)\cos[\omega x] * I(x) + \cos\phi\ G_{ab}(x;\sigma)\sin[\omega x] * I(x)$$

thus we can interpolate the response from a quadrature pair of Gabor filters to another Gabor filter with similar envelope characteristics but arbitrary phase.

Therefore, it is sufficient to sample with a pair of quadrature filters with phase modulation of 0 and $\pi/2$. It would, however, be possible to apply several quadrature filter pairs with the same center frequency but different spatial envelopes to obtain a least squares estimation of phase differences.

5.2 The presence of noise in correlated signals

Within the previous section, we have assumed that the statistical parameters that describe the quadrature components of band-pass noise are independent. These parameters can therefore be described by a bivariate Gaussian density function with zero correlation coefficient. However, consider two processes such as our stereoscopic signal pairs. In this case, the joint probability distribution functions are not necessarily independent i.e:

$$P(\mathbf{x}_l|\mathbf{x}_r) = \frac{P(\mathbf{x}_l,\mathbf{x}_r)}{P(\mathbf{x}_l)}$$

such that:

$$P(\mathbf{x}_l,\mathbf{x}_r) \neq P(\mathbf{x}_l)P(\mathbf{x}_r)$$

Under these conditions, the describing joint probability density (JPD) for these two functions ($[x_l, y_l]^T$ and $[x_r, y_r]^T$) becomes[36]:

$$P(I_s, I_a, I_{s\phi}, I_{a\phi}) = (2\pi\sigma^2)^{-2}(1-r^2)^{-1}\exp[-\frac{I_s^2 - 2rI_sI_{s\phi} + I_{s\phi}^2 + I_a^2 - 2rI_aI_{a\phi} + I_{a\phi}^2}{2\sigma^2(1-r^2)}]$$

where r is the correlation coefficient between the two pairs of variables that is related to the bandwidth of the bandpass filter such that for small bandwidth, $r \rightarrow 1$. Converting the above JPD into polar coordinates, which to our usual notation, requires the Jacobian matrix:

$$\frac{\partial(I_s, I_a, I_{s\phi}, I_{a\phi})}{\partial(E_l, \Phi_l, E_r, \Phi_r)} = E_r E_l$$

which may used to obtain the probability density function for the phase differences between the two dependent processes ($\phi = \phi_l - \phi_r$):

$$f_\phi(\phi) = \frac{1-r^2}{2\pi}(1-r^2\cos^2\phi)^{-\frac{3}{2}}[(1-r^2\cos^2\phi)^{\frac{1}{2}} + r\cos\phi\ (\frac{\pi}{2} + \sin^{-1}(r\cos\phi))]$$

Interestingly, with this density function the mode refers to the correlation coefficient r. It is easily verified that when $r = 0$ the probability density function (PDF) reduces to that of a uniform PDF (all phase difference angles are equally likely). The density function is also real and even ($f_\phi(\phi) = f_\phi(-\phi)$). This function is indeed similar to the Tichonov PDF, which we will also review.

5.2.1 Phase-locked loop with added noise

Foschini et al[21] were primarily interested in high speed communications efficiency encoded by pairs of quadrature components. Of interest to us, was their application of the Tichonov/Viterbi probability density to combined phase jitter and added Gaussian noise. The consequence of phase jitter rotates the vector $\mathbf{z}$ by an angle ϕ. The effects of phase jitter for our purposes may complement small perturbations in the camera system from robotic considerations of camera hardware, while Gaussian noise can be used to model the sensing noise from a physical device. Discontinuity in the disparity field appears at present to be a particularly difficult parameter to consider, and will therefore not be considered in this analysis. For noise that adds small increments of phase we have:

$$\mathbf{z}_l - \mathbf{z}_r = \mathbf{R}_\phi \mathbf{z}_r - I\mathbf{z}_r + \mathbf{n} \approx \phi \begin{bmatrix} -I_s \\ I_s \end{bmatrix} + \mathbf{n} \tag{5.8}$$

since $\cos\phi \approx 1$ and $\sin\phi \approx \phi$. Notice the last term on the right represents the normal vector ($\mathbf{z}_{r\perp}$) to $\mathbf{z}_r$ Averaging the norm square we have:

$$\mathcal{E}||\mathbf{z}_l - \mathbf{z}_r||^2 = \mathrm{N}_o + \sigma_\phi^2||\mathbf{z}_r||^2 \tag{5.9}$$

where σ_ϕ^2 is the variance of ϕ and $\mathcal{E}$ denotes the statistical average.

5.2.2 Tichonov Probability Density Function

We also require the probability density function of the phase error. A general closed form solution for such a PDF has yet to be found. Approximations based on the input to a first-order Phase-locked loop whose input is white Gaussian noise and a sinusoid was found by Viterbi[95]. He proposed that the phase advance could be described by a first order Markov process, whose instantaneous probability density $p(\phi, t)$ must satisfy the Fokker-Plank diffusion equation:

$$\frac{d(p(\phi,t))}{dt} = -\frac{d}{d\phi}[A(\phi)p(\phi,t)] + \frac{1}{2}\frac{d^2}{d\phi^2}[B(\phi)p(\phi,t)]$$

where $A(\phi)$ and $B(\phi)$ represent the change in phase ($\Delta\phi_{\Delta t \to 0}$) and variance of phase ($\Delta\phi^2$) as initial conditions. The solutions to this equation could only be found for the steady state distribution which was shown to be:

$$P(\phi) \approx \frac{1}{2\pi}\frac{\exp[\alpha\cos\phi]}{I_o(\alpha)}, \quad |\phi| <= \pi$$

Where I_o is the modified Bessel of the first kind and α is a positive number representing the ratios of signal to noise power. Viterbi, also noted the useful relationship that for $\alpha > 100$, $\sigma_\phi^2 = \frac{1}{\alpha}$. The reader, is referred to the cited literature for more information. The proofs are long, but the work of Foschini et al[21] holds important implications for this thesis. They derived two important equations. The first, described the PDF for signal transmission:

$$P(\mathbf{z}_l) = \frac{1}{2\pi \mathrm{N}_o}\exp[-\frac{1}{2\mathrm{N}_o}d^2(\mathbf{z}_l,\mathbf{z}_r)]$$

where $d(\mathbf{z}_l,\mathbf{z})$ referred to the distance between $\mathbf{z}_l$ and $\mathbf{z}_r$. The second equation is an interesting expression for this 'distance' as:

$$d^2 \approx ||\mathbf{z}_l - \mathbf{z}_r||^2 + 2 < \mathbf{z}_l,\mathbf{z}_r > +2\alpha\mathrm{N}_o - 2|\sqrt{||\mathbf{z}_l||^2||\mathbf{z}_r||^2 + 2\alpha\mathrm{N}_o < \mathbf{z}_l,\mathbf{z}_r > +(\alpha\mathrm{N}_o)^2}|$$

We are interested in a 'distance' obtained from the unit circle, which is interpreted from the phase angle difference. Their measure incorporates the actual 'distance' from the transmitted and received quadrature signals. Naturally, if we could adequately represent noise, we could also use this expression as a formulation for the probability of correct correspondence. Foschini et al, also point out that for closely packed signal constellations, equation (5.9) shows that signal points located further from the origin are subjected to a larger mean-square error. A similar error occurs with our method of disparity interpretation since if we consider a phase error owing to noise (ϵ). Then our interpretive disparity error is:

$$d_e = \frac{\epsilon}{\omega_i}$$

Clearly, the lower the spatial frequency, the larger the disparity error for a given phase error attributable to noise alone. In this way, the lowest spatial frequencies, which are capable of estimating the largest disparity differences without aliasing, also hold the disadvantage of providing the greatest potential error in disparity measurement. Unfortunately, it would also be difficult to extract the noise term in practice. Even the variation of energy associated with bandpass noise would be difficult to isolate, since we would have to separate the potential variation of energy with spatial position, which can occur with both frequency gratings and impulse image functions that are spatially separated owing to disparities.

5.2.3 Experimental results from added Gaussian Noise

To study the behaviour of noise, we have chosen to consider the simple situation of a static unit amplitude frequency grating with added zero-mean Gaussian noise using a standard noise generating package. In figure 5.2a, we present a polar plot of energy verses phase angle from a single sinusoid without added noise. Some specular noise exists from the image boundaries as we might expect. From this figure, we observe a constant energy response and equal probability of phase as we would expect.

Figure 5.2b shows the same plot but with zero-mean Gaussian noise added to the signal ($N_o = 3$). In this case, we observe markedly different results, which are indeed similar to the *drunkards walk*, which is predicted from the Rayleigh distribution. This is indeed confirmed in figure 5.3, where we show the *PDF* for the energy response to a static sinusoid (peak to peak amplitude equal to 3 image intensity units) with varying amounts of added Gaussian noise. Here, the consequence of added noise increases the variance of the energy response.

The distribution of phase error for zero-mean added noise is presented in figure 5.4. An examination of the distributions revealed that the best model to describe the phase error was the Tichonov probability density function. The closest fitting function was superimposed (crosses) onto the experimental data. The density function obtained from the Bivariate Gaussian distribution was found to decay too rapidly near to the expected phase (mean), and too little close to the aliasing limit. The difference is indicated in figure 5.4g where the Tichonov density is presented as crosses. We therefore conclude, that phase locked-loop dynamics can describe the phase error in the presence of Gaussian noise. In addition, the more noise, the greater the phase error which is hardly a surprising result. From figure 5.5b, we observe that the phase response still oscillates between $\pm \pi$ with a uniform probability distribution.

In figure 5.6 we show the local instantaneous frequency which as we have already discussed, bears a relationship to the magnitude of energy oscillations. We might expect this response profile to also represent a Rayleigh distribution. However, because of stationary phase and the presence of negative frequency elements within the image domain this is not the case.

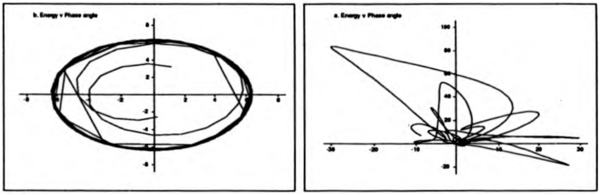


Figure 5.2: *Polar plot of Magnitude and phase angle for 512 points; (a) grating and filter tuned to 1/16 cpp. (b) Zero-mean Gaussian noise added with* $N_o = 3.0$.

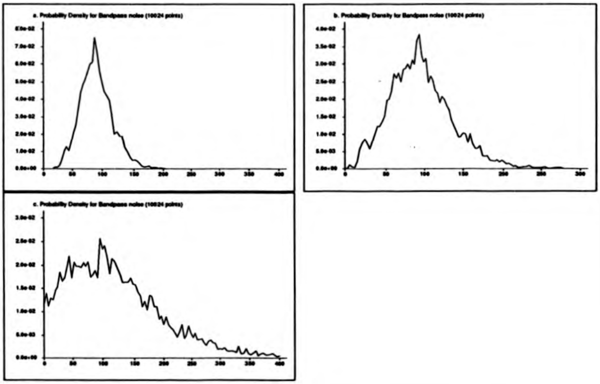


Figure 5.3: *Probability density function obtained from the energy response to zero-mean Gaussian noise added to a stationary grating (1/16 cpp). Filter preferred center frequency also 1/16 cpp. (a)* $N_o = 2.0$ *(b)* $N_o = 3.0$ *(c)* $N_o = 6.0$.

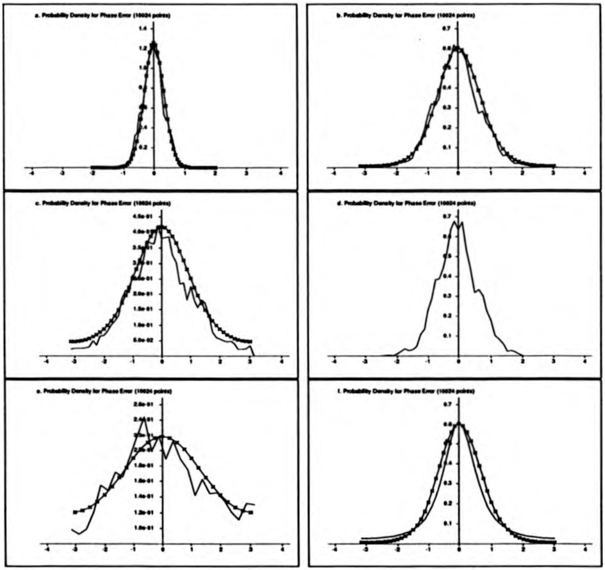


Figure 5.4: *Experimentally determined PDF for phase with zero-mean added Gaussian noise (filter and grating centered at 1/16 cpp). The best fitting Tichonov PDF shown superimposed (crosses). (a)* $N_o = 1.0$ $\alpha = 10$ *(b)* $N_o = 2.0$ $\alpha = 2.3$ *(c)* $N_o = 3.0 \alpha = 1.1$ *(d) As (b) smoothed by the truncated parabola (e)* $N_o = 6$ $\alpha = .3$ *(f) The Bivariate Gaussian and Tichonov PDF for* $\alpha = 2.3$ and $r = 0.77$.

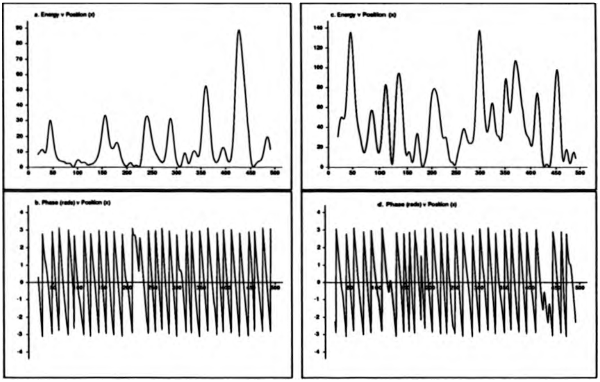


Figure 5.5: *Energy and phase plots as a function of position for the stationary grating used in figure 5.2 with zero-mean added Gaussian noise and (a) & (b) N_o = 3.0 (c) & (d) N_o = 6.0.*

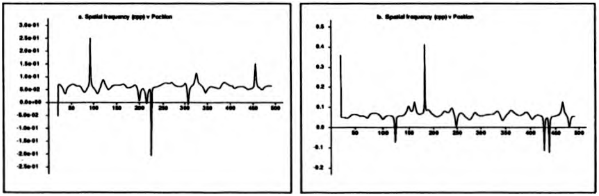


Figure 5.6: *Instantaneous frequency versus position from the data presented in figure 5.5. (a) N_o = 3.0 (b) N_o = 6.0.*

5.3 Interference

We will now discuss stationary phase and negative frequency within the context of bandpass filtering.

5.3.1 Stationary Phase and Negative Frequency

Stationary phase and negative frequency is a property not considered in the method of phase differences. It is therefore essential to understand the nature of this response. This is particularly true, since the Hilbert transform is by definition *"one-sided"* in its Fourier spectrum. The physical occurrence of stationary phase and negative frequency holds a direct analogy to the ***staircase stimulus*** that is well recognised with the Marr Laplacian[86]. Indeed, the Gabor function also exhibits these unwanted properties within these image regions. Under this type of image stimulus, the Gabor function is no longer in strict quadrature. Let us reconsider our phase non-linearity $\zeta(x)$ which is now a function of image position and represents the deviation from strict quadrature. With the Gabor function, this could easily occur under the convolution with a frequency modulated signal, but in practice, may be due to rounding errors, digitization or other artifacts including the complexity of natural image data. If we assume that $\zeta(x)$ only affects the imaginary part of the Hilbert transform as in equation (4.25), we have:

$$f_i = \frac{1}{2\pi}\frac{d}{dx}\tan^{-1}\left[\frac{\hat{h}(x)\zeta(x)}{h(x)}\right] \tag{5.10}$$

which is an equivalent form to equation(3.17). Expressing this equation in terms of the previous definition for instantaneous frequency we have a new measured value (f_{im}):

$$f_{im} \approx \frac{1}{2\pi}\left[\zeta(x)f_i + \frac{\dot{\zeta}(x)h(x)\hat{h}(x)}{w(x)^2}\right] \tag{5.11}$$

In this case, $w^2(x) = h(x)^2 + \hat{h}^2(x)$ also includes the non-linearity. In the case where $\zeta(x)$ is a constant, there will be a constant error in the measurement of instantaneous frequency. However, where the first and higher derivatives of $\zeta(x)$ exist, we have an additional term including the product of the magnitudes of the Hilbert pair. Since $\zeta(x)$ has not been specified, there is no guarantee that f_{im} will now always be positive. Interestingly, even this form of equation justifies the edge based approach, since at both the locations of real and imaginary edges the term on the right of the above equation is zero.

Now consider when $h(x) + j\hat{h}(x)$ is formed from the linear sum of several sinusoidal signals (which is quite justified because of the linearity of the Fourier expansion) formed by bandpass filtering a complex signal i.e:

$$h(x) \; + \; j\hat{h}(x) \; = \; [h_1(x) + h_2(x) + ...h_n(x)] \; + \; j[\hat{h}_1(x) + \hat{h}_2(x) + ... \; \hat{h}_n(x)] \tag{5.12}$$

we now calculate the Wronskian for the above equation and obtain:

$$W(x) \; = \; F(x) + M(x) \tag{5.13}$$

where:

$$F(x) \; = \; [h_1(x)\dot{\hat{h}}_1(x) - \hat{h}_1(x)\dot{h}_1(x)] + [h_2(x)\dot{\hat{h}}_2(x) - \hat{h}_2(x)\dot{h}_2(x)] ... + [h_n(x)\dot{\hat{h}}_n(x) - \hat{h}_n(x)\dot{h}_n(x)]$$

It is easily verified by comparison with equation (3.17), that the above terms enclosed in squared brackets are weighted instantaneous frequency terms. This is a method of describing the spatial frequency selectivity of a bandpass filter as an expansion of frequencies that the filter is sensitive to. Each term in $F(x)$ will be positive in sign, which we know from the action of the Hilbert transform. However, in this particular case, we have also included an additional term ($M(x)$):

$$M(x) \; = \; [h_i(x)\dot{\hat{h}}_j(x) - \hat{h}_i(x)\dot{h}_j(x)] \; + \; [h_j(x)\dot{\hat{h}}_i(x) - \hat{h}_j(x)\dot{h}_i(x)] \; + \; ..... \quad \text{for } i,j \; = \; 0...n \; , \; i \neq j.$$

which contains permutations mixed frequencies between $h_1(x)$ ... $h_n(x)$. It follows from equations (4.18) and (4.17), that we can write:

$$h_i(x) + jh_i(x) \; = \; E_i(\cos \omega_i x + j \sin \omega_i x)$$

If we consider the first pair of terms from our expansion of $M(x)$ then we can re-write this function as:

$$[h_i(x)\dot{\hat{h}}_j(x) - \hat{h}_i(x)\dot{h}_j(x)] + [h_j(x)\dot{\hat{h}}_i(x) - \hat{h}_j(x)\dot{h}_i(x)] \; = \; E_i E_j(\omega_i + \omega_j) \cos[(\omega_i - \omega_j)x + \phi_i - \phi_j] \tag{5.14}$$

Thus we observe that the Hilbert transform in this case is also sensitive to the differences of fundamental frequency elements. Consider the action of filtering a signal, which is complex but with fundamental frequency components placed to both the upper and lower cut-off regions of a given bandpass filter. We would expect a lower energy response in comparison to a signal with the preferred frequency of the filter. However, from the above equation

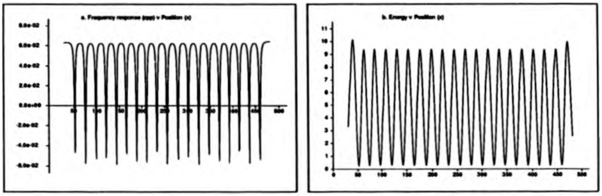


Figure 5.7: *Beats observed from the instantaneous frequency plot of a single filter (0.0625 cpp) and an image signal consisting of two spatial frequencies 0.0441 cpp and 0.0884 cpp. (a) Instantaneous frequency (y-axis) (b) Energy response (y-axis) against position (x-axis).*

we can observe, that it is possible in this case for the difference of individual frequencies to still lie close to the center frequency of the filter. The filter, which would have a weighted preference for such a signal would also be responding to the mixed frequencies and the mean fundamental components of the image signal. From equation (5.14), we observe that this type of interference would be a function of image position and the phase of the signals. The energy of the inferference frequency is also based upon the product of the energies from the individual components, which could be a greater weighting than many of the individual signal elements. This term may also assume both positive and negative values. To indicate this result, we consider the instantaneous frequency from a stimulus comprising of the addition of two sine gratings separated by one octave from which the instantaneous frequency is calculated using equation (3.17). The center frequency of the filter was chosen at the mid-point (in octaves) between the two gratings. The result is presented in figure 5.7. Here we should notice that the periodicity of the beat frequency is approximately 1/22 cpp, which we would predict from the differences of the frequency elements. The spatial frequency selectivity at the maxima of figure 5.7 is also 1/15.98 cpp which is the center frequency of the applied bandpass filter as we would expect from a mean weighting of these two particular spatial frequencies. We also notice that the beat frequencies have also tended toward -1/16 cpp.

5.4 Pre-envelope modulation

Equation (3.11) indicates that for Hilbert transform pairs to remain in quadrature, the pre-envelope should be approximately constant across spatial convolution. Figure 5.7b clearly shows that this is not a rigorous assumption. Here the pre-envelope is itself modulated. A similar result is well known in Telecommunication theory[13]. If we consider a simple model of an AM (amplitude modulated) signal where ω_c and ω_m refer to the carrier and modulation signals with amplitudes A_c and A_m:

$$I(x) \;=\; A_c[1 + A_m \cos\omega_m x]\cos\omega_c x \tag{5.15}$$

expanding gives:

$$I(x) = A_c \cos\omega_c x + \frac{1}{2}A_c A_m \cos[\omega_c + \omega_m]x + \frac{1}{2}A_c A_m \cos[\omega_c - \omega_m]x \tag{5.16}$$

which may be interpreted in the frequency domain as a carrier signal centered at ω_c with a pair of sidebands centered at $\omega_c + \omega_m$ and $\omega_c - \omega_m$. Difficulties arise when $A_c < A_m$ since the physical interpretation of such a signal includes a "*negative envelope*"[13]. This introduces considerable phase and amplitude distortion into the signal an example of which is observed in figure 5.7. Clearly if $\omega_m << \omega_c$ then the sideband frequencies are distant (in the Fourier domain) from the passband of interest. They may therefore be neglected.

Within Telecommunication theory it is usual to overcome this problem by applying a narrow band filter to extract the carrier signal and to amplify the carrier signal to an extent that when recombined with the sidebands: $A_c > A_m$ which relieves the problem. One way to achieve this task would be to subtract the Laplacian[89] which is often applied to enhance image data. Such an operator is both real and bandpass, and would serve to enhance the frequency content in the passband of interest.

Another method that can be used to reduce the occurrence of stationary phase behaviour involves weighting both frequency and phase difference measurements through 3 separate scales of bandpass filtering. Since frequency and phase differences do not retain the implict discontinuity in phase (between $[\pm\pi]$) they are useful parameter spaces to apply weighting techniques. In this latter case, we would be working on the premise that interference occurs because of the presence of two dominant sidebands in the frequency domain with respect to the center frequency of interest. This corresponds to the simultaneous detection of two or more features at different frequencies in Scale-Space by a filter that bisects the scales

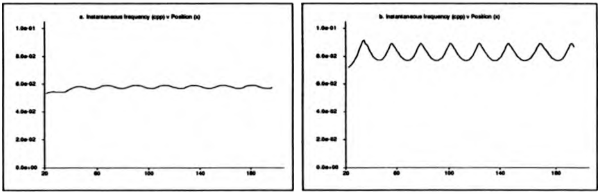


Figure 5.8: *Weighted instantaneous frequency (y-axis) against position. Image comprising of the summation of two sine gratings (a) Center frequency of filters at 0.0441 cpp, 0.0625 cpp and 0.0884 cpp. (b) Center frequency of filters at 0.125 cpp, 0.0884 cpp and 0.0625 cpp.*

of the features[33]. Such a weighting might be described by:

$$f_{av}SS(x) = \frac{\sum_{i=j-1}^{j+1} E_i(x,\omega_i,\sigma_i) f_i(x)}{\sum_{i=j-1}^{j+1} E_i(x,\omega_i,\sigma_i)} \tag{5.17}$$

where $f_{av}SS(x)$ refers to the weighted average of instantanous frequencies ($f_i(x)$) taken at the jth scale and the adjacent sampled scales ($j+1, j-1$) in the frequency domain. A similar expression can be formed for the phase differences between a stereo pair through Scale-Space. Such a scheme would then work on the premise, that a pre-envelope with a low magnitude in comparison to its frequency domain neighbours provides an indication that a given scale of filtering should not be considered. We have applied this technique to the two sine gratings considered previously in figure 5.7. In figure 5.8a we have taken the weighting from the instantaneous frequency of three filters with center frequencies 0.0441 cpp, 0.0625 cpp and 0.0884 cpp. In figure 5.8b we have center frequencies of 0.125 cpp, 0.0884 cpp and 0.0625 cpp. The smoothing of negative frequencies can be seen in comparison to figure 5.7.

Chapter 6

1-D ALGORITHM

6.1 Introduction

Having detailed the method for disparity interpretation, and examined the behaviour of the technique in the presence of added noise, we will now implement the method and explore some of the possible algorithms that we have applied using the method of phase differences.

It should be highlighted that in this section, we restrict our examination to 1-D filters. We will first detail the principles that we will apply in the implementation of each algorithm, and then indicate the exact methods and results as appropriate. We will begin by considering a mechanism for eye or camera vergence.

6.2 A mechanism for eye vergence

From our discussions so far, we are in a position to propose a simple method for eye vergence. The role of eye vergence towards the perception of depth is currently an area of debate. However, for our purposes, eye vergence holds some interesting properties. By verging the eyes, we bring features of small spatial extent into closer correspondence, so that they may then be detected by filters of finer resolutions. The benefits from bringing features closer into correspondence lies with the reduced possibility of aliasing, and also minimising the disparity error from the interpretation of frequency. This is because the error in disparity measurement grows proportionally owing to the difference between the form of the image signal and the frequency used to interpet disparity. The algorithm proceeds from coarse to the finest resolution filter applied to the image, where at each resolution we:

- Obtain the weighted mean disparity estimate at each resolution of filtering by convolving an image function with quadrature Gabor filters and weighting the disparity estimate at each resolution by the pre-envelope product from both image pairs.
- From the disparity estimate at each resolution, induce a vergence mechanism by shifting the images based upon the current mean disparity measurement.

At the finest resolution we would anticipate that vergence has indeed been completed.

Our measured mean disparity is therefore taken from:

$$D_{mean}(\sigma) = \frac{\sum_{x=-N}^{x=N} E(\sigma, x_o - x) d(\sigma, x_o - x)}{\sum_{x=-N}^{x=N} E(\sigma, x_o - x)} \tag{6.1}$$

Where $E(\sigma,x) = \sqrt{E_l(\sigma,x)E_r(\sigma,x)}$ represents the product of energy from the left and right image image pairs respectively at the scale indicated by σ, and $d(\sigma,x)$ refers to the disparity estimate at each pixel position. At each resolution we have obtained a weighted mean disparity estimate at each pixel position by applying equation (4.27) centered at the optical center (x_o) and extending -N to N pixels at a given resolution from the optical center. By progressively reducing N from coarse to fine resolutions of filtering, we therefore facilitate image features with small spatial extent towards correspondence.

A useful indication for the choice of N is the spatial standard deviation of the applied filter ($N = 2\sigma$) taken from the central point of the image. From our stereo examples this algorithm, gave final vergence estimates of 24, 9 and 20 pixels to the nearest integer. The measured disparity values where 25, 9, and 22, pixels for figures B.2, B.3 and B.4 respectively.

Equation (6.1) requires some modification should the local instantaneous frequency be used to measure a mean disparity for vergence. This is because of the presence of stationary phase. In this case we would have to modify equation (6.1) to the pre-envelope weighted form:

$$\hat{D}_{mean}(\sigma) = \frac{\sum_{x=-N}^{N}[E_l(\sigma,x_o-x)+E_r(\sigma,x_o-x)]\;\phi_d(\sigma,x_o-x)}{\sum_{x=-N}^{N}[E_l(\sigma,x_o-x)\;f_l(\sigma,x_o-x)\;+E_r(\sigma,x_o-x)\;f_r(\sigma,x_o-x)]} \qquad (6.2)$$

This filter scheme may well be the process that underlies the mechanism for convergence in binocularly driven mammals. Convergence is therefore achieved by reducing the disparity at some central position in the two images to zero. Providing that stereoscopic signals containing similar frequencies and contrast are present in both images, and that the disparity is less than half the wavelength of the lowest frequency component present, the disparity can be evaluated and a linear pixel shift induced. One method in which this may be achieved, is by taking the product of the pre-envelope from the image pair and obtain a weighted mean disparity at each resolution of filtering.

6.3 The Phase-Locked loop

There are two alternatives that we might propose to reduce the problems of disparity estimation by spatially tuned filters.

- Elastic stretching/compression of the image based upon local disparity estimates.
- Altering the spatial location of filters as a simulated phase locked loop.

The first notion, although extensively applied in speech processing as time warping holds limited possibilities in this approach (this approach is also the mechanism of Miller's[73] Phase-locked loop). This is a direct consequence of the uncertainty principle. Since we are applying multiresolution filters, we could anticipate warping the image from coarse to fine filtering. However, by warping the image we would expect to significantly alter the spectral properties of the original image. This was considered undesirable because of the pixelation of the original image and uncertainty in where to warp the image data. The second notion is interesting. Let us consider a pair of filters placed at corresponding retinal locations in both the left and right image. After convolution at a low resolution, we have a disparity estimate of d_{est} pixels. Let us induce a movement in the spatial location of each filter pair by $-\frac{d_{est}}{2}$ in both images. We have therefore reduced the disparity accordingly. We could then consider improving the disparity estimation, by convolving with the image again until a stable state is reached where there is zero-disparity from the filters response. This mechanism is similar to the initial proposal for binocular fusion. However, instead of initiating a vergence mechanism, we alter the spatial location of our filters based upon previous estimates at coarse resolutions. By coarse to fine matching, it is intended that spatial frequency estimates should converge to a stable 'fused' state, as we move to higher spatial frequencies (here we define a "fused state" as zero phase difference). For this reason, we might consider the proposal as locking in on the phase differences. We present results of this method in figure 6.1. Here, a single band-pass filter tuned to low spatial frequencies was convolved with a pair of images. Results are presented for two such image sequences. Stabilisation occurred after three iterations of the algorithm. Where a stable state was defined as no increments in disparity estimates greater than unity. The smooth disparity fields from the scan lines are also shown accordingly. However, the results indicate that phase wrap around and phase instability cause considerable problems. The phase wrap around problem is not only constrained to high spatial frequencies. This is apparent in figure 6.1b where the filter applied can in principle measure disparity to ±32 pixels. The largest disparity present in the image was approximately 24 pixels.

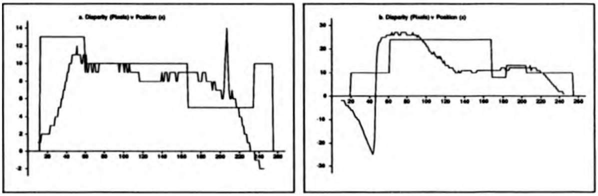


Figure 6.1: *Consequence of 3 iterations at the same resolution of bandpass filtering. Filters in both cases were tuned to 1/64 cpp with spatial position altered based upon previous estimates. (a) Figure B.3. (b) Figure B.2*

6.4 Algorithm 1

6.4.1 Method

The first algorithm that we have considered employs a local energy maxima constraint (taken from the pre-envelope product between stereo pairs). This algorithm works on the premise, that a large pre-envelope product at a given scale (in comparison to neighbouring scales of bandpass filtering) is indicative of a confident disparity measurement we store the local disparity estimate in a buffer. At finer scales, disparity estimates are then taken from filters with positions displaced by the current estimate of disparity. Updates of local disparity estimation are then constrained to a local energy constraint.

6.4.2 Local energy maxima

The algorithm is explained below in a Pseudo-code. The subscript σ_i will indicate the resolution of filtering with $i = 0$ indicating the coarsest and $i =$ last_scale the finest scale of filtering. $D_b[]$ will refer to a disparity buffer. More complex mathematical operations are defined within the Pseudo-code that cannot otherwise be made more explicit. The Pseudo-code is written in a way so that the ideas incorporated into the algorithm are concise. The Pseudo-code is not intended to be a rigid implementation of the algorithm details. For instance, to reduce machine storage it is not necessary to retain all the convolution coefficients throughout the program execution.

Pseudo-code

```
Define_Convolve()
For i = all_scales
For x = every_pixel_position
          {
/* Create two dimensional arrays containing the modulus-Argument representation
of both images convolved with a quadrature filter at each pixel and scale. */
          $\phi_l[x][i] = Arg[I_l(x) * \Psi(\sigma_i, \omega_i, x)];$
          $\phi_r[x][i] = Arg[I_r(x) * \Psi(\sigma_i, \omega_i, x)];$
          $E_l[x][i] = |I_l(x) * \Psi(\sigma_i, \omega_i, x)|;$
          $E_r[x][i] = |I_r(x) * \Psi(\sigma_i, \omega_i, x)|;$
          }

Begin_program
Read_image();
Convolve();
$D_b[] == 0;$
For i = second_scale to last_scale -1 step 1
{
For x = first_pixel to last_pixel step 1
          {
          $d = (\text{int}) D_b[x];$ /* integer disparity estimate */
          $\phi_d = \phi_r[x][i] - \phi_l[x-d][i]$ ; /* Phase differences from displaced filter positions */

/* Local energy maxima test */
          if ( $E_l[x-d][i-1]E_r[x][i-1] < E_l[x-d][i]E_r[x][i] > E_l[x-d][i+1]E_r[x][i+1]$)
                                                  $D_b[x] += \frac{\phi_d}{\omega_i}$ ;    /* update disparity buffer if true */
          }
Median_filter();
```

```
}
Print_out $D_b[]$;
End_program
```

Thus we anticipate that convergence will be obtained at the highest resolution of filtering.

6.4.3 Results

We present the iterated disparity measurements as the central position of the filters are changed, and the phase estimate from each filter pair at the locked state (fig. 6.2). The results from two separate image rasters are shown. The disparity estimates correlate very poorly with the measured disparities. The magnitude of the disparity responses are of the correct order, but we notice that we have oscillatory effects prior to disparate discontinuities and marked instability.

6.5 Conclusion

Their are many problems with this algorithm. In cases where phase advance is neither linear or monotonically increasing or decreasing, the phase difference will be in error, this is particularly true at disparity discontinuities where the spatial frequencies in both image planes are markedly different. A phase subtraction at such points would be unpredictable. We must therefore improve on our local knowledge of phase behaviour to reduce the noise present in our disparity estimates.

An early proposal for post-processing the phase response was based upon *median filtering* the disparity estimates after each resolution of bandpass filtering. Since median filtering holds the important property of removing spurious noise spikes, reducing oscillations, while preserving ramp and step edges, it seemed an appropriate operation to introduce into our algorithm. Figure (6.3) presents the results from a single raster sequence based upon 3 iterations of median filtering added to our algorithm. Disparity estimates were quantised to the nearest pixel. Median filtering has improved stability, and is effective in reducing the noise sensitivity of the phase locking iterations. However, the algorithm still appears unstable with potentially large disparity errors. Pursuing the notion of maximal energy response to

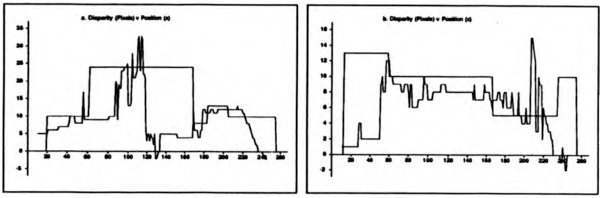


Figure 6.2: *(a) Phase locking based on local energy constraints from Figure B.2, $r = 0.59$; $\bar{x} = -6.3$; $\sigma_n = 0.96$ and (b) Same technique applied to Figure B.3, $r = 0.83$; $\bar{x} = -1.7$; $\sigma_n = 0.284$.*

justify the selection of spatial frequency channels for 'correctness' ceased to be of interest from this point hence. This is because of the inaccuracy of disparity interpretation and the poor representation of the local signal. We have already shown that quantising the spatial frequency components towards the center frequency of the filter causes a considerable error of 15-21% in disparity estimates. In complex data, this error is sufficient for higher resolution filters to lock on to incorrect signals in the image domain. This contributes a significant error to the method, which is also compounded by the inefficient use of bandpass filtering because we are only examining local energy maxima.

6.6 Algorithm 2

6.6.1 Introduction

To implement our next algorithm, we incorporate two added features into the basic technique of phase subtraction. First, we consider smoothing disparity estimates at each resolution of bandpass filtering and second, we introduce one method of recursively weighting disparity estimates between resolutions of filtering.

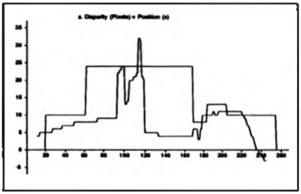


Figure 6.3: *Consequences of adding median filtering to the Phase Locking algorithm to Figure B.2. r=0.599;* $\bar{x} = -7.08$; $\sigma_n = 1.03$.

Smoothing

We know from Papoulis[81] that if we have a signal:

$$g(x) = g_f(x) + g_n(x)$$

where $g_n(x)$ is the output from a linear system, whose input n(x) is white noise, then for a constant $g_f(x)$ signal, the variance of g(x) is given by:

$$\sigma_g^2 = \mathcal{E}\{g_n^2(x)\}$$

The minimum mean-square estimation error of g_f smoothed with a window $w(x)$, is obtained if $w(x)$ is the truncated parabola given by:

$$w(x) = \frac{3}{4X}[1 - (\frac{x - x_o}{X})^2]p_X(x - x_o) \qquad (6.3)$$

where $p_X(x - x_o)$ is a pulse of width 2X. The operator referred to in equation (6.3) was derived from continuous mathematics and must be modified should $X < 4$ since the window loses the property of unit area. We argue that this measure is a suitable smoothing function to apply, because we have already shown that compression/expansion transformations between image pairs severely alters the instantaneous frequency between corresponding stereo pairs. This leaves correspondence by the method of phase differences difficult to interpret. We have also assumed that surfaces are opaque, and have therefore not considered transparent surfaces. Because of this, we are only interested in the local computation of a single disparity estimate. An optimal filter for the estimation of a constant parameter would then appear to be appropriate.

Recursive filtering

To implement a recursive filter implies that we have available to us a local parameter that can be applied to estimate the *strength* or confidence of correspondence. We have chosen to apply the square of the Willsky error measure:

$$Q(x) \;=\; \frac{1}{4}[1 \;+\; \cos\phi_d(x)]^2 \tag{6.4}$$

as the confidence measurent. We have chosen this function because the *phasewidth* (measured in the same proportional manner as bandwidth) is 0.695 radians and concentrated near to the origin on the phase circle. This function is then intended to retain a lock in phase differences during the progression from coarse to fine resolutions of filtering. We implement the recursive filter by defining our measured disparity by:

$$\bar{D}_{k+1}(x) \;=\; \frac{Q_k(x)\bar{D}_k(x) \;+\; Q_{k+1}(x)(D_{k+1}(x) + \bar{D}_k(x))}{Q_{k+1}(x) + Q_k(x)} \tag{6.5}$$

Where the measure k refers to the kth resolution of bandpass filtering, and $\bar{D}_{k+1}(x)$ refers to the measurement of disparity from the recent update measured by $D_k(x)$. Thus we are updating our measurements based on the *goodness of fit*, at successive resolutions. In keeping with recursive filter operations, which effectively incorporate information from all previous operations, we re-arrange and modify the above expression to obtain:

$$\bar{D}_{k+1}(x) \;=\; \bar{D}_k(x) \;+\; \frac{Q_{k+1}(x)}{Q_{k+1}(x) + Q_k(x)}D_{k+1}(x)$$

where we redefine Q_{k+1} to function recursively as:

$$Q_{k+1}(x) \;=\; \frac{Q_k(x)}{2} \;+\; \frac{\bar{Q}_{k+1}(x)}{2}$$

which we propose holds the properties that we require. Here $Q_{k+1}(x)$ refers to the measure of correspondence which is averaged with the previous value. Thus, if our disparity measurement at the current resolution gives a poor measure of correspondence i.e $Q_{k+1}(x) = 0$ then the disparity measurement does not change significantly. The converse naturally applies should the new weighting be close to unity. This measure has the advantages of incorporating resistance to the adverse affects of noise within a single bandpass channel because of the higher order weighting. Within resolutions of filtering, we have also applied the Willsky error measure to reduce the risk of phase wrap-around in the interpretation of disparity. By applying both a recursive filter, damping of disparity estimates and smoothing we run

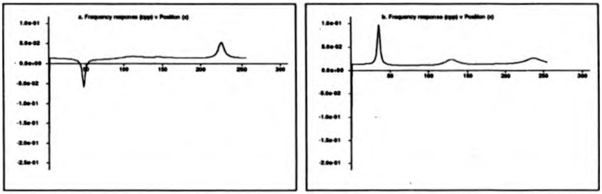


Figure 6.4: *(a) Instantaneous frequency response (y-axis) from a Gabor filter tuned to* $1.5x10^{-2}$ *cpp against position (x-axis) convolved with the intensity function in figure B.2a. (a) Original image (b) Mean intensity removed.*

the risk of overdamping the phase difference process so that a *phase lock* is never actually achieved. To compensate for this we propose to combine the camera vergence mechanism into our algorithm.

An additional feature of this algorithm lies with the subtraction of the mean image intensity from the image data. We have already seen in chapter 4, that d.c sensitivity does introduce difficulties in the representation of a linear phase response. Reducing the mean d.c from the image serves two purposes. First, it reduces the large pre-envelope response from filtering at the extremes of the image data. In this case, the phase responses are not biased by the edge effects of the image data as otherwise could be the case. Second, we have observed that the mean intensity levels can bias the phase response of filters tuned to very low spatial frequencies. For these reasons, we have removed the mean d.c level. This is highlighted in figure 6.4b, where the instantaneous frequency response is now positive from the removal of the mean image intensity level. Because we have removed the mean d.c, we have also inhibited the stationary phase behaviour that has been observed in figure 4.9a. This is because stationary phase is defined as a local sensitivity to d.c. We have in this case reduced the sensitivity to d.c for the Gabor filter tuned to $1.5x10^{-2}$ cpp, which is a low spatial frequency.

6.6.2 Methods

The 1-D algorithm based upon the measurement of disparity from the center frequency of the Gabor function is as follows:

- Subtract the mean intensity level from each pair of stereoscopic images.
- Convolve each image with the lowest resolution quadrature filters, and extract the mean disparity from equation (6.1). Store the weighting measures $Q(x)$ and increment the disparity buffer by the smoothed estimate of disparity.
- * Proceed to the next resolution and again extract phase differences, with filter locations at position I(x) displaced by the values stored in the disparity buffer at $D_b(x)$.
- Weight the new increments in the disparity buffer by disparity estimates obtained locally from phase differences *interpreted* from the center frequency of the filter, which are weighted by both the Willsky error measure and equation (6.4).
- Smooth the disparity buffer with the truncated unit area parabola from equation (6.3).
- Repeat from (*) until the filter centered at the the highest spatial frequency (0.25 cpp) has been applied.

Pseudo-code

```
Define_Convolve()
For i = all_scales
For x = every_pixel_position
          {
/* Create two dimensional arrays containing the modulus-Argument representation
of both images convolved with a quadrature filter at each pixel and scale. */
          $\phi_l[x][i] = Arg[I_l(x) * \Psi(\sigma_i, \omega_i, x)];$
          $\phi_r[x][i] = Arg[I_r(x) * \Psi(\sigma_i, \omega_i, x)];$
          $E_l[x][i] = |I_l(x) * \Psi(\sigma_i, \omega_i, x)|;$
          $E_r[x][i] = |I_r(x) * \Psi(\sigma_i, \omega_i, x)|;$
          }
```

```
Begin_program
Read_image();
Subtract_d.c_from_image();
Convolve();
/* Camera vergence mechanism returning $D_{mean}$ as a mean vergence */
Estimate_initial_vergence($D_{mean}$);
/* Initialise recursive weighting with mean phase difference at the coarsest scale */
$Q[x][0] = 0.25[1 + \cos(\omega_o D_{mean})]^2$ ;
$D_b[] == D_{mean}$;   /* Set disparity buffer to mean vergence value */
For i = first_scale to last_scale step 1
{
For x = first_pixel to last_pixel step 1
        {
        $d = (\text{int})D_b[x]$; /* integer disparity estimate */
        $\phi_d = \phi_r[x][i] - \phi_l[x - d][i]$ ; /* Phase differences from displaced filter positions */
        $Q[x][i] = 0.25(1 + \cos\phi_d)^2$ ; /* Willsky error measure */
        /* Update disparity buffer with recursive weighting */
        $D_b[x] = D_b[x] + Q[x][i]/(Q[x][i] + Q[x][i-1])\phi_d/\omega_i$; /* Update disparity buffer */
        $Q[x][i] = Q[x][i]/2 + Q[x][i-1]/2$ ;    /* Recursively define weighting through scales */
        }
Smooth_disparity_buffer_with_truncated_parabola();
}
Print_out $D_b[]$;
End_program
```

6.6.3 Results

We present some results based upon figure B.4. We have chosen this particular sequence because experimental work has showed from this stereo pair it is difficult to obtain stable depth estimates. This is because of the differences in image intensity recordings from the camera displacements. We also present some disparity measurements from other control data. It is apparent, that best results were obtained by initiating the final disparity measurements after the mean disparity measurements had been taken over all resolutions (fig. 6.5d). A full depth intensity surface is presented in figure 6.6.

6.6.4 Conclusion

Algorithm 2 has increased in stability by comparison with the similar data for Algorithm 1. However, the results are heavily damped in terms of the final depth estimates because of the heavy constraints from smoothing between scales of filtering, and also between scales of filtering. It is apparent with a 1-D filter, that the measurement of disparity using the center frequency of the filter is difficult because of phase error and local instability. We will now review the final 1-D algorithm which we have found to be the most successful for the 1-D approaches considered.

6.7 Algorithm 3

In this algorithm, we include the notion that the local instantaneous frequency can be used to interpret disparity. We also introduce the Compact Gaussian pyramid, which is a technique for subsampling image data. The primary advantage of subsampling image data lies with the efficient reduction in numerical computations. The technique also possess the advantage that fewer filter masks need to retained in computer storage because the same filters are applied through successive resolutions of filtering. However, the technique of interpreting disparity using the local instantaneous frequency also adds numerical difficulties because of stationary phase (i.e when the local phase gradient is zero). Under these conditions, the equations for interpreting disparity are undefined since there is no phase gradient. At present, we have applied two thresholds to prevent the interpretation of disparity within resolutions of bandpass filtering where the phase gradient is either zero or negative. The first threshold is based upon the disparity gradient limit[83] while the second threshold simply restricts all local phase gradients to lie within one octave of the center frequency of the filter. This is to prevent disparity estimation from instability because of the negative phase gradient which is not permissable within the Newton-Raphson iteration.

6.7.1 The Compact Pyramid

We have chosen to implement a method of image compression proposed by Burt and Adelson[15]. The technique takes an $C_o = M_C 2^N + 1$ by $R_o = M_R 2^N + 1$ image array of pixels and creates a subsampled image with new dimensions at the higher level (l) given by $C_l = M_C 2^{N-l} + 1$ and $R_l = M_R 2^{N-l}$. Here C_l and R_l refer to the dimensions of the subsampled image, and M_c, M_r, N are constants. Using this method, there are $0 < l < N$ possible levels of subsampled image. The method of subsampling is based on convolution i.e:

$$I_l(i,j) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} w(m,n) I_{l-1}(2i+m, 2j+n) \text{ for } i,j \ 0 \le i \le C_l \ , \ 0 \le j \le R_l \qquad (6.6)$$

where $I_l(i,j)$ refers to the subsampled image function at level l, and $I_{l-1}(2i+m, 2j+m)$ the unsampled image. The filter $w(m,n)$ is polar separable and can be described by:

$$w(m,n) = \hat{w}(m)\hat{w}(n) \qquad (6.7)$$

with:

$$\sum_{m=-2}^{2} \hat{w}(m) = 1 \text{ and } \hat{w}(i) = \hat{w}(-i)$$

Burt and Adelson[15] also suggested an *equal contribution* constraint which enforces each pixel to contribute the same weight towards each scale of subsampled image. They suggest:

$$\hat{w}(0) = a \ \hat{w}(1) = 0.25 \text{ and } w(2) = 0.25 - \frac{a}{2} \qquad (6.8)$$

with $a = 0.4$ as the best approximation to a digitised Gaussian. This technique has been used to compress image data in one-dimension in this chapter, and in two-dimensions in chapter 8.

6.7.2 The Disparity gradient

A geometrical constraint, which also has implications within the filter profiles lies with the disparity gradient hypothesis. The constraints that this principle implies allow a natural thresholding from phase differences applied to stereoscopic correspondence. To recap, the disparity gradient hypothesis, assumes that image data is obtained from opaque surfaces. Under these constraints, it is not possible for the disparity gradient to exceed $|2|$. A disparity gradient limit of $|1|$ is found to be both biologically and theoretically acceptable[83].

Consider the implications of this constraint in terms of phase differences and local instantaneous frequency analysis. First we consider an average disparity. This we define as

$(d_{av}(x))$:

$$d_{av}(x) = \frac{1}{2}\left[\frac{\phi_d(x)}{\omega_l} + \frac{\phi_d(x)}{\omega_r}\right] \tag{6.9}$$

The disparity gradient (Γ)[14], which is the spatial derivative of disparity becomes:

$$\Gamma = \frac{\omega_r^2 - \omega_l^2}{2\omega_l\omega_r} \tag{6.10}$$

Letting $\Gamma = 1$ and solving as a quadratic equation, we arrive at the simple relationship:

$$\frac{\omega_r}{\omega_l} = 1 \pm \sqrt{2}$$

The positive root, states that for the disparity gradient constraint to be satisfied, the ratio of matched frequency elements must not differ by more than 1.25 octaves. Precisely the relationship found by Julesz[45] with added bandpass noise to a pair of random dot stereograms.

6.7.3 Interpretation of disparity

Interpreting disparity from the local phase gradient[55] can be achieved by either the spatial derivative of phase, or by applying a quadrature pair of filters and their derivatives. In the latter case, it is also possible to obtain a least squares estimate of disparity by:

$$\mathbf{W_l}\mathbf{W_r}^T[\mathbf{W_r}\mathbf{W_r}^T]^{-1} = \mathbf{R_d} \tag{6.11}$$

$$\mathbf{W_r}\mathbf{W_l}^T[\mathbf{W_l}\mathbf{W_l}^T]^{-1} = \mathbf{R_d}^T \tag{6.12}$$

with:

$$\tan\phi_d(x) = \frac{\sin\phi_d}{\cos\phi_d} \tag{6.13}$$

as the least squares phase difference (ϕ_d), with coefficients taken from:

$$\mathbf{R_d} - \mathbf{R_d}^T = 2\sin\phi_d \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{6.14}$$

and

$$\mathbf{R_d} + \mathbf{R_d}^T = 2\cos\phi_d\ \mathbf{I} \tag{6.15}$$

Here $\mathbf{W_l}$, $\mathbf{W_r}$ refer to the 2x2 matrices of the Wronskian $(W(x,\omega,\sigma))$ in left and right image and R_d contains the least squares phase difference. One possible advantage of this technique lies in the presence of phase singularities[41] in Scale-Space, which do not necessarily occur at the same scale because the derivatives of Gabor function have a different spatial envelope.

6.7.4 Method

The basic algorithm that we now suggest is summarised as follows:

- subtract the mean image intensity from each image.
- Compress image data using the Compact Pyramid code.
- Obtain the mean disparity difference as an initial vergence mechanism stored in the disparity buffer using the vergence mechanism.
- * Calculate disparity estimates using the differences in local phase and weighted instantaneous frequency with filter pairs displaced by the current local estimate of disparity.
- Apply thresholds based upon the upper and lower cut-off frequency of the bandpass filter and the disparity gradient hypothesis.
- Smooth disparity estimates at a given resolution.
- Increment the disparity buffer based upon the curent disparity estimates and repeat from (*) at the next resolution until the highest spatial frequency.

Pseudo-code

```
Define_Convolve()
For i = all_scales
For j = Number_of_filters_at_each_scale
For x = every_pixel_position_at_each_scale
    {
/* Create three dimensional arrays containing the modulus-Argument representation
of both images convolved with a quadrature filter at each pixel and scale. */
    $\phi_l[x][i][j] = Arg[I_l(x,i) * \Psi(\sigma_j, \omega_j, x)];$
    $\phi_r[x][i][j] = Arg[I_r(x,i) * \Psi(\sigma_j, \omega_j, x)];$
    $E_l[x][i][j] = |I_l(x,i) * \Psi(\sigma_j, \omega_j, x)|;$
    $E_r[x][i][j] = |I_r(x,i) * \Psi(\sigma_j, \omega_j, x)|;$
    det_Wronskian($W_l[x][i][j], W_r[x][i][j]$);  /* Returns det|Wronskian| for both images */
```

```
    $f_{il}[x][i][j] = \frac{1}{2\pi} W_l[x][i][j] / E_l[x][i][j]$ ; /* Instantaneous frequency calculation */
    $f_{ir}[x][i][j] = \frac{1}{2\pi} W_r[x][i][j] / E_r[x][i][j]$ ; /* Instantaneous frequency calculation */
    }
Begin_program
Read_image();
Subtract_d.c_from_image();
$D_b[] = 0$;
Compress_image_with_Compact_Pyramid($I_l(x,i), I_r(x,i)$); /* Returns compressed image data */
Convolve();
For i = first_scale to last_scale step 1
{
    For j = first_in_each_scale to last_in_each_scale step 1
        {
        Forx = first_pixel to last_pixel step 1
            {
            $d = (\mathrm{int}) D_b[x]$; /* integer disparity estimate */
            $\phi_d = \phi_r[x][i] - \phi_l[x-d][i]$ ; /* Phase differences from displaced filter postions */
            /* Average Instantaneous frequency */
            $f_{av} = f_{ir}[x][i][j] E_r[x][i][j] + f_{il} E_l[x][i][j] / (E_l[x][i][j] + E_r[x][i][j])$;
            if ( All thresholds satisfied? )
/* Disparity estimation from instantaneous frequency with Willsky error weighting ($\eta$) */
                $D_b[x] += \eta \phi_d / f_{av}$;
            }
Smooth_disparity_buffer_with_truncated_parabola();
        }
Expand_disparity_buffer(); /* Inverse of compress */
$D_b[] = D_b[] * 2$ ; /* Mutiply disparity estimates by 2 */
}
Print_out $D_b[]$;
End_program
```

---

we have also implemented this algorithm on our control data by applying the Hilbert transform pair of the first derivative of the Gauusian. This we define by:

$$\hat{F}(\omega) = |\omega| \exp[-\frac{\omega^2}{4\sigma^2}] \quad (6.16)$$

which we expect to respond with linear phase when applied as a quadrature pair with:

$$F(\omega) = -j\omega \exp[-\frac{\omega^2}{4\sigma^2}] \quad (6.17)$$

which requires an inverse Fourier Transform to define the filter weights.

6.7.5 Results

We present results of our current algorithm shown in in figures 6.8 to 6.10. We can see that our results have improved considerably, particularly in comparison to our previous sets of data (Appendix A). It is particularly apparent that the Hilbert transform of the derivative of the Gaussian has in this case given improved results. To suggest the optimal filter for disparity interpretation is out of the scope of this thesis, however, we present the results to verify that the technique can be applied to several quadrature filter pairs. We also at this point present an intensity surface of disparity in figure 6.10. Intensity varies proportionally to disparity. The lamp in the foreground, window background and room walls have been separated. However, the algorithm still retains local instability, which is not surprising when considering that only a 1-D filter has been applied.

6.7.6 Conclusion

We have formally stated our algorithm for a phase based mechanism toward solving the correspondence problem. Implicit to the solution lies the notion of eye vergence to bring features into close correspondence. Although not absolutely necessary for all image data studied, we use it to reduce the error involved with disparity interpretation from phase differences.

There are four influential parameters that we can apply to reduce instability from the method of phase differences. First, Willsky's error measure on a circle reduces the risk of gross error from phase wrap-around which is highly possible, not only from large disparity differences between image data, but from phase subtractions taken from markedly different spatial frequency image signals. This application was of course suggested by Wilson and

Knutsson. Fundamental to stability, lies the notion of smoothing the disparity measurements based upon optimal smoothing filters. This notion is the key causal factor to the interpretation of disparity. To assist with stability, we proposed removing the d.c component from our image data, and inducing a weighted vergence mechanism based upon the mean disparity estimates. We could equally begin phase subtraction from a known predetermined vergence. In either case the end result increases the relative disparity obtained from finer resolutions of filtering. This is because each filter can measure both positive and negative disparities. By initiating a vergence mechanism, we effectively double the potential range of disparities that each filter can measure since we are not then restricting our search to the measurement of positive disparities. This process is applied *prior* to any disparity interpretation at any given resolution of filtering. This mechanism, we have found necessary to avoid overdamping our disparity interpretations because of the heavy constraints placed upon phase extraction.

6.8 A critical view of the method of Phase differences

From our results, it is clear that the method of phase differences is particularly suited for obtaining stereoscopic correspondence from large image features, or constant disparities across an image field. Under these conditions, good results can be obtained for very large disparities which are limited only by aliasing. In the presence of noise, disparity error can only be minimised because Gaussian noise affects the response of a phase sensitive process. Even with noiseless image data, the method of phase differences also has difficulty should the local instantaneous frequency properties between image pairs differ significantly. We have showed that the method of phase differences is analogous to that of the Newton-Raphson iteration. The limitations of this method are clearly defined. To reduce the divergence of this method, it is not desirable to iterate at a given resolution, unless the system is heavily constrained. Disparity averaging, recursive weighting and the Willsky error measure all serve to assist in the stabilisation of the process, in particular when progressing to the highest frequency elements of filtering. In addition, the methods presented here rely heavily on coarse filters initiating the correct vergence mechanism and that they themselves are not subject to phase wrap around. This is the fundamental weakness of the technique at present.

6.9 Future work

We have dealt with considerable work in this chapter. A 1-D theory of stereopsis has been developed based upon phase differences, with some encouraging results. The application of phase itself, however, is far from exploited. There are several issues still to consider. In particular, we might consider weighting our disparity estimates at a given resolution based upon the responses from bandpass filters centered at the next higher and lower spatial frequencies than the resolution of interest. Thereby reducing the probability of analysing regions with negative frequency responses, which from the method of phase differences are not analytic. Because of the nature of the stationary phase problem applying filters in triplets will undoubtably reduce this source of error. We have also inferred that there may be distinct advantages in applying multiple bandpass filters whose phase differences are in quadrature but have different spatial envelopes for a given modulation frequency. This may also possess improved signal to noise ratio.

Additionally, we might consider the application of phase unwrapping. The advantages of phase unwrapping (which primarily involves integrating over the differential of phase) lies with the recursive nature of this process, which effectively removes the phase aliasing problem. This approach is theoretically constrained by the assumption that the point from which unwrapping commences has itself not wrapped around. Coarse to fine vergence again reduces this risk. Although phase unwrapping showed considerable promise, stationary phase and negative frequency regions once again proved to be difficult to cope with in practice. Wilson (personal communication) also points out that phase unwrapping is difficult in two dimensions. In chapter 6, however, we will argue that useful phase information primarily lies in parallel to the alignment of the eyes.

We have, however, shown that the method of phase differences can provide accurate results on real image data.

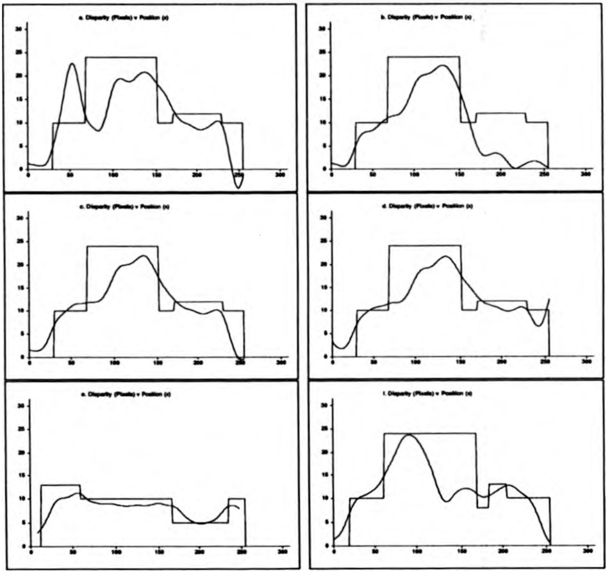


Figure 6.5: *Disparity from fig. B.4 interpreted from the filter center frequency. (a) Without recursive weighting, r=0.82;* $z = -2.23$*;* $\sigma_n = 0.49$ *(b) Mean image intensity included, r=0.88;* $z = -5.05$*;* $\sigma_n = 0.67$ *(c) Mean intensity subtracted, r=0.92;* $z = -2.32$*;* $\sigma_n = 0.40$*. (d) Mean vergence of 12 pixels obtained from successive coarse to fine filtering introduced as an the initial vergence, r=0.93;* $z = -1.75$*;* $\sigma_n = 0.34$*.(e) As (b) but applied to Figure B.3, r=0.95;* $z = -5.59$*;* $\sigma_n = 0.83$*. (f) As (b) but applied to Figure B.2, r=0.88;* $z = -3.03$*;* $\sigma_n = 0.52$.

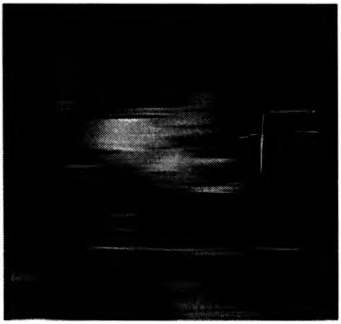


Figure 6.6: *Intensity disparity image produced from figure B.1 using the center frequency of the Gabor function to interpret disparity.*

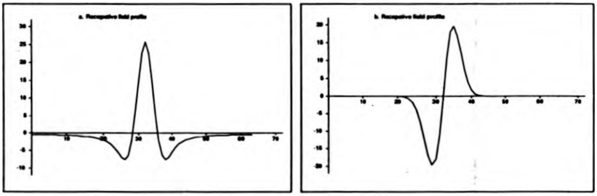


Figure 6.7: *The receptive field of the Hilbert pairs from the 1st derivative of the Gaussian. (a) Real part (b) Imaginary part.*

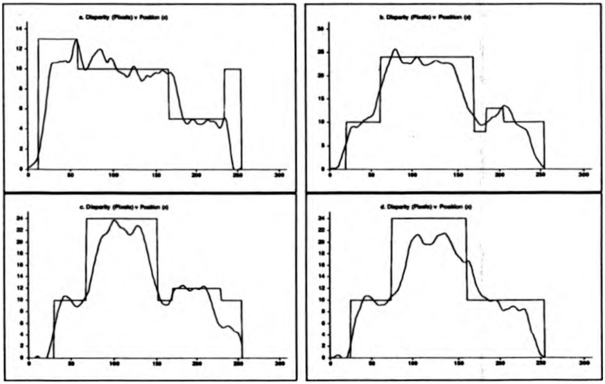


Figure 6.8: *Disparity interpreted using the local instantaneous frequency with Gabor filters and the Compact Pyramid. Center frequency of the Gabor function at 1/8 cpp and* $\frac{\sqrt{2}}{8}$ *cpp. (a) Figure B.3, r=0.89;* $z = -0.41$; $\sigma_n = 0.18$. *(b) Figure B.2, r=0.95;* $z = -1.41$; $\sigma_n = 0.28$. *(c) Figure B.5, r=0.96;* $z = -2.23$; $\sigma_n = 0.35$. *(d) Figure B.4, r=0.93;* $z = -2.60$; $\sigma_n = 0.41$.

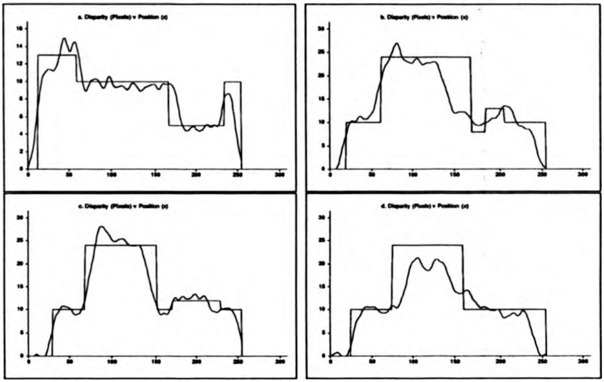


Figure 6.9: *Stereoscopic depth estimates using pyramidal subsampling and the Hilbert Transform pairs of the 1st derivative of the Gaussian. Center frequency at 1/8 and* $\frac{\sqrt{2}}{8}$ *cpp. Disparity interpreted by incorporating the stereoscopic continuity constraint and the truncated parabola for smoothing. (a) Figure B.3, r=0.92;* $\bar{x} = 0.45$; $\sigma_n = 0.15$. *(b) Figure B.2, r=0.93;* $\bar{x} = -1.53$; $\sigma_n = 0.33$. *(c) Figure B.5, r=0.96;* $\bar{x} = -0.65$; $\sigma_n = 0.20$. *(d) Figure B.4, r=0.93;* $\bar{x} = -3.00$; $\sigma_n = 0.46$.

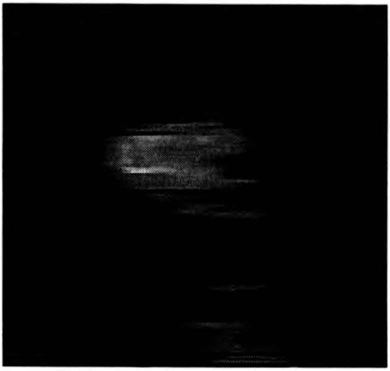


Figure 6.10: *Intensity disparity image produced from figure B.1 using the local instantaneous frequency from Gabor filters to interpret disparity.*

Chapter 7

FEATURE BASED ANALYSIS

7.1 Introduction

The method of phase differences, has been extensively applied in Chapter 5 towards the goal of stereoscopic correspondence. We have discussed the benefits, and limitations of this particular approach. Before progressing directly into 2-D filtering, it is worthwhile reviewing some work undertaken with regard to orientationally selective filtering. In particular, we will discuss methods for both corner and edge detection. We will also discuss curvature. Although curvature differences are not directly applied in our work concerning stereopsis at present, it is appropriate to include this type of computation for completeness, while identifying that the use of curvature may be included in future work.

7.2 Orientational Selective Mechanisms

We have so far only considered 1-D stereo matching. Thus we have completely omitted orientation from the work so far. That neurons in the visual cortex appear to respond to orientation preferences has been known since the early pioneering work of Hubel and Wiesel[39]. Blakemore et al[8] using single cell recording techniques, searched for other cues that might be utilised by the visual system for depth discrimination. As opposed to horizontal disparities, they were particularly interested in other geometrical transformations that might be used to disambiguate the perceived world. They found binocular cells, which when monocularly stimulated showed orientation preferences that differed by up to 15 degrees. They also showed that maximum response from each binocular neuron occurred with each receptive field stimulated by its optimal orientation. They concluded that certain binocular cells respond specifically to tilted objects in 3-D space. Early psychophysical evidence, comes from Braddick[11], who showed that lines with orientational separation of up to 15 degrees could be fused irrespective of their length. These results, were subsequently interpreted in terms of the disparity gradient hypothesis by Burt and Julesz[14]. Braddick interpreted his results by suggesting that stereopsis is dependent on orientation differences and not point by point disparity, a notion that complies with the known architecture of the visual cortex. Indeed, orientation can be a useful mechanism for disambiguating possible false matches, a notion known to physiologists almost a decade before computational vision workers applied the principle to computer algorithms.

7.3 2-D Filter Design

Physiological evidence for the relationships between orientation and spatial frequency tuning show considerable variation. There is evidence, however, that simple cells show preferences for orientation tuning[43] over spatial frequency. Typically, physiological data suggests that simple cells have elliptical profiles with width/length aspect ratios $\frac{\sigma_u}{\sigma_v} = \lambda$ ranging from 1/4 to 1 in favour of orientation selectivity. Here σ_v and σ_u refer to the principle axes defining the standard deviation of an elliptical Gaussian envelope .

Movshon[77] found that orientation and spatial frequency selectivity were well correlated: orientation half-bandwidths increased by about 10 degrees for each octave increase in spatial frequency, which is consistent with a fixed filter envelope aspect ratio ($\frac{\sigma_u}{\sigma_v}$). Thus, Daugman[18] concluded that while the ratio $\frac{\sigma_u}{\sigma_v}$ remained relatively stable, the product $\sigma_v \sigma_u$ corresponded to at least 1000:1 range of receptive field areas. In view of these observations, we will for the purpose of this thesis constrain ourself to an aspect ratio $\lambda = 0.65$ as a convenient mean of observed psychophysical responses. This is a limitation of our investigation. Thus, by fixing the aspect ratio, and implementing the conditions for linear phase quadrature Gabor filters from equation (8.8), we have sufficiently described the practical implementation of our filters.

To retain continuity with the previous chapters, we shall sample the 2-D frequency plane radially as in the 1-D case so we place filters centered at $\sqrt{2}n r$ intervals where:

$$r = \sqrt{\omega_x^2 + \omega_y^2}$$

refers to the modulus of the center frequency of the applied filters.

7.4 Corner and edge detection

Recent attention has been given to the detection of corners in low-level picture processing. Indeed, observations by Noble[78] have pointed out that conventional edge detecting operations fail to locate corners, and intersections of local image properties. While some measure of "cornerness" may have important applications in many aspects of image processing (e.g pattern recognition), the motivation for this work is derived as an extension of the work in the stereoscopic and motion aperture problem.

In particular, we wish to isolate image regions where the aperture problem might arise. Therefore, we are particularly interested in the location of edges, but we will also consider corners. It is shown experimentally in chapter 8 for a sine plaid that the best unbiased estimates for horizontal disparity occur with image signals separated by $\frac{\pi}{2}$ radians. We propose that measures of cornerness or by inverse logic finding a measure to describe an absence of edge information, provides an unambiguous confidence measure of precise motion and in particular, disparity vectors in an image plane.

7.5 Corner detection

Corner detection has previously involved the product of both edgeness, and curvature (rate of change of gradient direction with gradient magnitude)[78]. However, these techniques have primarily involved edge extraction followed by differential operators, which are noise sensitive and often require smoothing prior to picture processing. Clearly, any smoothing of an image, will also reduce the magnitude of gradient direction, and hence the possibilities of detecting corners. We propose a new method for locating corners in image pictures, by applying a band of orientationally selective 2-D Gabor filters tuned to a constant radial frequency, but different orientation preferences.

We consider the convolution of an image function with orientationally selective Gabor filters with an envelope aspect ratio $\frac{\sigma_v}{\sigma_u} = 0.5$. The aspect ratio is chosen to reduce the noise sensitivity at the presence of an image corner, but at the expense of increased processing time. Let us consider a circular band $\Psi(x, y, u_i, v_i)$ of Gabor filters where:

$$\Psi(x, y, u_i, v_i) \;=\; \frac{1}{\sqrt{2\pi\sigma_x^2\sigma_y^2}} \exp[-\frac{x^2}{4\sigma_x^2} - -\frac{y^2}{4\sigma_y^2}]\ \exp[2\pi j(u_i x + v_i y)] \quad 0 \;\leq\; i \;\leq\; N-1 \qquad (7.1)$$

and

$$\sqrt{u_i^2 + v_i^2} \;=\; r \qquad (7.2)$$

$$\tan^{-1}(\frac{v_i}{u_i}) \;=\; \theta_i \qquad (7.3)$$

$$\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \;=\; 16 \qquad (7.4)$$

such that each filter is applied to a constant radial frequency r, and tuned to an orientation θ_i.

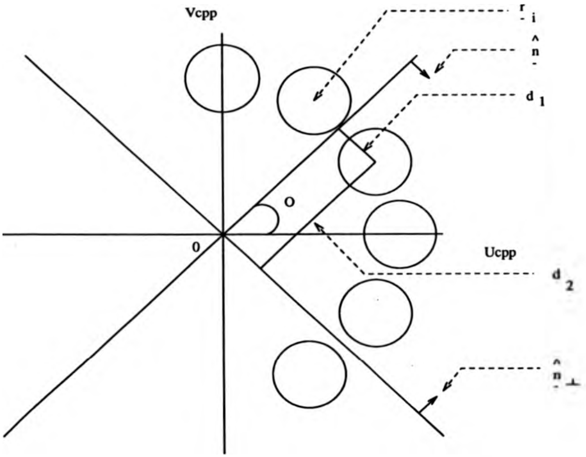


Figure 7.1: *A band of $\Psi(x, y, u_i, v_i)$ filters applied in a circle and considered in the 2-D Fourier plane, with equal spatial frequency but different orientation preferences. If we model a corner as a pair of orthogonal lines represented by the unit vectors $\hat{n}$ and $\hat{n}_\perp$ in the Fourier domain, then we minimise the product of the normal distance between each filter, and the image stimulus.*

Consider a multi-orientated pre-envelope taken at a localised corner in the image plane (fig. 7.1). This is formed by the linear operation:

$$E(\theta_i, u_i, v_i; x, y) = |\Psi(x, y, u_i, v_i) * I(x, y)| \tag{7.5}$$

where $I(x, y)$ refers to a two dimensional image function and we consider the filter $\Psi(x, y, u_i, v_i)$ applied in several different orientations. Here the two filters with orientation preferences separated by $\frac{\pi}{2}$ radians in the closest proximity to the edges of the corner will respond with maximal energy. Using a simple model, we assume that the energy response will be inversely proportional to the normal distance (in the the frequency domain) between the center frequency of the filter and the edge which in this case is approximated by a line in the Fourier domain. This is also equivalent to the statement that the filter will respond to the local frequencies in the 2-D Fourier plane that are perpendicular to the center frequency of the filter. For corner detection, we apply weighted least squares fit to the energy output of each quadrature filter pair, and noting that we must minimise the *product* of a pair of vectors taken from each filter to the line of best fit i.e:

$$Q(\theta; x, y) \;=\; \sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y)\ [|<\vec{r}_i, \vec{n}>|^2| <\vec{r}_i, \vec{n}_\perp>|^2] \tag{7.6}$$

where r_i is a position vector representing the i_{th} filter in the frequency domain (fig. 7.1), and $\vec{n}$ is a unit vector with whose normal $(\vec{n_\perp})$ is used to model the orientation of Fourier energy present at a corner. $Q(\theta; x, y)$ refers to a weighting taken from all the filters applied at a point (x, y) in the image domain. In figure 7.1, the lengths d_1 and d_2 are the lengths whose squared products we wish to minimise. This gives:

$$Q(\theta; x, y) \;=\; \sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y)\ r^2 \cos^2[\theta_i - \theta] \sin^2[\theta_i - \theta] \tag{7.7}$$

differentiating with respect to θ we have:

$$\begin{aligned}\frac{dQ(\theta; x, y)}{d\theta} &= \sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y) r^2 4 \cos 2[\theta_i - \theta] \sin 2[\theta_i - \theta] \\ &= \sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y) r^2 \sin 4[\theta_i - \theta]\end{aligned} \tag{7.8}$$

setting the derivative equal to zero we finally obtain:

$$\tan 4\theta \;=\; \frac{\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y)\ \sin 4\theta_i}{\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y)\ \cos 4\theta_i} \tag{7.9}$$

Equation (7.9) provides a definition for the orientation of the local corner in the picture plane. However, it is only possible to distinguish between corners separated by mod $\frac{\pi}{4}$ radians in the Fourier domain. Thus by considering local operations it is not possible to separate "X" from "L" junctions based upon an energy response alone. This is because of the one-sided nature of the Hilbert Transform. Similarly, it is not possible to distinguish an angular separation between a pair of lines of η and $\pi - \eta$ by considering a local pre-envelope. Should we consider processing the pre-envelope response from a circular band of filters with a *radial* spatial separation of r pixels, then we would expect to resolve these ambiguities. Knutsson, Wilson and Granlund[48] have also shown that edge orientation may be obtained using a similar method. They took Willsky's[97] error measure on a circle, here we again consider a weighted least squares fit of the filter responses to a line through the origin to give a similar result[54]:

$$\tan 2\theta = \frac{\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y) \ \sin 2\theta_i}{\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y) \ \cos 2\theta_i} \tag{7.10}$$

As an interesting observation we note that the zero-crossing from the response patterns of an odd directionally selective 2-D function, which is also applied in a circle can also indicate the orientation of an edge. This is because the zero-crossing in this case corresponds to the normal to the edge. Which is a simple method for extracting the edge tangent. Such a scheme, however, cannot itself give a probabilistic measure for edgeness and therefore may not be useful from this point of view.

We present the responses taken from a circular band of 2D-Gabor filters to both an ideal edge, and corner (fig. 7.3). Notice the energy responses indicate a periodicity of $\frac{\pi}{4}$ radians for a corner, and $\frac{\pi}{2}$ radians for an edge which is obtained by mapping the energy responses from Gabor filters onto the real line $\theta \in [0, \pi]$.

As a measure of edge and corner confidence, we take the linear Fourier transform of the pre-envelope from each band of filters at $2\theta_i$ and $4\theta_i$, which corresponds to the expected pre-envelope response indicating either edges or corners. This simplified model does have restrictions for Gabor filters because the model proposed is only an approximation. This is because we have not considered the envelope of the filter as a parameter of the model.

Consider a sine grating (or edge) sampled by 2-D Gabor filters with a fine orientation bandwidth. The actual pre-envelope response is a narrow Gaussian energy function whose peak is centered at the filters with similar orientation preference (fig. 7.4). The duration

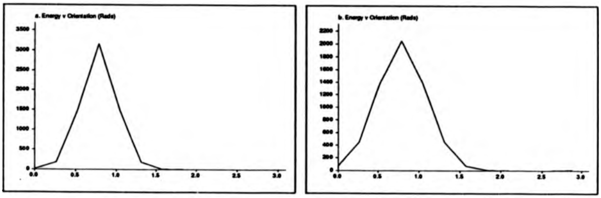


Figure 7.2: *(a)(y-axis) Energy responses from a simple 2-D sine grating against orientation (x-axis). The grating (1/16 cpp) was orientated at $\frac{\pi}{4}$ radians. Filters tuned to 1/8 cpp where convolved with the data with width/length spatial aspect ratio of 0.65. Edge probability estimated at 90.9%, orientation :1.571 radians. (b) As (a) but with orientation bandwidth reduced with a width/aspect ratio =1.0. Edge probability estimated at 83.1%, orientation: 1.570 radians.*

of the response is dependent on the properties of the filter envelope. For fine orientational tuning the pre-envelope responce as a function of orientation will tend to an impulse, while for coarse orientational tuning, the envelope will correspondingly spread out as a function of orientation. By the scaling theorem, the Fourier transform of a narrow Gaussian signal covers a wide spectrum with significant energy at both $2\theta_i$ and $4\theta_i$ sampled frequencies, thus making edge and corner measures inseparable. This is because equations (7.9) and (7.10) require the ordinary Fourier Transform applied to the energy responses of filters in the circle. Since the Fourier spectrum will be broad in this case, we cannot expect to isolate both corner and edge measures from this type of response. Fortunately, orthogonal edge intersections provide a unique response. This can be seen by considering the discrete Fourier Transform from the response at both corner and edge (fig. 7.4c). Even with the limitations of the model we have considered, it is clear that the frequency component that we are primarily interested in for the detection of corners responds with the largest magnitude (ignoring d.c). In this case, corner and edge measures are now separable. By reducing the orientation preference of the Gabor filters, the orientational bandwidth of the filter responses can be controlled so that there is little overlap between the frequency responses of interest, but at the expense

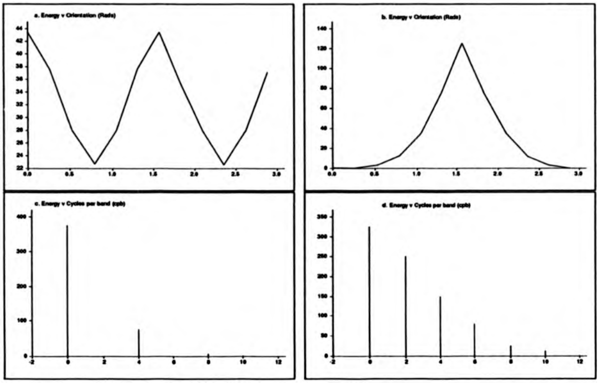


Figure 7.3: *Energy response from a band of 2-D Gabor filters. Aspect ratio 1.00 and center frequency 1/4 cpp, (a) to an ideal edge (b) corner. (c) The discrete Fourier transform of the response in (a). (d) The discrete Fourier transform to the response in (b).*

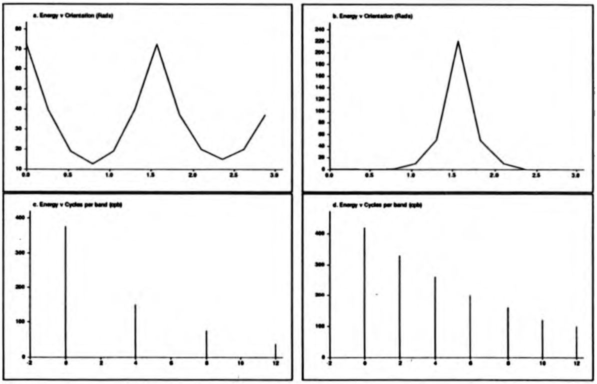


Figure 7.4: *Energy response from a band of 2-D Gabor filters. Aspect ratio 0.65 and center frequency 1/4 cpp. (a) to an ideal edge (b) corner. (c) The discrete Fourier transform of the response in (a). (d) The discrete Fourier transform to the response in (b).*

of increased noise sensitivity (fig. 7.3) to corner detection. Therefore, to reduce the noise sensitivity, filters should have fine orientation bandwidth, and the probability of detecting an edge must be established a priori before a corner can be declared. This can be represented by the following equations:

$$E_{dge} = \frac{[\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y)\ \cos 2\theta_i]^2 + [\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y)\ \sin 2\theta_i]^2}{[\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y)]^2} \quad (7.11)$$

$$C_{orner} = \frac{[\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y)\ \cos 4\theta_i]^2 + [\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y)\ \sin 4\theta_i]^2}{[\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y)]^2} \quad (7.12)$$

Equations (7.11) and (7.12) normalise the energy responses from the band of filters for both corner and edge detection. Since the edge response is small at a corner intersection, then a measure of corner confidence may be calculated from:

$$C_{onfidence} = C_{orner} - E_{dge} \quad (7.13)$$

By setting an appropriate threshold, measures of both edge and corner measurements can be obtained. In practice, a Gaussian matched filter (frequency domain) was applied to improve the probability of corner detection. A small pre-envelope threshold was also applied to filter out low pre-envelope image regions.

7.5.1 Orientation Results

We have observed the energy response by applying a band of orientationally selective filters to both an ideal edge and corner. Let us now apply equations 7.9 and 7.10 to estimate orientation. We have already applied this equation in figures 7.2 to 7.4. We consider added isotropic Gaussian noise to the estimates of orientation. In figure 7.5 we show the error in orientation estimate with added noise. As expected, the variance increases with the variance of added noise. Because there is in principle no difference between the phase error between directional quadrature filter pairs and the phase angle obtained through orientation estimates, the distribution of phase error is very similar to that observed in chapter 5. However, in this case the best fitting density function (shown as crosses) was found to be the Bivariate Gaussian density function. The most important result lies with increased stability for orientation estimation. The signal to noise ratio between the quadrature phase estimation used in chapter 5 and here was maintained. However, there is a noticeable difference between the performance

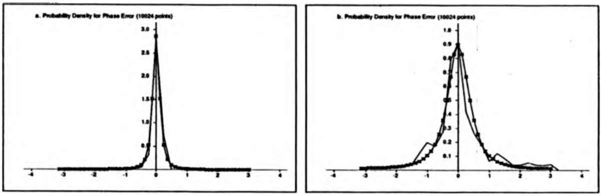


Figure 7.5: *Orientation estimates for a single grating (1 intensity unit peak to peak) using 8 orientationally selective filters to obtain orientation for (a)* $N_o = 3.0$ $r = 0.985$ *and (b)* $N_o = 6.0$ $r = 0.87$.

between figure 5.4e and 7.5b. This is to be expected, because we are both taking a least squares approximation to orientation and using 2-D filters to estimate the image properties. We also present results indicating the estimation of corner orientation with the same quantities of added noise (fig. 7.6). Care was taken to ensure that the signal to noise ratio was maintained but with an image stimulus comprising of the sum of two orthogonal sinusoidal gratings. Interestingly there is little difference between the noise sensitivity estimates for corners and edges. However, we point out that at ideal image corner each orientationally selective filter can only integrate along a half of its own length. This would still explain the noise sensitivity of corner detection as opposed to a "X" in the image domain.

We are particularly interested in the noise sensitivity from directional filtering, since we wish to interpret disparity in the least squares sense from a band of filters applied in a circle. Any error in orientation measurement, will correspondingly induce a measured error in disparity measurement.

As a further example, we present an orientation field from figure B.1a. In this case, we show orientation at regularly spaced intervals as a unit vector, whose magnitude is represented by the edge probability measure that we have defined in equation (7.10). In this instance, the center frequency of the Gabor filter was |0.25| cpp and circularly symmetric in terms of the envelope. Circular symmetry was chosen because we were primarily interested in representing edge orientation.

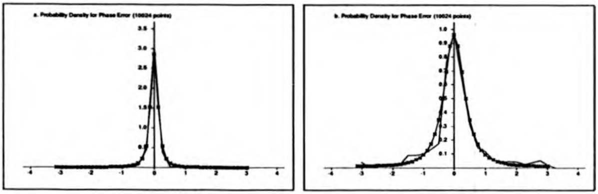


Figure 7.6: *Orientation estimates for a pair of orthogonal gratings ($\frac{1}{\sqrt{2}}$ intensity units peak to peak) for (a) $N_o = 3.0$ $r = 0.985$ (b) $N_o = 6.0$ $r = 0.885$.*

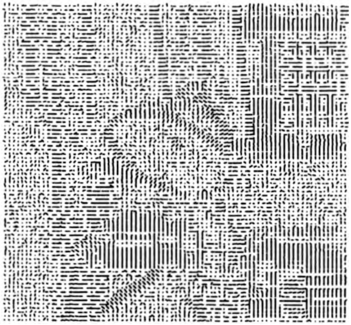


Figure 7.7: *Orientation represented as a vector from figure B.1. 8 filter pairs were applied with unity orientation/spatial frequency preferences.*

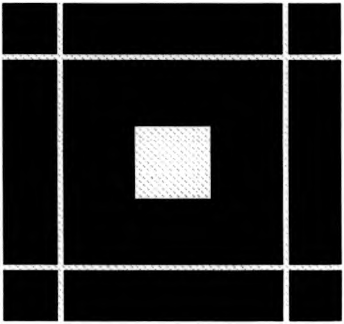


Figure 7.8: *Ideal square image used for both edge and corner extraction.*

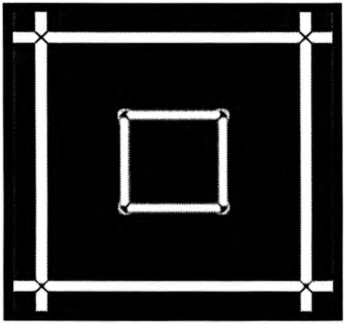


Figure 7.9: *Edge extraction probability from "square" image.*

Figure 7.10: *Corner extraction probability from "square" image.*

7.5.2 Corner and Edge Results

We present the results of the work concerning both corner and edge detection. We show the processing of a synthetic square and our *room* to both edge and corners confidence measures. These are presented in figures 7.8 to 7.12. Here the bright image regions represent the magnitude of the corner and edge probability measures. Notice that we have correctly indicated image regions that exhibit energy describing both edges and corners.

Unlike traditional edge and corner extraction algorithms, we are not necessarily interested in the exact location of the corner or edge. This is because the technique for disparity measurement that we are considering does not require an accurate measure of edge location. We will argue in chapter 8, that it is useful to retain a probabilistic measure of edges because of an aperture problem[1] in stereopsis. The aperture problem arises because of the ambiguity in inferring a translation over a two dimensional region in the presence of local image signals that are primarily oriented in a single direction. We will also use the probabilistic function: $1 - P_e$ to infer local image regions that are either isoluminant or contain more than one local orientation of dominant energy. In the latter case this could be a corner.

7.5.3 The general case

We can extend our previous arguments to the general case of finding the internal angle of separation between a pair of edges or a piecewise linear approximation to the change in angle of a curve. If we consider some arbitrary curve passing through a pixel, then we assume a piecewise approximation to the curve, as two edges separated by an unknown angle 2θ. We then define a unit vector $\vec{\hat{n}}_c$ which is perpendicular to the piecewise linear approximation to one of the edges of the curve and passes through $r_i = (x_i, y_i)$, which is a position vector representing the ith filter. We also define a unit vector $\bar{N}$ which is obtained from convolving in a circle spatially and represents the normal of the curve at the point of interest (x, y). The orientation of the normal can be found from:

$$\tan 2\theta = \frac{\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y) \sin 2\theta_i}{\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y) \cos 2\theta_i} \tag{7.14}$$

Since we have defined the normal to the curve which we will assume bisects the angular separation of our two edges. We also define $\vec{\hat{n}}_{cr} = [\cos\theta, -\sin\theta]^T$ as the reflection of $\vec{\hat{n}}_c = [\cos\theta, \sin\theta]^T$ about the normal by an angle θ. It is necessary to redefine our position vectors of the band of filters in the frequency domain relative to the normal of the curve:

$$\vec{r}_{i\theta} = \mathbf{R}_{\theta\perp} \vec{r}_i \tag{7.15}$$

Where $\mathbf{R}_{\theta\perp}$ represents the rotation matrix of the orientation of the normal to the pair of lines and:

$$\bar{N} = [\cos\theta_n, \sin\theta_n]^T \tag{7.16}$$

Here θ_n represents the orientation of the normal to the curve. Using the same notation that we used to define the orientation of a corner we have:

$$d_1 = < \vec{r}_{i\theta}, \vec{\hat{n}}_c > \tag{7.17}$$

and d_2:

$$d_2 = < \vec{r}_{i\theta}, \vec{\hat{n}}_{cr} > \tag{7.18}$$

We then form a measure $Q(\theta, x.y)$ as before based upon the weighted response from all filters applied to the circle, which we wish to minimise i.e:

$$Q(\theta; x, y) = \sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y) \,|< \vec{r}_{i\theta}, \vec{\hat{n}}_c >|^2 |< \vec{r}_{i\theta}, \vec{\hat{n}}_{cr} >|^2 \tag{7.19}$$

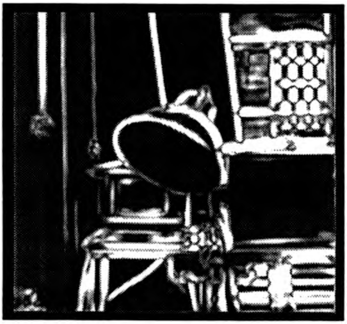


Figure 7.11: *Edge probability from figure B.1.*

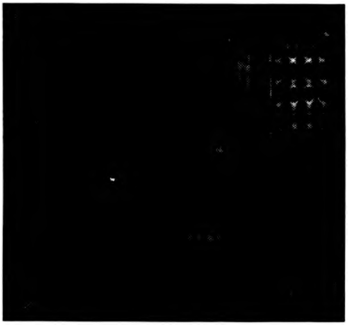


Figure 7.12: *Corner probability from figure B.1 .*

We can also see that:

$$\bar{r}_{i\theta} = r[\cos(\theta_i - \theta_n), \sin(\theta_i - \theta_n)]^T \tag{7.20}$$

where r represents the radial distance of the ith filter from the origin in the frequency domain i.e:

$$Q(\theta; x, y) = \sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y) \ \cos^2[\theta - (\theta_i - \theta_n)] \cos^2[\theta + (\theta_i - \theta_n)] \tag{7.21}$$

Differentiating $Q(\theta; x, y)$ with respect to θ, we obtain:

$$\begin{aligned} \frac{dQ(\theta; x, y)}{d\theta} &= \sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y) r^2 [4 \cos^4(\theta_i - \theta_n) \sin^3 \theta \cos \theta - \\ &\quad 4 \sin^4(\theta_i - \theta_n) \cos^3 \theta \sin \theta - \frac{1}{4} \sin 4\theta \sin^2 2(\theta_i - \theta_n)] \end{aligned}$$

Setting the derivative to zero at the minimum, then we arrive at:

$$\tan^2 \theta = \frac{\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y) \sin^2(\theta_i - \theta_n)}{\sum_{i=0}^{N} E^2(\theta_i, u_i, v_i; x, y) \cos^2(\theta_i - \theta_n)} \tag{7.22}$$

The above expression is a generalisation of the internal angle between a pair of edges or a segment of a curve. A little examination reveals that the form of the solution lies in terms of the variance or spread of energy from the band of filters. We know from signal processing theory, that the variance or second moment in the frequency domain, is equivalent to differentiating twice in the spatial domain, which is indeed the form of the rate of change of angle. The quadratic solution for the change in angle also suggests the possible ambiguity introduced by the application of orientationally selective filters at the same retinal location.

This expression, however, assumes that the orientational selectivity of each filter can be modeled as an impulse function. It is not difficult to introduce additional constraints which also accounts for the spread of energy as a cosine function from the preferred orientation. However, we do not pursue this topic further. This is because the measure for orientation differences requires an appropriate model to describe the exact orientational tuning of each filter. Such a model would indeed be sensitive to both the orientation and spatial frequency bandwidth that each filter possessed and therefore not generally applicable as a phase relationship. We will, however, review the work of Koenderink and Richards[50], which although not explicitly used in this thesis, bears significance as an important phase relationship.

7.6 Curvature

It would be appropriate to include a small section reporting some work on image curvature. The curvature K of a contour or edge is defined to be the rate of change of slope of the tangent vector to the contour with respect to distance along the contour. Where the radius of curvature is simply the reciprocal of K. We can define curvature in vector space:

$$|K| = \frac{d\mathbf{T}}{ds} \tag{7.23}$$

as the first Serret-Frenet equation of differential geometry[40]. Koenderink and Richards[50] showed that the second derivative taken over an isoluminant region L(x,y) is:

$$\frac{d^2y}{dx^2} = \frac{-L_{xx}L_y^2 + 2L_{xy}L_xL_y - L_{yy}L_x^2}{L_y^3} \tag{7.24}$$

with

$$\frac{dy}{dx} = -\frac{L_x}{L_y} \tag{7.25}$$

as the tangent of the edge. This equation has been successfully implemented using the *Canny edge detector*[16]. Since the curvature K over L(x,y) may also be shown to be:

$$K = \frac{\frac{d^2y}{dx^2}}{[1 + (\frac{dy}{dx})^2]^{\frac{3}{2}}} \tag{7.26}$$

then in the special case where the x axis is tangent to the local boundary of a curved edge, $L_x = 0$ and by comparing the above two equations we see that:

$$K = -\frac{L_{xx}}{L_y} \tag{7.27}$$

Koenderink and Richards then proposed convolving orientationally selective filters in a circle. They argued that the filter with maximal output then indicates the orientation of the boundary from which the local coordinate frame may be calculated to apply their equations. The expression obtained by Koenderink for curvature, is only valid at edges. The same may also be attributed to the calculation of orientation using the Canny edge operator from equation (7.25).

Using the Hilbert transform, we can extend the calculation of orientation to an image field of arbitrary phase. We form:

$$\Psi(x,y) = L(x,y) + j\hat{L}(x,y) = k \tag{7.28}$$

where $\Psi(x,y)$ refers to the 2-D analytic signal with $L(x,y)$ and $\bar{L}(x,y)$ as Hilbert transform pairs. k is a constant, which enables us to define an orientation from $\frac{dy}{dx}$. This equation gives an orientation (θ) which is reduced to a resolution of $\frac{\pi}{2}$ radians. i.e:

$$\tan\theta \;=\; -\frac{|L_x + j\hat{L}_x|}{|L_y + j\hat{L}_y|} \tag{7.29}$$

which would directly obtain orientation throughout the phases of an image field. This expression merely requires the derivatives of a Hilbert pair orthogonal in orientation. The phase information which is not considered in this definition is still valid for the method of phase differences. This approach indicates that first order directional derivatives of a Hilbert pair may well yield improved least-squares estimates for orientation, since the expression is indeed a phase relationship. That Biology applies first order kernels to extract spatial orientation, has recently been determined psychophysically[27]. This notion may equally apply to improve the calculation of velocity from the relationship $\frac{\omega_t}{\omega_x}$ as a spatio-temporal orientation.

By similar arguments, we can also define the curvature as:

$$K \;=\; -\frac{|L_{xx} + j\hat{L}_{xx}|}{|L_y + jL_y|} \tag{7.30}$$

which would also be independent on the quadrature phase of the signal. The expression for curvature about regions of even symmetry that Koenderink and Richards considered, required the further examination of third and second order partial derivatives of an image function, which would be highly noise sensitive.

7.7 Phase relationships in Image processing

It is interesting to notice that equations for orientation, and curvature all require an inverse tangent operation. Similarly, in chapter 4, we have also extensively applied inverse tangent operations to locate the *displacement* of a signal from the location of our quadrature filter pairs using the method of phase differences. Instantantaneous frequency may also be locally defined as the spatial derivative of phase, which may be obtained by the change in phase from spatially separated filters or equivalently, by applying the 1st order quadrature derivatives of a Hilbert transform pair of operators. Orientation and curvature are also obtained from phase relationships. In particular, Koenderink and Richard's expression for curvature in equation (7.27) is a rather unusual phase relationship in the circle of orientation differences, since it requires the ratio from an odd cell to even cell, who themselves are orientated in

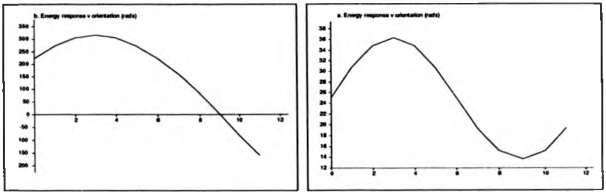


Figure 7.13: *An example of Koenderink's equation for curvature extraction. (a) The circular energy responses from the 1st directional derivative of the Gabor function. (b) Second directional derivative. Isoluminant circle with curvature of 0.03. Estimated curvature 0.043. 12 filters applied in a circle tuned to 1/8 cpp. Notice the zero-crossing from the imaginary filter response corresponding to edge normal. Edge orientation from least squares estimate at 1.5707 radians.*

quadrature in the spatial domain. This scheme we have extended by further use of phase relationships. Indeed, the essential characteristic of a phase lies with the division between orthogonal parameters. In terms of convolution filters, the division often removes the envelope from consideration, leaving a descriptor based upon the image function alone.

7.8 Classical End-stopped Cell

Recent physiological investigations provide evidence for a curvature sensitive cell referred to as the end-stopped cell. Koenderink and Richards[50] observed that a similar receptive field may be obtained from the third derivative of a Gaussian (L_{xxy} in their notation) we, however, show that this cell may be obtained quite simply from the addition of two sets of orientationally selective Gabor filters in orthogonal directions.

First we will define and show the end-stopped cell (fig. 7.14) by:

$$F_I(x,y) \;=\; K\exp[-\frac{x^2}{4\sigma_x^2}-\frac{y^2}{4\sigma_y^2}]\sin(\omega_x x)\cos(\omega_y y) \tag{7.31}$$

by noticing that we can expand the product of the oscillatory terms we have:

$$\sin(\omega_x x)\cos(\omega_y y) \;=\; \frac{1}{2}(\sin(\omega_x x+\omega_y y)+\sin(\omega_x x-\omega_y y)) \tag{7.32}$$

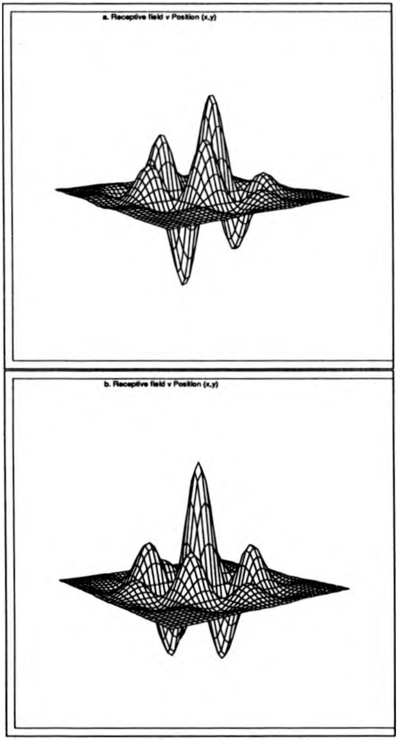


Figure 7.14: *(a) Classical form of the "Imaginary" end stopped cell formed from the addition of two Gabor functions with unity width/length aspect ratio but orthogonal orientation preferences. (b) "Real" end-stopped cell.*

which when multiplied with the Gaussian envelope is the sum of two orthogonal Gabor functions providing $\omega_x = \omega_y$. In the frequency domain, this representation is equivalent to 4 Gaussian distributions centered at $[|\omega_x|, |\omega_y|]$ in the imaginary plane. For a similar cell in the real Fourier plane, we can similarly show that:

$$F_R(x,y) \;=\; K\exp[-\frac{x^2}{4\sigma_x^2} - \frac{y^2}{4\sigma_y^2}]\cos(\omega_x x)\cos(\omega_y y) \tag{7.33}$$

Is also a real component of the classical end-stopped cell and may also be shown to be equivalent to the sum of two real Gabor functions with orthogonal directional preferences. We should, however, observe that these receptive field profiles, are equivalent (in the case of unity aspect/width ratio) to the operators we have applied to obtain corner responses.

7.9 Summary

We have shown in principle how localised 2D-Gabor filters may be applied to the problem of corner and edge detection. We have also obtained a function for describing the internal angle between pairs of line elements using Gabor filters. However, the nature of this last relationship incorporates rigid assumptions regarding the orientational bandwidth of the filter. These assumptions do not permit a correct model to describe the internal angle between a pair of lines or edges in the image domain.

We have also reviewed work for obtaining edge orientation, which we have inferred is particularly important in the extension of the theoretical body of this work into two dimensional stereoscopic computation. It is proposed that post-processing a band of 2D-Gabor filters for corner detection provides a confidence measure for methods that claim to solve both the stereoscopic and motion aperture problem.

Chapter 8

TWO-DIMENSIONAL ANALYSIS

8.1 Introduction

The extension of the theory to two dimensions is not simple. In particular, 2-D image functions immediately expand the computation power by at least a factor of mN^2: where N is the spatial number of computations in the 1-D case and m is the number of orientationally selective filters (we have ignored polar separability), and contain their own special properties to which increased attention has been paid. There is evidence that stereopsis also has some components of vertical disparity induced by the horizontal separation of the eyes, although the application of this form of disparity interpretation remains uncertain. It is well accepted, however, that the vertical component of disparity is considerably less than its horizontal counterpart, and as a consequence has been ignored in many stereo algorithms[83]. The role of vertical disparity may have considerable benefits in its estimation. Using vertical disparities Longuet-Higgins[60] has shown in principle, that for precise depth to be extracted from a stereo pair it is essential that at least three points must be correctly matched and that they must not lie in a vertical plane to the eyes. Longuet-Higgins, however, assumed that only relative disparity was available and that the spatial location of disparity on image sensors an unknown parameter. Such an assumption may well be correct with mammalian systems, and therefore implies that without the additional vertical disparity component precise depth information cannot be obtained. Relative depth, however, would be possible. This question has caused some problems for experimental psychologists to determine.

A further causal effect of vertical disparity lies with the physical displacement of the optical sensors. For a visually guided autonomous vehicle, it is possible over a period of time, that cameras become misaligned. Indeed, the biological ocular system also experiences a similar problem[70]. In order to self calibrate, it would be important to have a stable reference point such as a nose and an intrinsic sensitivity towards vertical disparities.

8.2 Phase Information from 2-D Gabor filters

Daugman[18] extended the work of Gabor and Papoulis[82] and showed that in the image domain Gabor filters also minimise the uncertainty between the signal and spatial-frequency, but that the new uncertainty inequality becomes:

$$\Delta y \Delta x \Delta u \Delta v \geq \frac{1}{16\pi^2} \tag{8.1}$$

where similar to the 1-D case, $\Delta y, \Delta x$ and $\Delta u, \Delta v$ refer to the duration of the signals in the space and frequency domains respectfully.

The reason for using 2-D filters is clear. Consider an image function:

$$f(\vec{r}) = \sin(2\pi\vec{u}.\vec{r}) \tag{8.2}$$

Such an image function will have an associated spatial frequency and preferred direction. Consider an identical signal, which is mirrored about the axis we are examining. In 1-D analysis, both signals will respond with an identical energy and phase response the phase differences will also be similar in a stereo pair but the signals will be markedly different on the grounds of orientation. Clearly, this is a potential source of false matching. For this reason, it is essential to apply 2-D filters with an orientation preference to the correspondence problem. In the same manner as the 1-D filter consider:

$$\begin{aligned} I_e(x,y,u_i,v_i) = {} & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma_x\sigma_y)^{\frac{1}{2}}} \exp[\frac{-(x-\tau)^2}{4\sigma_x^2} - \frac{-(y-\xi)^2}{4\sigma_y^2}] \\ & \cos[2\pi u_i(x-\tau) + 2\pi v_i(y-\xi)]F(x,y)d\tau d\xi \end{aligned} \tag{8.3}$$

$$\begin{aligned} I_o(x,y,u_i,v_i) = {} & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma_x\sigma_y)^{\frac{1}{2}}} \exp[\frac{-(x-\tau)^2}{4\sigma_x^2} - \frac{-(y-\xi)^2}{4\sigma_y^2}] \\ & \sin[2\pi u_i(x-\tau) + 2\pi v_i(y-\xi)]F(x,y)d\tau d\xi \end{aligned} \tag{8.4}$$

The convolutions in the shifted image may be computed and placed in a similar form as the 1-D case with:

$$F(x,y;u_o,v_o) = \cos[2\pi(u_o x + v_o y)] \tag{8.5}$$

and

$$F(x,y;\phi_x,\phi_y,u_o,v_o) = \cos[2\pi(u_o x + v_o y) + \phi_x + \phi_y] \tag{8.6}$$

It is important to notice that ϕ_x and ϕ_y represent unrotated phase differences corresponding to vertical and horizontal disparities in a binocular image. By proceeding in an identical manner, it can be shown that the rotation matrix of equation (4.25) is preserved with:

$$\left[\begin{array}{c} I_{s\phi}(x,y,u_i,v_i;u_o,v_o) \\ \frac{I_{a\phi}(x,y,u_i,v_i;u_o,v_o)}{\zeta} \end{array}\right] = \left[\begin{array}{cc} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi, \end{array}\right] \left[\begin{array}{c} I_s(x,y,u_i,v_i;u_o,v_o) \\ \frac{I_a(x,y,u_i,v_i;u_o,v_o)}{\zeta} \end{array}\right] \tag{8.7}$$

and

$$\frac{I_a(x,y,u_i,v_i;u_o,v_o)}{I_s(x,y,u_i,v_i;u_o,v_o)} = \frac{1}{\zeta}\tan[2\pi(u_o x + v_o y)] \tag{8.8}$$

Where we use the same notation as that used in chapter 4. Interestingly, the method of phase differences is now directional, which is a feature that we will use to obtain both horizontal and vertical components of disparity. The new 2-D correction factor for the rotation matrix now becomes:

$$\zeta = \tanh(8\pi^2(u_i u_o \sigma_x^2 + v_i v_o \sigma_y^2)) \tag{8.9}$$

with the phase rotation difference equal to:

$$\Phi(x,y,u_i,v_i;u_o,v_o) = \sqrt{\phi_x^2(x,y,u_i;u_o) + \phi_y^2(x,y,v_i;u_o)} \tag{8.10}$$

The new linear phase condition for a 2-D Gabor filter becomes:

$$\frac{u_i^2}{\sigma_u^2} + \frac{v_i^2}{\sigma_v^2} = 16 \tag{8.11}$$

The condition for linear phase now becomes an ellipse of minimum area.

Because the new rotation element ϕ is dependent on the Pythagorian combination of vertical and horizontal phase differences then a filter's phase difference may be derived from an infinite combination of vertical and horizontal disparities that lie on a circle of constant magnitude.

A similar occurrence is well known in motion analysis as the aperture problem but has not previously been linked to stereopsis[53].

8.3 The Motion and Stereoscopic Aperture Problem

The evidence from the geometry of retinal images supports the notion that objects not aligned on either the vertical or horizontal horopter will experience both a vertical and horizontal disparity. If we consider a sinusoidal grating, then, any physical displacement corresponding to a phase difference can best be resolved as a shift parallel to the grating's direction because of orientational selectivity. Phase differences parallel to the grating's orientation will yield no phase change.

combination of vertical and horizontal disparities (assuming the epipolar constraint). The solution to the aperture problem may be directly extracted from motion analysis and predicts: *if both horizontal and vertical disparities are required for absolute depth perception and binocular fusion, then each binocular image pair must contain features with similar frequency content and at least two orientations of dominant energy response.*

To extract this information, it is necessary to apply 2-D Gabor filters in bands of constant frequency but different orientation preferences in at least two quadrants in the Fourier plane. From the work of Longuet-Higgins, we observe that a stereoscopic image signal that does not satisfy this criterion may only provide relative as opposed to precise depth cues. The calculation of disparity may be obtained by taking a weighted least-squares fit from a band of filter's energy responses, and phase differences applied in a circle at the same pixel location i.e:

$$\mathbf{E}\ \mathbf{\Phi D} = \mathbf{E}\ \mathbf{d} \tag{8.12}$$

Where $\mathbf{E}$ is the MxM diagonal matrix of the energy responses from the $i^{th}(i = 1 to M)$ filter in a circle whose leading elements are formed from:

$$E_{ii} = \sqrt{E_l(\theta_i, u_i, v_i; x, y) E_r(\theta_i, u_i, v_i, ; x, y)} \tag{8.13}$$

$\Phi = [\cos\phi_i, \sin\phi_i]$ is the Mx2 matrix of directional orientation, and d represents the vector of measured disparity at all M orientations. $E_l(\theta_i, u_i, v_i; x, y)$ and $E_r(\theta_i, u_i, v_i; x, y)$ represent the pre-envelope of the i^{th} filter in both image domains, which follows from equation (7.5). $D = [D_x, D_y]^T$ represents the disparity (in pixels) estimated at the image point of interest in both image pairs from the filters at the ith orientation.

The above equation may be solved by numerical methods related to over determined sets of equations. In particular, the pseudo-inverse method gives us:

$$\mathbf{D} = (\mathbf{\Phi}^{\mathbf{T}}\mathbf{P}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathbf{T}}\mathbf{E}\mathbf{d} \tag{8.14}$$

from which we may estimate disparity in both vertical and horizontal directions. This equation holds several drawbacks. In particular, the equation has no weighting corresponding to the epipolar constraint. Indeed, should the dominant energy response from the band of filters lie along the vertical axis, then there will be a bias towards zero horizontal disparity, since vertical disparities are negligible for the finite visual angle of current camera hardware[83].

Equation (8.13), has some interesting properties. It has the effect of weighting down the energy responses from filters with similar orientation preferences but markedly different energy responses.

8.3.1 Methods

We will now consider some brief experimental work in the estimation of disparity using orientationally selective filters.

Experiment 1

We consider the error in horizontal disparity measurement for a horizontally oriented, two dimensional sinusoidal grating in both images, for a fixed horizontal shift between the image pairs as a function of grating orientation (both images retained the same orientation). The disparity estimates for a horizontally oriented sinusoidal grating that is shifted horizontally is also shown to indicate the phase differences obtained from orientationally selective filters as a function of orientation at a single point in the image.

Experiment 2

We consider the estimation of disparity from two superimposed sinusoidal gratings separated by a different internal angle. Both image gratings retained the same modulus of spatial frequency but were orientated in different directions. Here we are primarily interested in the estimation of disparity as the internal angle that separates two gratings is altered. We also indicate that both components of vertical and horizontal disparity can be obtained by considering an ideal "L" junction, which is displaced by both vertical and horizontal components of disparity.

Experiment 3

We now consider the estimation of horizontal disparity as a function of image position for the case of a horizontally oriented grating with a horizontal shift, and the same grating rotated by $\pi/4$ radians with a similar disparity.

8.3.2 Results

Experiment 1

Table 8.3.2 indicates the problems that occur with respect to the stereoscopic aperture problem. Here the correct estimate of horizontal disparity is only available when the grating is horizontal. A little examination of table 8.3.2 will reveal that the error in interpreting a horizontal disparity estimate can be described by a cosine function. Figure 8.2 highlights the phase differences as a function of orientation with respect to a horizontal disparity introduced into a horizontally orientated pair of gratings.

Experiment 2

Table 8.2 shows the results of this experiment. Here, the best disparity estimates occur when gratings are separated by $\frac{\pi}{2}$ radians. This may well be considered as the intersection of two orthogonal edges in a real image. To disambiguate a stereoscopic aperture problem, it is essential for an image to contain least two spatial frequencies with orientation preferences that are dissimilar. That the best stimulus for solving the aperture problem occurs at image corners is intuitive, because a single 2-D filter will have difficulty in disambiguating the contribution from two sinusoidal functions whose orientation and spatial frequency characteristics lie within the bandwidth of a single filter.

Corner detection is clearly important within stereopsis as indeed it is in motion since the best resolution of orientational differences between local signals arises at corners. Several authors have also stressed the importance of image features with high curvature in motion and stereopsis (e.g [12]). During scene analysis, corner intersection often occurs at areas of depth discontinuity and therefore is of considerable interest as an isolated image feature.

In figure 8.3 we show the phase differences from Gabor filters in both images as a function of orientation, taken from a pair of orthogonal lines displaced horizontally by 5 pixels and vertically by 2.5 pixels. The filters were located exactly on the corner of one image and disparity measurements were taken accordingly. The overall measures of disparity error in this case was 0.3 pixels in the x-axis and 0.15 pixels in the y-axis. From the energy response in figure 8.4, we can observe that the two energy maxima are unequal. This has occurred because the filters in the second image lie an unequal distance from each pair of lines in the stereo image.

Orientation (rads)	D_h	D_v
$-3\pi/8$	0.45	-1.07
$-\pi/4$	1.52	-1.52
$-\pi/8$	2.60	-1.07
0	3.05	$5.6E^{-7}$
$\pi/8$	2.60	1.07
$\pi/4$	1.52	1.52
$3\pi/8$	0.45	1.07
$\pi/2$	0	0

Table 8.1: *Disparity estimates for a single grating and 8 orientationally selective filter pairs tuned to 1/8 cpp. Sine grating varied in orientation. Horizontal disparity of 3 pixels introduced into the stereo pair.*

Experiment 3

In this experiment, we present the results for disparity estimation, as a function of image position for a pair of horizontal sinusoidal gratings, one of which is a shifted version of the other. The actual disparity introduced was 3 pixels, which has been recovered. If we now orientate our image pairs to $\frac{\pi}{4}$ and repeat the experiment, we observe that the estimate for disparity has resolved into the minimum path required to translate both gratings into correspondence, i.e both vertical and horizontal disparities are obtained. If we consider a disparity of -3 pixels introduced into a pair of gratings both orientated at $\pi/4$ then we would expect to measure a horizontal disparity of $-3\cos\pi/4 = -2.12$ because of the aperture problem. This is the result that we obtain.

8.3.3 Conclusions

For the simple sinusoidal grating, considered in figure 8.2, disparity estimates have behaved similar to a step, whose magnitude corresponds to the disparity estimate. This result is easy to interpret since a sinusoidal function transforms to an impulse in the Fourier plane. It follows from equation (8.8) that each filter can only respond to the local phase and frequency component of the grating. Therefore, disparity estimates from phase differences are a constant in magnitude. Orthogonal to the grating, however, the phase non-linearity term ζ prevents

Orientation difference (rads)	D_h	D_v
0	1.52	1.52
$\pi/8$	1.90	0.71
$\pi/4$	2.57	0.18
$3\pi/8$	2.84	0.03
$\pi/2$	2.96	$6.7\ E^{-3}$

Table 8.2: *Disparity estimates from a pair of sine plaids with orientation differences (orientation is defined in the frequency domain). One grating was maintained at an orientation of $\pi/4$ radians. Both gratings and 8 orientationally filter pairs tuned to 1/8 cpp. Horizontal disparity of 3 pixels introduced into the stereo pair.*

a correct interpretation of phase differences. This is because:

$$\tanh^{-1}(\zeta) = 8\pi^2 \begin{bmatrix} u_o \\ v_o \end{bmatrix} \cdot \begin{bmatrix} u_g\sigma_x^2 \\ v_g\sigma_y^2 \end{bmatrix} \tag{8.15}$$

which varies as an inner product between the orientation of the filter and grating.

The displaced pair of vertical lines indicate a different phase difference response profile as a function of orientation in comparison to the displaced sinusoid. In this case the difference can be described by a cosine function whose magnitude corresponds to the disparity. The displaced "L" junction indicates the presence of higher harmonics within the disparity estimates obtained from a circle of filters because of the introduction of both horizontal and vertical disparities.

These results can be interpreted from our work on corner and edge extraction in chapter 7 . Because the Fourier transform of a bar is a constant 2-D linear impulse function orthogonal to the orientation of the bar in the spatial domain, we know from our feature based work, that each filter responds to the frequency component whose distance is shortest from the center frequency of the applied filter. From Chapter 7, the preferred frequency that each filter responds to is $|\vec{r}_i - \vec{n}_\perp . \vec{r}_i|$. For filters applied in a circle the preferred image signal frequency varies as a cosine function. Therefore, for a translated pair of image bar functions, our disparity estimation:

$$D_{est} = \frac{\phi_d}{\omega_g} \tag{8.16}$$

contains an error similar to chapter 4 because of the differences in edge spatial frequency

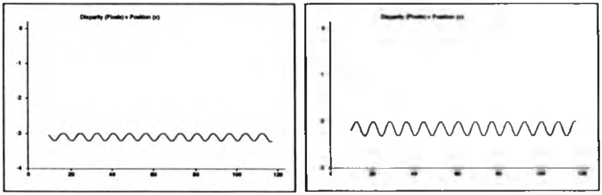


Figure 8.1: *(a) Horizontal disparities estimated from 2-D filters applied in a circle over a single raster line. Filters tuned to the frequency of the grating at 1/8 cpp. Disparity introduced at -3 pixels. Orientation calculated at 0 radians with confidence 88.0%. Error in estimation calculated as 3.3%. There was no component of vertical disparity measured with this stimulus. (b) Horizontal shift of -3 pixels introduced into a stereo pair of gratings orientated at* $\pi/4$*radians. Confidence measure 88.8% and orientation measured at 1.5701 radians.*

sensitivity for filters applied in a circle. In this special case:

$$\omega_{di} = \omega_g \cos(\theta_i - \theta_n) \tag{8.17}$$

where $\theta_i - \theta_n$ represents the angle difference between the normal to the bar (θ_n) and the orientation of the ith filter (θ_i), and ω_{di} represents the spatial frequency that each filter is sensitive to. In this case, we have the unusual situation, where the displacement (d) is constant, but the image frequency to which each filter is sensitive to varies as a cosine function (eqn. 8.17) as we rotate our applied filter pairs. To maintain, the same disparity displacement, the phase difference must correspondingly decrease. Indeed, the filter which lies parallel to the tangent of the bar has an indeterminate disparity measurement because the directional instantaneous frequency is zero (i.e d.c). If we were to assume an epipolar constraint, then an error in the measurement of horizontal disparity has been created by the nature of applying orientationally selective filters.

Indeed, if one observes a RDS formed from the two primary colours red and green and superimposed, then by rotating the angle of view, we observe that stereopsis is rapidly destroyed as the viewing angle increases past $\frac{\pi}{4}$ radians, thereafter the stereoscopic image

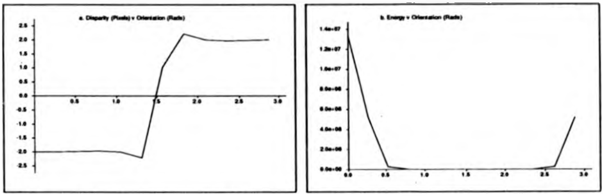


Figure 8.2: *Pair of horizontally oriented sinusoidal gratings (1/8 cpp). Disparity of two pixels induced into one stimulus. Filter center frequency also 1/8 cpp. (a) Disparity estimates from a band of orientated filters located at a single point. (b) Energy response taken from the product of pre-envelopes from left and right image pairs.*

pairs are in rivalry. The destruction of the stereoscopic signal in this case appears to be related to a cosine function of orientation. Thus it would appear that the human visual system is indeed insensitive to vertical translational disparity and that the perception of stereoscopic depth varies cosinusoidally with the orientation of the disparity signal. Such a percept would indeed be consistent with the extraction of a horizontal component of disparity from the mechanisms that we have proposed.

With motion analysis, a body may translate in any arbitrary path across the image plane. Therefore, to minimise the error in the computation, one would expect the perception of a translating grating to appear in the direction of the grating. However, stereoscopic translations are predominantly horizontal owing to the displacement of the eyes. This poses a dilemma that questions the application of orientationally selective filters towards the computation of disparity. Possibly stereoscopic systems could use a system equivalent to equation (8.14). However, such a system would incorrectly compute disparity at any other than a vertically orientated edge unless iteration methods were also used. Indeed, such computations are unnecessary when viewing the geometry of stereoscopic space. Thus, the question arises towards the need for the calculation of vertical disparities. Traditionally, the main evidence for vertical disparities lies with the *Induced effect* first introduced by Ogle[79]. Rogers and Koenderink[87] refuted these findings as we do here. Our argument, however, is more simple.

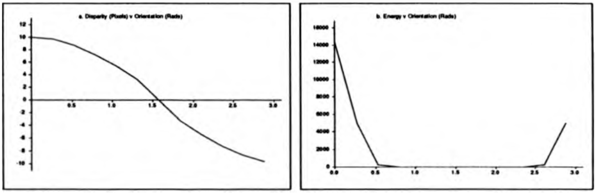


Figure 8.3: *Disparity estimates from an ideal vertically oriented bar. Disparity of 10 pixels introduced in the horizontal axis. 12 Filters applied in a circle center frequency of 1/22 cpp and aspect ratio 1:1. Disparity estimates obtained as 10.3 pixels (x-axis) and 0.000065 pixels (y-axis). (a) Disparity estimates from a circle of filters applied to the same retinal location. (b) Energy responses.*

We argue that if vertical disparities are to be obtained then we require disparity computation from filters applied in a circle to resolve both orthogonal directions. However, by doing so, we immediately introduce an aperture problem, whose effects would seriously affect stereoscopic computation. Therefore we might propose that the aperture problem is not generally considered in biological stereopsis, at the expense of vertical disparity estimates and possibly absolute depth computation during correspondence.

Alternatively, we could argue that the visual system only applies equation (8.14) at corners, or in the presence of orthogonal signals. Otherwise it resolves disparity into the horizontal component. The motivation for this type of calculation may be to resolve the local vertical displacements owing to sensor misalignment, a feature that has been recorded by physiological studies[70]. Indeed, the process of applying 2-D filters onto an image plane serves the purpose of permitting a certain degree of vertical disparity because of the spatial extent of the 2-D filters.

In view of the severe criticisms of equation (8.14), we can propose a justification for orientation selection, through all scales of filtering. We can add to the final disparity estimate an additional weighting that corresponds to the orientational differences between the left and right image pairs. There is also an additional advantage in the application of orientationally

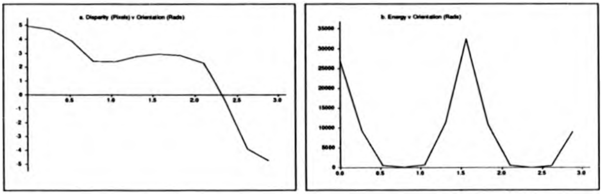


Figure 8.4: *An ideal corner formed from two orthogonal linear spatial impulse functions (bars). 12 Filters tuned to 1/22 cpp with aspect ratio 1:1. Initial disparity of 5 x-axis and 3 y-axis pixels input. Disparity estimates obtained as 4.77 pixels x-axis, and 2.88 pixels (y-axis). (a) Disparity estimates from a band of filters applied in a circle at the same point. (b) Energy response*

of orientationally selective filters that relates to the response of the filter as a function of orientation. A filter that is not orientated in the direction of the image signal will respond with a small energy response. As we have already shown in chapter 5, such a response will be highly noise sensitive since for small increments of added noise, there is approximately a linear relationship of phase error with respect to added quadrature noise. This follows from equation (5.3) with the maximum error (ϕ_{max}) in quadrature phase given by:

$$\phi_{max} \approx \frac{|\mathbf{n}_r(x,y,\omega,\sigma)|}{|\mathbf{z}_r(x,y,\omega,\sigma,\theta)|} \tag{8.18}$$

here we assume the image contains added isotropic noise ($|\mathbf{n}_r(x,y,\omega,\sigma)|$). If the magnitude of the pre-envelope ($|\mathbf{z}_r(x,y,\omega,\sigma,\theta)|$) is dependent on the filter orientation, then the filter with the least energy response will be most sensitive to added isotropic noise. Therefore it is essential to evenly sample the 2-D Fourier plane with orientionally selective filters.

In view of the narrow orientation bandwidths for stereoscopic correspondence that were found by both Blakemore and Braddick[11], we propose using the square Willsky weighting factor (η):

$$\eta \;=\; \frac{1}{4}(1+\cos 2\theta_d)^2 \tag{8.19}$$

where θ_d corresponds to the difference in orientation between the corresponding filter pairs in both the left and right image to constrain stereo matching. This term will weight down the correspondence of the phase locking process from image features that differ markedly in orientation differences. Our disparity interpretation vector becomes:

$$\mathbf{D} = [(\mathbf{\Phi^T E \Phi})^{-1} \mathbf{\Phi^T E d}]\eta \tag{8.20}$$

Where η is obtained from the orientational differences between quadrature pairs of filters from equations (7.10) and (8.19) respectively.

Care must be taken in the application of the orientation differences between our image pairs. This is because orientation is measured as a local image property, and we would not wish to prevent phase locking from a 2-D feature which is horizontally displaced, but also exhibits local orientation differences because of the location of the respective left and right image filter pairs relative to the feature. Consider phase locking a 2-D feature such as a box, with one band of filters located close to the vertical edge and the corresponding pair of filters in the stereo pair located close to the horizontal edge owing to disparities. Here we would erroneously prevent phase locking by the differences in local orientation. In dealing with this problem, we suggest iterating the phase locking procedure twice, and on the second pass use the orientation differences as a weighting to prevent false matching.

8.4 Method

Here we present the algorithm that was used to estimate disparity using 2-D Gabor filters. The basic details of the algorithm are explained in chapter 6. The algorithm functions with a simulation of a Phase-locked loop with a single estimate of horizontal disparity taken at each pixel of a compressed image. We therefore have assumed an epipolar constraint. As previously, the algorithm works from coarse to fine scales of filtering. For the remainder of this chapter, we will discuss alternative methods for disparity measurement. The basic structure of the Pseudo-code will not change, and will therefore only be presented once.

Pseudo-code

```
Define_Convolve()
For i = all_scales
```

```
For j = Number_of_filters_at_each_scale
For x,y = every_pixel_position_at_each_scale
For k = Every_orientation
        {
/* Create five dimensional arrays containing the modulus-Argument representation
of both images convolved with a quadrature filter at each pixel, scale and orientation. */
        $\phi_l[x,y][i][j][k] = Arg[I_l(x,y,i) * \Psi(\sigma_j,\theta_k,x,y)]$;
        $\phi_r[x,y][i][j][k] = Arg[I_r(x,y,i) * \Psi(\sigma_j,\theta_k,x,y)]$;
        $E_l[x,y][i][j][k] = |I_l(x,y,i) * \Psi(\sigma_j,\theta_k,x,y)|$;
        $E_r[x,y][i][j][k] = |I_r(x,y,i) * \Psi(\sigma_j,\theta_k,x,y)|$;
        }
Begin_program
Read_image();
Subtract_d.c_from_image();
$D_b[] = 0$;
Compress_image_with_Compact_Pyramid($I_l(x,y,i), I_r(x,y,i)$); /* Returns compressed data */
Convolve();
For i = first_scale to last_scale step 1
{
        For j = first_in_each_scale to last_in_each_scale step 1
            {
            For x,y = first_pixel to last_pixel step 1
                {
                $d = $(int)$D_b[x,y]$; /* integer disparity estimate */

/* Pass $E_r[x,y][i][j][], E_l[x-d,y][i][j][], \phi_r[x,y][i][j][], \phi_l[x-d,y][i][j][]$ */
/* to a proceedure that returns the horizontal component of disparity ($D_{dx}$) only */
/* Disparity measured using the center frequency of the filter */
                Solve_weighted_fit_disparity_estimation($D_{dx}$);
                $D_b[x,y] += D_{dx}$;
                }
Smooth_disparity_buffer_with_truncated_parabola();
```

```
            }
Expand_disparity_buffer(); /* Inverse of compress */
$D_b[] = D_b[] * 2$ ; /* Mutiply disparity estimates by 2 */
}
Print_out $D_b[]$;
End_program
```

8.4.1 Discussion and Results

We present our first results to the 4 test raster lines that we have used in chapter 4 (fig.8.5). Results have both increased in stability and also improved in accuracy by this method of filtering. This can be seen by comparing the data presented in Appendix A. This is not the case for figure 8.5a. Here we notice that we have incurred an anomalous depth estimate around pixel position 80. Examination subsequently revealed that this was owing to aliasing at the lowest spatial frequencies. The remainder of the results, however, are significantly improved from the 1-D study. In particular, we have not observed any aliasing in figure 8.5.d, which was a problem in 1-D for this particular sequence. We point out, that no vergence mechanism was incorporated into this particular algorithm, and therefore we would expect to improve on our results significantly by adding this feature to our algorithm. We have not included a recursive filter operation or a Willsky phase damping term because we have not found these additional damping factors essential, at least for these stereo pairs. This is partly owing to the weighted fit that we can now use to improve the measurement of disparity. We stress that resolving the horizontal component of disparity implicitly introduces an orientational damping which could be conceived as a similar weighting in 2-D as the Willsky error measure that was used in chapter 4 and proposed by Wilson and Knutsson[100]. Since, we have already improved our results on this basis alone, we immediately present a depth response from our control pair of images shown in figure B.1. This is shown in figure 8.6, where we have also shown the disparity estimates through the scales of resolution.

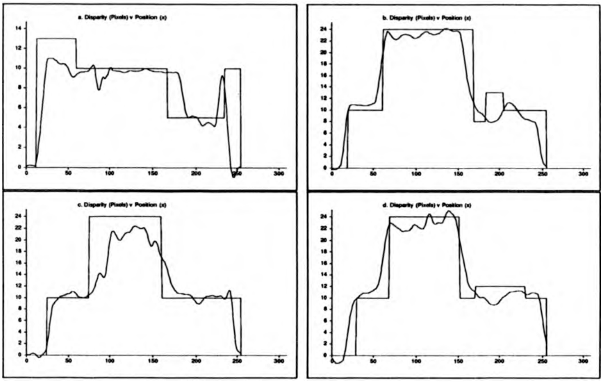


Figure 8.5: *Disparity estimates obtained by applying equation (8.20) to (a) Figure B.3, r=0.88;* $\bar{x} = -0.51$; $\sigma_n = 0.20$. *(b) Figure B.2, r=0.96;* $\bar{x} = -0.99$; $\sigma_n = 0.23$. *(c) Figure B.4, r=0.93;* $\bar{x} = -1.87$; $\sigma_n = 0.34$. *(d) Figure B.5, r=0.96;* $\bar{x} = 0.17$; $\sigma_n = 0.20$. *8 orientationally selective filters were used with aspect ratio 1:1 and tuned to 1/8 cpp. Images compressed to 16x16 at the lowest resolution.*

8.5 Biological plausibility

Perhaps it is advisable to turn to biology as a reference guide for possible solutions to the stereoscopic aperture problem. Maske et al[70], studied the response patterns from orientationally selective cells in the striate cortex of the cat to simple bar stimuli. They were interested in investigating the claims by Bishop and Pettigrew[6] that only cells with preferred stimulus orientation near to the vertical can make significant horizontal disparity interpretations. They showed that cells that were sufficiently end-stopped can make precise horizontal disparity discriminations, independent of the optimal stimulus orientation. They also showed that end-free cells were only *effective* for the measurement of horizontal disparity providing

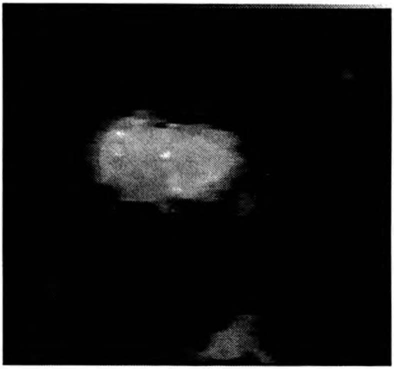
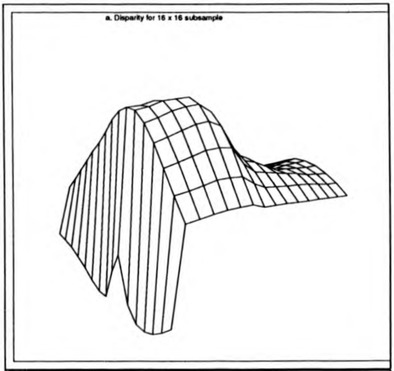
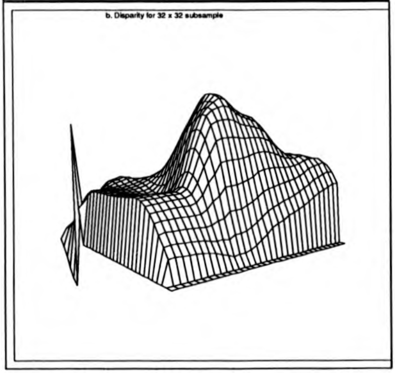


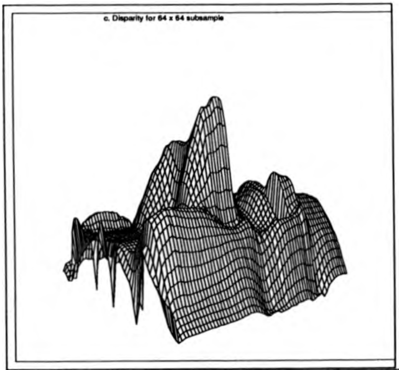
Figure 8.6: *Intensity disparity image produced from figure B.1. 8 orientationally selective filters with aspect ratio 1:1 were used tuned to 1/8 cpp. Images compressed to 16x16 at the lowest resolution.*



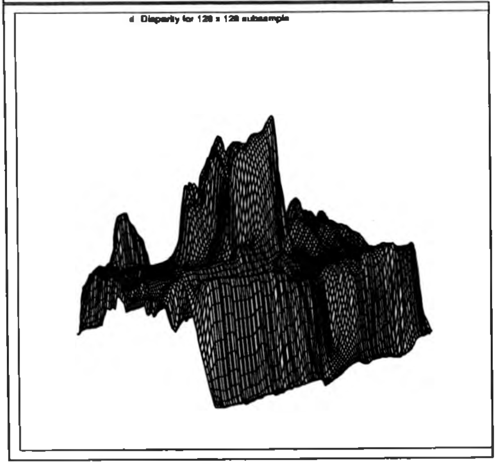
a. Disparity for 16 x 16 subsample



b. Disparity for 32 x 32 subsample



c. Disparity for 64 x 64 subsample



d. Disparity for 128 x 128 subsample

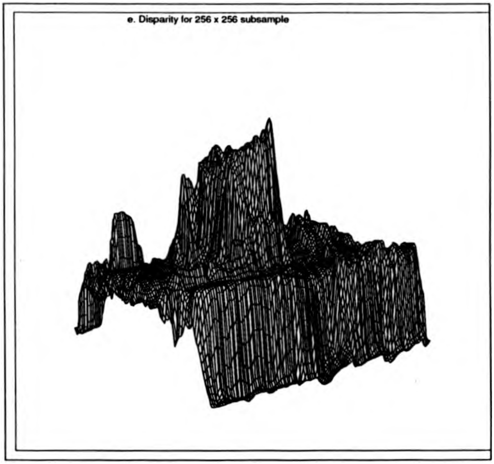


Figure 8.7: *Surface map showing disparity from figure 8.6 through the resolutions of subsampling with the Compact Pyramid Code.*

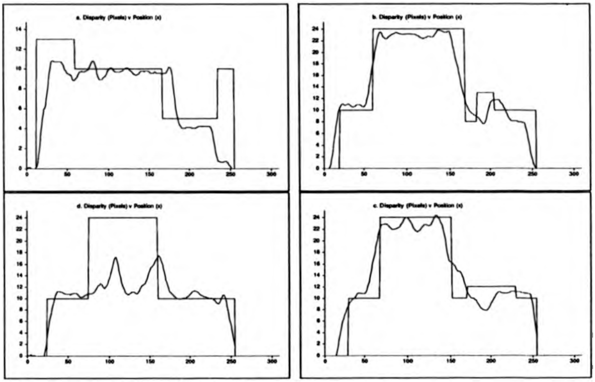


Figure 8.8: *Disparity estimates obtained by applying equation (8.20) to (a) Figure B.3, r=0.86;* $\bar{x} = -1.0$; $\sigma_n = 0.22$. *(b) Figure B.2, r=0.96;* $\bar{x} = -0.97$; $\sigma_n = 0.23$. *(c) Figure B.4, r=0.87;* $\bar{x} = -3.47$; $\sigma_n = 0.56$. *(d) Figure B.5, r=0.96;* $\bar{x} = -0.05$; $\sigma_n = 0.20$. *12 orientationally selective filters were used with aspect ratio 1:1 and tuned to 0.25 cpp. Images compressed to 16x16 at the lowest resolution.*

their preferred orientations were near to the vertical. We have also obtained similar results which we interpret as a direct consequence of the stereoscopic aperture problem. We have already shown that the receptive field profile of our corner detector has a similar form to the end-stopped cell. We suggest there is no justification by applying the receptive field profile of the end-stopped cell as a single operator to obtain translational disparity. This is because we lose the orientational phase information from the orthogonal filter pairs. Indeed, this is the very notion that we are relying on from equation (8.14) to obtain unambiguous measures for both horizontal and vertical disparity.

It appears that biological systems have also recognised the stereoscopic aperture problem and has dealt with the problem in a logical manner: In the presence of regions which

Figure 8.9: *Intensity disparity image produced from figure B.1. 12 orientationally selective filters with aspect ratio 1:1 were used tuned to 0.25 cpp. Images compressed to 16x16 at the lowest resolution.*

exhibit orthogonal signals (i.e corners) the stereoscopic aperture problem may be resolved into both horizontal and vertical components of disparity. This naturally follows from the motion aperture problem. However, unlike the motion aperture problem we know from the epipolar geometry constraint that disparities expected in the direction of the horizontal axis. Since we have a confidence measure for both corners and edges, we can therefore propose to disambiguate this situation. We suggest:

$$D_h \;=\; C_{edge} d_{he} \;+\; (1 - C_{edge}) d_{hc} \qquad (8.21)$$

where D_h is the measure for horizontal disparity at a given scale and position in the image function, and C_{edge} is the probabilistic measure for edgeness and d_{he} ,d_{hc} represent the horizontal estimates for disparity in the presence and absence of edge information respectively. We choose this form, because we have seen that edge information gives rise to a particularly large energy response. Since the aperture problem is consequence of directional image energy we suggest it is sufficient to detect the presence and absence of edge information only. This holds the advantage of not requiring a description to any arbitrary response pattern (textures) that may occur from the circle of filters. The vertical component of disparity is then similarly obtained from:

$$D_v = (1 - C_{edge})d_{vc} \tag{8.22}$$

where the subscript 'v' denotes a vertical component[55].

8.6 Dealing with the stereoscopic aperture problem

Having spent some time discussing the issues involved with stereopsis and the aperture problem, we require a solution. First, we could simply restrict our analysis to the 1-D case, by the application of a single 2-D filter with horizontal spatial frequency preferences. This immediately removes all the issues that we have discussed. We would still expect considerable improvements over the results presented in chapter 4, since we would now be dealing with a filter whose spatial characteristics were two-dimensional. Indeed this is also the case which we observe in figures 8.12 and 8.13. Results have improved in stability by comparison to chapter 4, since we have iterated the phase locking procedure using only the center frequency of the filter to interpret disparity, the Willsky error measure to damp phase differences, and also applied our parabolic window to smooth the data. This is similar to the algorithm proposed by Wilson and Knutsson, however, but we have refrained from iterating more than once through each scale and still smoothed at each resolution to avoid local instability.

As an alternative consider applying filter pairs so that we restrict the orientations from which we extract horizontal disparities. Thus we restrict the application of filters with spatial frequency orientations between $\pm\frac{\pi}{4}$, which immediately permits a weighted estimate of disparity in the same manner prescribed before, but with the additional advantage that we resolve our estimated disparities by:

$$D_{hi} = \frac{d_i}{\cos\theta_i} \; ; \; \theta_i \neq \frac{n\pi}{2} \quad n = 1, 3, .. \tag{8.23}$$

Where $\cos(\theta_i)$ represents the orientation of the ith filter from which disparity is interpreted. We would then anticipate estimating the final disparity estimate in a least squares sense from:

$$D_h = \frac{\sum_{i=-\frac{\pi}{4}}^{\frac{\pi}{4}} E_i(\theta_i, v_i, u_i; x, y) D_{hi}}{\sum_{i=-\frac{\pi}{4}}^{\frac{\pi}{4}} E_i(\theta_i, v_i, u_i; x, y)} \tag{8.24}$$

which we would anticipate to assist in the phase locking process. The problem that we have introduced by applying orientationally selective filters, becomes clear by the division of the cosine of the angle, since we cannot define a horizontal disparity for filter pairs with vertically oriented spatial frequency preferences. This risks entering a pole into the above expression, which cannot be tolerated.

A further option would be to include a weighting function which approximated $\frac{1}{\cos(\theta)}$ for say the first $\pm\frac{\pi}{4}$ radians and then rapidly tends to zero at $\pi/2$ radians, i.e:

$$\eta = 1 \ + \ \theta_i \sin\theta_i \tag{8.25}$$

which approximates the property that we require. This would then enable us to extract the phase information from all our filters in a circle. Alternatively, we could apply a cosine weighting function to our energy responses i.e:

$$E_i(\theta_i, v_i, u_i; x, y) \ = \ E_i(\theta_i, v_i, u_i; x, y)\cos\theta_i \tag{8.26}$$

to reduce the bias for vertical disparity sensitivity. We choose our solution based upon the neurophysiological data. i.e we apply equation (8.14) to obtain our vertical and horizontal estimates of disparity, and we also extract an estimate of disparity horizontally based upon filters with orientation preferences of $\pi/4$ towards the horizontal, which are then resolved to interpret horizontal components of disparity from the least squares measure obtained from equation (8.24). This scheme holds the desirable property, that all the other oriented filters that are also applied in the circle, are in fact orthogonal to at least one filter whose disparity is resolved horizontally. We could then apply equation (8.21) to obtain our horizontal component of disparity, with the corresponding vertical estimate obtained from equation (8.22).

The horizontally resolved estimates for disparity interpretation are presented in Table 8.3 for the same stimulus as in Table 8.2. Improved accuracy may also be obtained from extracting the horizontally resolved disparity from equation (8.24) from a circular band of filters and excluding the filter with horizontal orientation preferences (Table 8.4), since equation (8.24) will still contain an error in disparity interpretation from a feature oriented at an angle less than $\frac{\pi}{4}$ radians. We presents results from this experiment in figure (8.10).

Orientation difference (rads)	D_h	D_v	P_e
0	2.77	0.09	0.95
$\pi/8$	2.56	0.09	0.88
$\pi/4$	2.75	0.05	0.68
$3\pi/8$	2.80	0.03	0.37
$\pi/2$	2.96	$6.7\ E^{-3}$	0.01

Table 8.3: *Horizontally resolved disparity estimates from a pair of sine plaids with orientation differences. One grating was maintained at an orientation of $\pi/4$ radians. Both gratings and 8 orientationally filter pairs tuned to 1/8 cpp. Horizontal disparity of 3 pixels introduced into the stereo pair. Far right column shows edge probability estimate given.*

Orientation (rads)	D_h	D_v
$-3\pi/8$	3.10	-0.07
$-\pi/4$	3.08	-0.09
$-\pi/8$	3.04	-0.06
0	3.05	$3.10E^{-8}$
$\pi/8$	3.04	-0.06
$\pi/4$	3.10	0.06
$3\pi/8$	3.04	0.09
$\pi/2$	0	0

Table 8.4: *Horizontally resolved disparity estimates for a single grating and 8 orientationally selective filter pairs tuned to 1/8 cpp. Only the horizontally oriented filter was prevented from contributing in this case. Sine grating varied in orientation. Horizontal disparity of 3 pixels introduced into the stereo pair.*

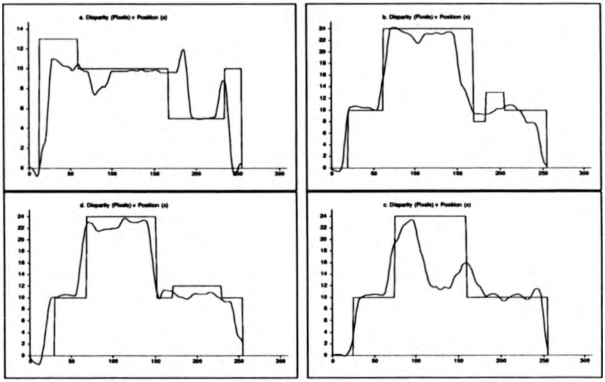


Figure 8.10: *Disparity estimates from applying 8 orientationally selective 2-D filters tuned to 1/8 cpp with aspect ratio 1:1 to (a) Figure B.3, r=0.86;* $\bar{x} = -0.42$; $\sigma_n = 0.21$. *(b) Figure B.2, r=0.96;* $\bar{x} = -1.60$; $\sigma_n = 0.29$. *(c) Figure B.5, r=0.95;* $\bar{x} = -0.58$; $\sigma_n = 0.22$. *(d) Figure B.4, r=0.89;* $\bar{x} = -2.11$; $\sigma_n = 0.41$. *Edges resolved horizontally.*

8.7 2-D Spatial derivatives of Phase

There are now two options that we might now consider by applying the spatial derivative of phase to interpret our measure of instantaneous frequency. In the first instance, we could consider applying our filter pairs arranged in a circular cluster. Therefore, instead of extracting information from a band of filters located at the same pixel location, we extract phase information from a band of filters which lie on a circle of radius (r) on the image domain. relative to a point of interest. This technique would enable us to extract the directional derivative of phase from filter pairs with similar orientation preferences, but separated by the spatial diameter of the circle that they lie on. Naturally, to implement this notion correctly and efficiently, we would require an image that has been sampled isotropically. The

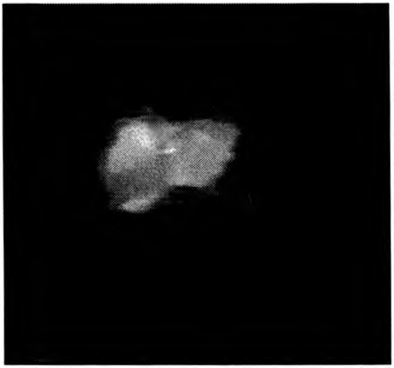


Figure 8.11: *Intensity depth image using the horizontally resolved component of disparity as displayed in figure 8.10.*

hexagonal lattice would provide a natural image basis for this notion because it holds these properties, and is presumably the reason why biological systems also use a similar sampling array on the retina of the eye. Unfortunately, conventional sampling arrays do not exhibit these isotropic properties. In view of this, we apply the 1st directional spatial derivatives of our Gabor functions to estimate the instantaneous frequency of our image field[56]. However, stationary phase presents considerable problems for disparity interpretation by this method. At an edge, instantaneous frequency varies as a Willsky error (fig. 8.16). This response is owing to two contributing factors. First, there exists a cosine weighting based upon the difference in orientation between filter and stimulus. Second, there lies an additional cosine weighting owing to the radial sensitivity of the filter in the frequency domain. This results in a Willsky error. Resolving horizontal and vertical components of disparity under these conditions is difficult. However, in this case we can propose a frequency domain model of:

$$\begin{bmatrix} \frac{df_x}{d\eta} \\ \frac{df_y}{d\eta} \end{bmatrix} = \frac{1}{2}\left(\mathrm{div}\omega_o\, I + \mathrm{def}\omega_o\, \mathbf{M} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{M}^{-1}\right) \tag{8.27}$$

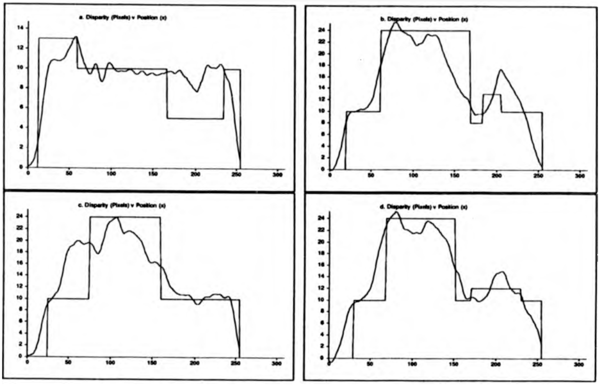


Figure 8.12: *Disparity estimates from applying a single 2-D filter tuned to 1/8 cpp with aspect ratio 1:1 to (a) Figure B.3, r=0.90;* $\bar{x}$ = 1.03; σ_n = 0.21. *(b) Figure B.2, r=0.94;* $\bar{x} = -0.95$; σ_n = 0.29. *(c) Figure B.4, r=0.91;* $\bar{x}$ = 0.31; σ_n = 0.29. *(d) Figure B.5, r=0.95;* $\bar{x}$ = 0.27; σ_n = 0.23.

where M refers to the principle axis of the deformation and η is a paramter. Unfortunately, using the directional instantaneous frequency in this case does not lend naturally to the direct interpretation of disparity. This can be seen from figure 8.16. Perpendicular to the orientation of the signal lies a d.c component which is indicated by the instantaneous frequency term passing through zero. It is not possible to continuously interpret disparity orthogonal to an edge or line using instantaneous frequency because the phase advance is stationary. We can also see from the assumptions made on corner detection, that the instantaneous frequency at a corner will oscillate as a function of orientation at twice the frequency of an edge. There is therefore no direct model of instantaneous frequency behaviour as a function of orientation.

Figure 8.13: *Intensity disparity image produced from figure B.1 using a single horizontally oriented 2-D filter as in figure 8.12.*

At an edge/bar we could apply a weighted least squares fit:

$$\mathbf{E}\mathbf{\Phi}\mathbf{F} = \mathbf{E}\mathbf{f} \tag{8.28}$$

where f refers to the measured instantaneous frequency vector $f_i(\theta_i)$ E refers to the product pre-envelope weighting used in equation (8.13), $\Phi = [\cos\theta_i, \sin\theta_i, 1]$ refers to the orientation of the ith filter with the added constant that requires evaluation (i.e the d.c term now present) with $F = [a_1, a_2, a_3]^T$ referring to the coefficients of the constant and oscillatory part of f that we wish to measure. Because of the difficulty in representing a suitable model, we have chosen a weighted mean instantaneous frequency to interpret disparity. For example in the left image we would have (f_{lav}):

$$f_{lav} = \frac{\sum_{i=0}^{M} E(\theta_i, u_i, v_i; x, y) f_i(\theta_i)}{\sum_{i=0}^{M} E(\theta_i, u_i, v_i; x, y)} \tag{8.29}$$

We have also applied the same thresholds as we have applied in the 1-D case, based upon the disparity gradient limit, and the upper and lower cut-off frequencies of our filter pairs. At present, we take a weighted average of spatial frequency as the threshold parameter.

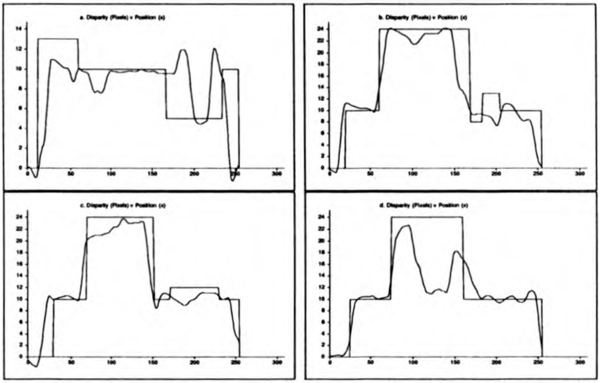


Figure 8.14: *Disparity estimates from applying 8 orientationally selective 2-D filters tuned to 1/8 cpp with aspect ratio 1:1 to (a) Figure B.3, r=0.82;* $\bar{x} = -0.98$; $\sigma_n = 0.23$. *(b) Figure B.2, r=0.95;* $\bar{x} = -1.5$; $\sigma_n = 0.30$. *(c) Figure B.5, r=0.95;* $\bar{x} = -0.88$; $\sigma_n = 0.24$. *(d) Figure B.4, r=0.91;* $\bar{x} = -2.4$; $\sigma_n = 0.40$. *The directional component of spatial frequency was weighted and used to interpret disparity differences. Thresholds were applied similar to those used in chapter 4.*

Some results from applying this technique are presented in figure 8.14. Where equation (8.12) is now modified to:

$$\mathbf{E}\,\mathbf{\Phi}\mathbf{D} = \mathbf{E}\,\phi_{\mathrm{d}} \tag{8.30}$$

where ϕ_{d} now represents a vector that contains the phase differences between quadrature filter pairs at similar orientations in the left and right images.

We can propose a better method of disparity measurement that is based upon the derivative of phase resolved in the horizontal direction. Here we extract phase from our circle of filters, and obtain the horizontal derivative of phase from the output response of each filter

Figure 8.15: *Intensity disparity image using the same parameters as figure 8.14.*

pair. i.e we form:

$$\frac{1}{2\pi}\frac{d\phi(x;\theta)}{dx} = f_{ih}(x) \tag{8.31}$$

where $f_{ih}(x)$ refers to the horizontally resolved frequency component extracted from the quadrature pair of filters whose orientation is given by θ. It is also possible to obtain the phase gradient of equation (8.31) by using quadrature filters and their derivatives with an isotropic envelope by a modification of equation (3.17):

$$\frac{d\phi(x,\theta)}{dx} = \frac{h(u_x;x,y)\dot{\tilde{h}}(u_i,v_i;x,y) - \tilde{h}(u_x;x,y)\dot{h}(u_i,v_i;x,y)}{E^2(u_x;x,y)} \tag{8.32}$$

with $h(u_x;x,y) + j\tilde{h}(u_x;x,y)$ as Hilbert Transform pairs with horizontal directional preferences, and $\dot{h}(u_i,v_i;x,y) + j\dot{\tilde{h}}(u_i,v_i;x,y)$ as Hilbert transform pair derivatives taken at an arbitrary orientation.

We therefore obtain the frequency for interpreting horizontal disparity by:

$$f_h = \frac{\sum_{i=0}^{M} E_{li}(\theta_i,v_i,u_i;x,y)f_{lih} + \sum_{i=0}^{M} E_{ri}(\theta_i,v_i,u_i;x,y)f_{rih}}{\sum_{i=0}^{M} E_{li}(\theta_i,v_i,u_i;x,y) + \sum_{i=0}^{M} E_{ri}(\theta_i,v_i,u_i;x,y)} \tag{8.33}$$

and obtain a least squares phase difference from equation (8.30).

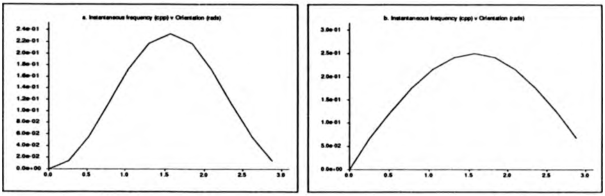


Figure 8.16: *Instantaneous frequency from 12 oriented filters and their directional derivatives with center frequency at 1/4 cpp. (a) Edge response showing a Willsky error in instantaneous frequency. (b) Sine grating (1/4 cpp) with vertical orientation (Fourier domain): only a cosine error exists in this case.*

8.8 Disparity Error in Two Dimensions

A final method of disparity measurement in two dimensions follows from the observations in chapter 4, that the method of phase differences reduces to a Newton-Raphson iteration. From equations (8.12) and (8.27), it is clear that we are assuming that the disparity can be represented by a vector, and instantaneous frequency can be modeled as a second order tensor. It is straight forward to realise that at a given scale of filtering the iteration of disparity measurements can be defined by:

$$\mathbf{D}^{(n+1)} = \mathbf{D}^{(n)} + [\mathbf{\Phi}_i']^{-1}\mathbf{\Phi}_d \tag{8.34}$$

where $\mathbf{D}^{(n)} = [d_x, d_y]^T$ refers to nth iteration of disparity using the horizontally and vertically resolved phase difference vector $\mathbf{\Phi_d}$ from equation (8.30) and the tensor $\mathbf{\Phi}_i'$ is:

$$\mathbf{\Phi}_i' = \left| \begin{array}{cc} \frac{\partial \phi_x}{\partial x} & \frac{\partial \phi_x}{\partial y} \\ \frac{\partial \phi_y}{\partial y} & \frac{\partial \phi_y}{\partial y} \end{array} \right| \tag{8.35}$$

where once again, we can apply the local instantaneous frequency from either image to interpret disparity. Thus by creating the partial derivatives of the instantaneous frequency field, it is possible to iterate into convergence. Similarly to the one dimensional case, the disparity error (ϵ) from the measurement of disparity from the center frequency of the filter after one iteration becomes:

$$\epsilon = (\mathbf{I} - \frac{1}{|\omega|}[\Phi_l']^{-1} \; + \; \ldots \;)\mathbf{D} \tag{8.36}$$

which retains both horizontal and vertical components. Here $|\omega|$ refers to the modulus of center frequency of the filter, and $\mathbf{D}$ the disparity.

8.9 Conclusion

Within the context of this thesis, there exist two main approaches which are based upon the methods used to interpret disparity:

- Instantaneous frequency
- Center or expected filter frequency

Both of these methods contain problems in their interpretation. In the first case, an improved description of the image signal is possible, and also statements regarding the convergence of the phase locking iteration. Unfortunately, thresholds are at present applied to this technique, which would be preferable to avoid. The technique of applying a directional component of instantaneous frequency to interpret disparity is also difficult because of stationary phase problems. It is, however, possible to estimate a mean horizontal component of instantaneous frequency by considering the spatial phase derivative taken in the horizontal direction from filters with arbitrary orientation. A weighted mean instantaneous frequency can then be obtained to measure disparity.

With one dimensional filtering, and using the center frequency of the filter to interpret disparity does increase the error in disparity measurement and also relies heavily on the signal characteristics. Convergence cannot yet be assured with this technique, although we have noted some considerable improvements in performance by applying recursive filtering, and additionally the Willsky error measure. The two dimensional extension does improve results considerably.

The greatest problems that have been found occur with the instability of a phase process in the presence of noise. This is particularly true at the higher spatial frequencies of analysis. The premise that surrounds a coarse to fine strategy assumes that the lowest spatial frequencies present in an image signal will not themselves alias. We have seen, that this premise is not entirely reliable and forms the basic weakness of the algorithms proposed. We have, however, reduced this possibility by the examination of the spatial phase gradient. Indeed, noise forms an additional constraint, which makes image interpretation particularly difficult. Fortunately, within the 2-D framework, we have been able to apply least squares functions which reduce the consequences of noisy signals. Which has enabled a considerable improvement to stereoscopic computation.

Chapter 9

DEFORMATION

9.1 Introduction

We will review some work on the computation of deformation under a stereoscopic transformation. This work has not yet been incorporated into an algorithm to obtain both translational and higher order properties of the disparity field. The methods discussed here, however, are relevant in the context of this thesis because they apply some techniques considered in chapter 6, towards the computation of differential invariants of the disparity field.

9.2 The geometry of the disparity field

Let us consider the translation of a non-rigid laminar surface ($\dot{\mathbf{R}}$), given by:

$$\dot{\mathbf{R}} = \mathbf{V} + \mathbf{\Omega} \mathrm{x} \mathbf{R} + \mathbf{F} \tag{9.1}$$

We have rotational ($\mathbf{\Omega}$x$\mathbf{R}$), translational ($\mathbf{V}$) and deformation ($\mathbf{F}$) components which can be used to model the possible transformations of an image plane. Generally, we should consider the six possible degrees of freedom that a solid surface can translate through and its representation on the image plane. It is more difficult, is to extract any combination of these degrees of freedom in representing an object stereoscopically. Koenderink and Van Doorn[49] considered these transformations in binocular space. They came to the conclusion that the local disparity vector field ($\vec{\chi}(\vec{r})$)can be represented by:

$$\vec{\chi}(\vec{r} + d\vec{r}; \vec{r_o}) = \vec{\chi}(\vec{r}; \vec{r_o}) + \frac{\partial \vec{\chi}}{\partial \vec{r}} d\vec{r} + B d\vec{r} + C d\vec{r} + \ldots \tag{9.2}$$

Where $\vec{r}$ represents the 3-D vector of a point in disparity space from the cyclopean eye. The first term describes a rigid translation, the second a rotation proportional to the curl. The third term represents the divergence of the image function, while the fourth term relates to a deformation. This decomposition is, however, well established within the context of continuum mechanics and dynamic equations[69]. We will therefore review this field with the purpose of re-interpreting the conclusions of Koenderink and Van Doorn.

9.2.1 Interpretation of the Local Disparity Field

From Koenderink and Van Doorn, we assume that the local image field is differentiable at least twice. In addition we assume that the image field possesses a singularity at the origin of our current frame of reference with:

$$\mathcal{J} = |\frac{\partial \chi^i}{\partial x^i}| \tag{9.3}$$

as the *Jacobian* of the transformation. We assume that it is possible to Taylor expand the local velocity or disparity field (there is no distinction between the time and stereoscopic sampling for a time invariant system) in the case where the transformations are autonomous. We arrive at equations with the form:

$$\dot{x}^i = \mathcal{J} x^i \tag{9.4}$$

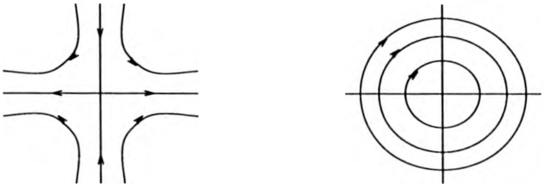
where the spatial gradient of the velocity vector ($v_i = \frac{dx_i}{dt}$) may be decomposed into symmetric and skew-symmetric parts according to:

$$\frac{\partial v_i}{\partial x_j} = \frac{1}{2}(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}) + \frac{1}{2}(\frac{\partial v_i}{\partial x_j} - \frac{\partial v_j}{\partial x_i}) = \mathbf{D} + \mathbf{V} \tag{9.5}$$

where D and V refer to the *rate of deformation* tensor and *vorticity* tensor respectively. The behaviour of the above system of equations are well understood within the regions of the linear approximation in the context of continuum mechanics[69] and *small deformation theory*. In particular, since we assume stereoscopic fusion has been obtained at the fixation point (singularity) or origin of our coordinate frame of reference, the local vector field is analogous to the *material derivative* of a continuum. The isoclines ($\frac{dx^i}{dx^j}$) under the linear approximation exhibit properties that are completely described by the eigenvalues of the matrix $\mathcal{J}$ shown above. We show three examples (fig. 9.1) which are particularly relevant:

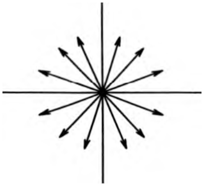
- Saddle point: Eigenvalues real and of opposite sign.
- Centre: Eigenvalues wholly imaginary.
- Focus: Eigenvalues real and equal sign.

Within this context, compression/expansion transformations correspond to wall surfaces (rotations about a vertical axis) that deforms the orientation of local line elements with path lines described by a saddle with principle axis in horizontal and vertical directions (relating to world coordinates). For points that lie along the principle axis of the deformation (i.e vertical and horizontal lines) there are only normal strains (*stretch ratio*). Lines not oriented along the principle axis exhibit rotation. If we assume the epipolar constraint, then



a) Saddle

b) Center



(c) Node (unstable)

Figure 9.1: *The local differential field of an image plane obtained from motion or stereoscopic transformations represented as isoclines (arrows). The interpretation assumes a singularity at the origin corresponding to stereoscopic fusion or indeterminate velocity with* $\det|\mathcal{J}| = 0$.

compression/expansion surfaces can be described by both components of deformation and divergence. For ground planes (rotations about a horizontal axis), line elements exhibit either vertical or horizontal shear. Shear may be decomposed from components pertaining to rotation and deformation, with the axis of deformation rotated by $\frac{\pi}{4}$ radians in relation to the deformation field owing to wall planes.

9.2.2 The Disparity Vector Field Model

If we consider equation (9.4) in terms of the linear approximation by a Taylor expansion on an image plane we arrive at:

$$\begin{bmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{9.6}$$

where in this case t is a parameter

The matrix:

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \tag{9.7}$$

can be re-written in a pseudo-canonical form as:

$$\frac{1}{2}\begin{bmatrix} d & 0 \\ 0 & d \end{bmatrix} + \frac{1}{2}\begin{bmatrix} w & 0 \\ 0 & -w \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 & c \\ c & 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 & e \\ -e & 0 \end{bmatrix} \tag{9.8}$$

where the decomposition in this case is found from:

$$d = a_1 + b_2 \ , \ w = a_1 - b_2 \ , \ c = a_2 + b_1 \ , \ e = b_1 - a_2 \tag{9.9}$$

and from left to right we have terms approximating to: DIV, $DEF^{0,\frac{\pi}{2}}$, $DEF^{\pm\frac{\pi}{4}}$ and ROT[1].

By restricting consideration to the changes in orientation of local line elements, we do not consider the contribution owing to the divergence of the image field. By changing the coordinate frame by the Jacobian:

$$\frac{\partial(x_1, x_2)}{\partial(r, \theta)}$$

we have:

$$\dot{\theta} = \frac{1}{2}[e - c\cos 2\theta + w\sin\theta] \tag{9.10}$$

and also:

$$\dot{r} = \frac{r}{2}[d - c\sin\theta + w\cos\theta] \tag{9.11}$$

---

[1] superscripts denote the deformation axes.

Notice that the deformation component $\theta = f(\theta)$ is a form of orientation modulation and analogous to frequency or phase modulation that we discussed in chapter 4. Rogers and Graham[88] showed that the magnitude of perceived slant for a surface oriented about a horizontal axis (ground plane) was twice that of an equivalent slant but rotated about a vertical axis (wall plane). They explained their results by suggesting that the human visual system uses the local differences in orientation between stereoscopic images to infer the magnitude and direction of slant. This is easily explained by equation (9.10). The disparity gradient owing to a wall plane is equivalent to an image compression which comprises of a divergence and deformation part. In contrast, a ground plane is equivalent to an image shear which comprises of rotation and deformation parts. In terms of the local changes in orientation, the shear transformation predicts the anisotropy because of the additional rotation term.

9.2.3 Ogle's Induced Effect

Ogle[79] introduced the psychophysical observation that a frontoparallel image field that is magnified along the vertical direction by a cylindrical lens appears rotated about a vertical axis. This interpetation is not difficult in terms of the differential invariants of the stereoscopic field, since the action of a single direction of expansion to an image function can be decomposed into equal parts of deformation and divergence[87]. Such a deformation by definition will also be an *irrotational field* since by definition the vorticity tensor (rotation) will on average vanish everywhere in the field.

9.3 Rotational Invariants

Central to the expansion of the local stereoscopic field, lies the concept of vergence, or fusion at a fixation point. We consider two techniques[57]. First, we obtain only the rotation component of the stereoscopic field using a method that requires a minimal number of filters to estimate orientation independent of quadrature phase. Second, we consider a technique that permits the computation of both rotation and deformation.

9.3.1 Rotation

We begin by considering a direct method for obtaining the rotation component about the fixation point as the differences in local orientation that is independent of quadrature phase. This can be seen as the extension to some of the ideas presented in chapter 7. This can be effectively obtained by using the interpolative properties of Gaussian derivatives and their Hilbert transforms. Freeman and Adelson[23] considered the second directional derivative of the Gaussian which unfortunately requires two additional filters to compute orientation. We use the natural tensor representation[47]:

$$\left| \begin{array}{cc} x.x & x.y \\ y.x & y.y \end{array} \right] \text{i.e} \left[\begin{array}{cc} l_x^2 + \hat{l}_x^2 & \hat{l}_{x.y}^2 + l_x l_y \\ \hat{l}_{y.x}^2 + l_x l_y & l_y^2 + \hat{l}_y^2 \end{array} \right] \tag{9.12}$$

from which the eigenvectors give:

$$\tan 2\theta_l = \frac{[(\hat{l}_{x.y}^2 - \hat{l}_{y.x}^2 + 2l_x l_y)]}{(l_x^2 + \hat{l}_x^2) - (l_y^2 + \hat{l}_y^2)} \tag{9.13}$$

where l_x refers to the operation $\frac{dG(x,y)}{dx} * I(x,y)$ with $G(x,y)$ as a Gaussian smoothing kernel, $I(x,y)$ an image function, and subscripts refer to the directional derivative. $\hat{l}_x$ refers to the Hilbert transform of the directional derivative of the Gaussian. It is well known that the orientational energy response of this kernel[16] with an isotropic envelope varies as a function of $\cos(\theta - \eta)$ where η refers to the orientation of the signal. The energy response of its Hilbert transform therefore varies as a function of $\sqrt{\frac{1}{2}(1 + \cos(2(\theta - \eta))}$. The real part includes a constant term which requires a further filter to estimate. Thus we note that we require the minimum of five filters to obtain a phase independent estimation of orientation (two odd-symmetric three even symmetric). In equation (9.12) we have applied six to avoid interpolation. Since orientation is energy based, we define orientation over $\theta \in [0, \pi]$. We also note in this case, that $\hat{l}_x^2$ may be replaced by $|l_{xx}|$ but with a loss in quadrature. The rotation (O_θ) may then simply be found from a careful phase subtraction (since $\theta \in [0, \pi]$): $O_\theta = \theta_l - \theta_r$ which can be used to approximate the magnitude of the disparity gradient (Γ)[83]:

$$\Gamma = \frac{\sin(\theta_l - \theta_r)}{\sqrt{\frac{1}{4}\sin^2(\theta_r + \theta_l) + \sin^2\theta_l \sin^2\theta_r}} \; ; 0 < \theta_l, \theta_r < \pi \tag{9.14}$$

but only in the absence of deformation. This interpretation of the disparity gradient has also been applied by Brint and Brady[12] to infer a vector direction for edge based stereoscopic computation. The application of an isotropic Gaussian derivative does exibit singularity problems in the estimation of orientation. One example lies with the intersection of orthogonal lines. This can be overcome by the introduction of elliptical Gaussian envelopes but at the expense of the geometrical properties of the filter in estimating orientation.

9.3.2 Rotation and deformation

To obtain estimates of both rotation and deformation, we propose to add a higher level of processing based upon the energy responses from orientationally selective quadrature filter pairs. We form the analytic signal $\vec{Z}(x_1, x_2, \omega, \theta)$:

$$\vec{Z}(x_1, x_2, \theta, \omega) = \Psi(x_1, x_2, \vec{\omega}_g, \vec{\theta}_g) * I(x_1, x_2, \vec{\theta}_i, \vec{\omega}_i) = L(x_1, x_2, \theta, \omega) + j\hat{L}(x_1, x_2, \theta, \omega) \quad (9.15)$$

where $\Psi(x_1, x_2, \vec{\omega}_g, \vec{\theta}_g)$ represents a quadrature filter pair with both spatial frequency and orientational preferences. L and $\hat{L}$ are directional Hilbert transform pairs formed by bandpass filtering.

$\theta = | < \vec{\theta}_i, \vec{\theta}_g > |$ and $\omega = | < \vec{\omega}_i, \vec{\omega}_g > |$ indicate the sensitivity of the envelope of $\vec{Z}(x_1, x_2, \theta, \omega)$ to the 2-D Fourier components in the local neighborhood of the image at $I(x_1, x_2, \vec{\theta}_i, \vec{\omega}_i)$. At an orientation (θ_g) the directional instantaneous frequency $(f_{\theta i})$ may be found from the directional derivative:

$$f_{\theta i} = \frac{1}{2\pi} \mathcal{I}m \frac{d \ln \vec{Z}}{d\vec{\hat{r}}_g} \quad (9.16)$$

where $\vec{\hat{r}}_g = [\cos\theta_g, \sin\theta_g]^T$. That horizontal differences in spatial frequency can be interpreted as a slant about the vertical axis (wall plane) was first proposed by Blakemore leading to his *diffrequency* hypothesis. This is the fundamental operation of the phase-locked loop implemented by Miller[73] in hardware. This can be easily seen assuming linear phase and the expression used to interpret disparity $D(x_1)$ from the method of phase differences in 1-D[90, 42, 100]:

$$D(x_1) \approx \frac{2}{\omega_l + \omega_r}(\omega_l x_1 - \omega_r x_1 + \Phi_l - \Phi_r) \quad (9.17)$$

where ω_l and ω_r refers to the instantaneous angular frequency used to obtain disparity in the left and right image respectively[56]. The horizontal disparity gradient becomes:

$$\frac{dD(x_1)}{dx_1} \approx 2 \cdot \frac{\omega_l - \omega_r}{\omega_l + \omega_r} \quad (9.18)$$

which with simple signals is also valid for narrow band disparity modulation. We also make the observation that in the temporal domain:

$$\omega_t \approx \frac{\Phi_{t_1} - \Phi_{t_2}}{t_1 - t_2} = \frac{\delta\Phi}{\delta t} \tag{9.19}$$

and therefore in this representation stereopsis is indeed a discrete form of velocity ($\vec{v}$) since in 1-D:

$$\vec{v} \;=\; \frac{\omega_t}{\omega_x} \approx \frac{\frac{\delta\Phi}{\delta t}}{\omega_x} \tag{9.20}$$

Equation (9.18) suggests a mechanism for obtaining the stretch ratio between the left to right stereoscopic transformation in a specified orientation by applying equation (9.16) to both images. In principle it is possible to use the directional instantaneous frequency processed in a circle to compute the divergence and deformation between stereoscopic transformations. This turns out to be numerically unstable because of singularities in phase space[41]. This form of analysis is analogous to computing the stress forces on a planar surface. In the same manner as the *rosette* formation is commonly used in material science using strain guages, we can explore spatial arrangements of filters to obtain deformation. At present, we restrict ourself to local operations to illustrate the basic idea.

To avoid the computation of instantaneous frequency, we use the envelope from equation (9.15) and assume both image sensors are fixating at the same point in the world at the origin $x_s(0)$. We form:

$$|\bar{Z}_l(\theta, \omega_l; x_s(0))| = |\Psi[\vec{\theta}_g; \bar{\omega}_g, x_s(0)] * I_l[x_s(0)]|$$

which is the energy response from 2-D filters with similar spatial frequency preferences applied at a single point but different orientations. This defines an energy function from a circle to an interval on the real line $[0, \pi]$. A similar representation is also computed for the right image. Under stereoscopic transformation, the response from the directionally selective filters will encode the 1st order slope of the surface by an orientational modulation. We create an additional Hilbert Transform pair defined in the left image by:

$$P_l(\theta) = |\bar{Z}_l(\theta, \omega_l; x_s(0))| * \chi(\theta, v_g) = s(\theta) + js(\theta) \tag{9.21}$$

where $\chi(\theta, v_g)$ is a 1-D bandpass quadrature filter with centre angular frequency v_g. Thus we are decoding the disparity signal by the group delay in the envelope[91] under stereoscopic transformation. This also implies that the first order characteristics can also be obtained from a direct quadrature phase relationship. We consider the phase in response to Gaussian noise[10]:

$$\upsilon_g\theta + \Phi(\theta) = \tan^{-1}\frac{\hat{s}(\theta)}{s(\theta)} \tag{9.22}$$

The instantaneous frequency is defined by[22]:

$$f_i(\theta) = \frac{1}{2\pi}Im\frac{dP_l(\theta)}{d\theta} = \frac{1}{2\pi}\frac{s(\theta)\dot{\hat{s}}(\theta) - \hat{s}(\theta)\dot{s}(\theta)}{s^2(\theta) + \hat{s}^2(\theta)} \tag{9.23}$$

Thus we may apply a quadrature pair of filters and their derivatives to obtain the local instantaneous frequency[56]. Because $|\bar{Z}_l(\theta, \omega_l; x_1, x_2)|$ is bandpass, $P_l(\theta)$ is also:

$$P_l(\theta) = E_l(\upsilon_g, \omega_l, \theta)\exp[\upsilon_g\theta + \Phi(\theta)] \tag{9.24}$$

Notice the envelope $E_l(\upsilon_g, \omega_l, \theta)$ is sensitive to the directional spatial frequencies present in the original image. We expand the angular part of equation (9.24), ignoring second and higher order terms:

$$\upsilon_g\theta + \Phi(\theta) \approx \Phi(\theta_o) + \upsilon_g\theta + \Phi'(\theta_o)(\theta - \theta_o) \tag{9.25}$$

the local instantaneous frequency is then:

$$f_{il}(\theta) = \frac{1}{2\pi}(\upsilon_g + \Phi'(\theta)) = \frac{1}{2\pi}\upsilon_l \tag{9.26}$$

Now assume that $f_{il}(\theta)$ is approximately constant between θ and $\theta + \delta\theta$, as is the envelope[81]:

$$E_l(\upsilon_g, \omega_l, \theta) \approx E_l(\upsilon_g, \omega_l, \theta + \delta\theta) \tag{9.27}$$

The violation of this assumption is well known in telecommunications theory and its consequence introduces phase distortion in the carrier signal. To overcome these restrictions it is usual to amplify the frequency components present in the passband of interest. This may be achieved by the subtraction of the Laplacian. This proposal provides a direct application for cells with similar profiles present in biological systems.

To illustrate these ideas, consider the transformation due to the slant of a planar surface in the right image relative to the left. We have:

$$P_r(\theta) = E_r(\upsilon_g, \omega_l + \delta\omega, z)\exp[j\upsilon_l z] \tag{9.28}$$

where $z = \theta + \frac{d\theta}{dx_1}$ and the local change in spatial frequency ($\delta\omega$) occurs because of stretches explained by equation (9.11). The physical interpretation is a signal that is both phase (PM) and amplitude (AM) modulated. PM occurs because of equation (9.10) while the AM occurs because of equation (9.11). Since the spatial frequency term in the energy response of a 2-D bandpass filter only responds to the modulus of the departure of the fundamental image signal from the centre frequency of the filter, the transformed energy response or AM signal cannot be uniquely determined in this representation. This is a consequence of local stretches of the image field. We therefore have to rely on the orientational phase differences to demodulate the disparity signal. Consider the instantaneous frequency of the right image by differentiating the angular part of equation (9.28):

$$f_r(\theta) \approx \frac{1}{2\pi}\frac{d(\upsilon_l z)}{dz}\frac{dz}{d\theta} = \frac{1}{2\pi}\upsilon_l[1 + \frac{d(\frac{d\theta}{dX_1})}{d\theta}] \tag{9.29}$$

The local *diffrequency* ($f_d(\theta)$) between the left and right stereoscopic image pairs becomes:

$$f_d(\theta) \approx \frac{1}{2\pi}\upsilon_l\frac{d(\frac{d\theta}{dX_1})}{d\theta} \tag{9.30}$$

Using equation (9.24) we can estimate the orientation disparity ($O_d(\theta)$) using the method of phase differences:

$$O_d\,(\theta_o) \approx \frac{2}{\upsilon_l(\theta_o) + \upsilon_r(\theta_o)}[\Phi_l(\theta_o) - \Phi_r(\theta_o)] \tag{9.31}$$

We can track orientational disparities in the form of a 1-D phase-locked loop or take a weighted estimate through resolutions of filtering.

Consider from continuum mechanics, the term analogous to the *material derivative*:

$$\frac{d(\frac{d\theta}{dX_1})}{d\theta} = (a_{22} - a_{11})\cos 2\theta + (a_{12} + a_{21})\sin 2\theta \tag{9.32}$$

We can re-write this term as:

$$\mathcal{R}e\ \sqrt{(a_{12} + a_{21})^2 + (a_{22} - a_{11})^2}\exp[j2\theta - \eta] \tag{9.33}$$

where $\eta = \tan^{-1}\frac{a_{12} \pm a_{21}}{a_{22} - a_{11}}$. The modulus of the coefficient in the above term gives the magnitude of the slant. The phase gives the direction of the slant.

An examination of equations (9.18,9.30,9.31) indicates that we can estimate the above term by:

$$\frac{d\frac{d\theta}{dX_1}}{d\theta} \approx \frac{2}{f_l + f_r}f_d(\theta) \tag{9.34}$$

since we can consider taking a reference from either image. Therefore diffrequency can be used to approximate the slant of the surface, by using the anisotropic behaviour of deformation. However, because of numerical difficulties in the computation of instantaneous frequency, we restrict our consideration to equation (9.31).

9.4 Methods

To implement this scheme, we have applied Gabor function as our 2-D orientation and spatial frequency bandlimited filter. To preserve orientational selectivity over spatial frequency, we have applied filters whose envelope aspect ratio is 2:1 in favour of orientation.

We assume that orientational disparities will be small. The risks from aliasing which are problematic with a phase differencing technique are therefore reduced. We therefore apply a weighted estimate and cyclotorsion as opposed to Phase-locking. From the energy responses from the band of filters, we obtain the weighted mean disparity, and diffrequency estimates by convolving with the quadrature Gabor function in both images by three resolutions of coarse to fine filtering:

$$O_{est}(\theta) = \frac{\sum_{j=1}^{n} E_{lr}(\theta, v_j) O_d(\theta)}{\sum_{j=1}^{n} E_{lr}(\theta, v_j)} \tag{9.35}$$

where $E_{lr}(\theta, v_j)$ refers to the product of energy from the 1-D Gabor filters used to interpret disparity (fig. 9.2). A similar expression was applied to estimate the local diffrequency. At a given scale, we estimate the least squares mean phase difference (curl) using a similar mechanism proposed for camera vergence[55], which will act in the same manner as cyclotorsion:

$$\mathbf{S_l S_r}^T [\mathbf{S_r S_r}^T]^{-1} = \mathbf{R}_\theta \tag{9.36}$$

$$\mathbf{S_r S_l}^T [\mathbf{S_l S_l}^T]^{-1} = \mathbf{R}_\theta^{\,T} \tag{9.37}$$

with:

$$\tan O_\theta = \frac{\sin O_\theta}{\cos O_\theta} \tag{9.38}$$

as the least squares phase difference, with coefficients taken from:

$$\mathbf{R}_\theta - \mathbf{R}_\theta^{\,T} = 2\sin\mathbf{O}_\theta \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{9.39}$$

and

$$\mathbf{R}_\theta + \mathbf{R}_\theta^{\,T} = 2\cos\mathbf{O}_\theta \ \mathbf{I} \tag{9.40}$$

Here $\mathbf{S_l}, \mathbf{S_r}$ refer to the 2xM matrices of Hilbert transform pairs $(s(\theta) + js(\theta))$ and R_θ contains the least squares phase difference. Disparity in this case is interpreted using the centre frequency of the bandpass filter $\chi(\theta, v_\theta)$. We note that the the curl also defines the magnitude of slant about a horizontal axis. Removing the mean rotation also reduces both the risks from phase aliasing and disparity error[56] because we increase the range of orientational disparity detection. The remaining disparity signal is now easy to compute. To extract the magnitude and phase of the deformation component we make the analogy to the computation of orientation[47] using tensor field filtering. We therefore consider the eigenfunctions of the tensor:

$$\sum_{\theta=0}^{M} O_{est}(\theta)[T_k - \frac{1}{2}I] \tag{9.41}$$

where $T_k = [\cos\theta, \sin\theta]^T[\cos\theta, \sin\theta]$. In this representation, the eigenvalues correspond to the magnitude of the deformation, while the eigenvectors correspond to the surface slant. Alternatively, we could apply a weighted least squares fit[55]:

$$\mathbf{E\Phi D} = \mathbf{Ed} \tag{9.42}$$

where $\mathbf{E}$ is an $M\mathrm{x}M$ matrix with leading diagonal formed from the product of energy responses. $\mathbf{d}$ represents the measured disparities ($M\mathrm{x}1$), $\mathbf{\Phi} = [\cos 2\theta_i, \sin 2\theta_i, 1]$ represents the orientation of the ith disparity measurement and the constant refers to the rotation component , and $\mathbf{D} = [a_1, a_2, a_3]^T$ is a vector containing coeficients that relate to half the magnitude of deformation resolved into horizontal and vertical components (a_1, a_2) with the rotation given by (a_3).

9.4.1 Results

We present some simple data to highlight the principle that we have introduced. In figure 9.2 we show the energy response to an ideal corner that has been subjected to a horizontal shear. To obtain the total orientational disparity requires the summation of the rotation term to the vertical component of deformation. The energy response maxima correspond to the mutual alignment in orientation of quadrature filters located at an ideal and sheared corner. The estimated disparity and local diffrequency from subsequent postprocessing the band of filters is also shown. We also present a second example taken from a random dot stereogram that has also been subjected to horizontal shear. The Needle diagram (fig. 9.3) indicates the direction and magnitude of the disparity gradient in the absence of translational disparity.

9.5 Conclusion

We have shown, that the first order differential transformations of slanted surfaces can reduce to differences in spatial frequencies between left and right stereoscopic image pairs. Thus we

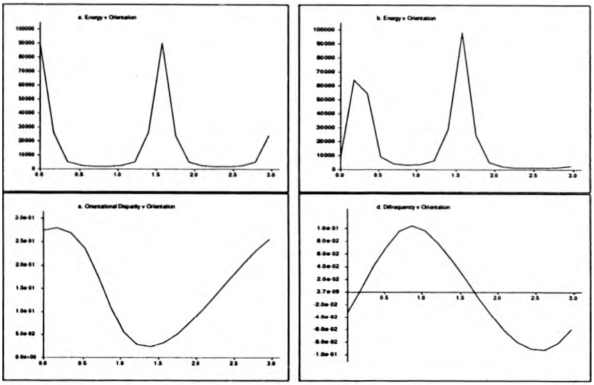


Figure 9.2: *(a) Energy response at an ideal corner. (b) Energy response after horizontal shear taken at the fixation point. (c) Orientation disparity. Curl estimated at 0.148 rads, Def was estmated as:* $D_{x1} = -0.0065$, $D_{x2} = 0.1224$ *rads. Actual disparity gradient at 0.254 rads with an error of 0.016 rads. (d) Weighted diffrequency estimates.*

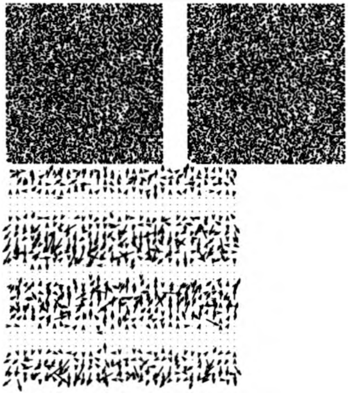


Figure 9.3: **Top** *Random dot stereogram subjected to a modulated horizontal shear.* **Bottom** *Disparity gradient vector field.*

have combined the theories of Blakemore[7] and Koenderink and Van Doorn[49] into an equivalent representation. Naturally, this scheme can equally be applied to visual motion.

There are two situations that require clarification given the methods proposed in this paper. Firstly, consider the case that the local image field contains strongly directional energy. The local image field will only experience rotations or divergences depending on the slant of the surface, and orientation of the signal. In the case of an isolated but elongated line, the local stereoscopic field will only amount to a rigid rotation. Secondly, consider the case of an elongated pair of lines that bisect the principal axes of the deformation component. Under these conditions, the local image field will also only be observed as a rigid rotation. We have proposed a technique that restricts consideration to rotational differences.

These approximations to the local behaviour of the disparity vector field draw interesting conclusions regarding the architecture of the human visual cortex. In view of the large receptive field of orientationally selective filter that we apply, we do not expect a considerable change in orientational energy between neighboring points. To avoid circular convolutions at each point, we can propose that the 1-D phase locking iteration can be performed by stretching the M orientation samples at each point onto a linear array, and applying the 1-D operators over neighboring pixel locations. From this point of view, the architecture of the human striate cortex is highly suited towards the type of processing proposed in this paper.

To incorporate these ideas into the algorithm presented in chapter 8 is an area of further research. A correct solution is out of the scope of this thesis. However, creating a further Hilbert Transform pair from the Energy responses of quadrature filter pairs and tracking orientational disparities from phase differences is a natural application of the method of phase differences.

Chapter 10

SAMPLE IMAGERY

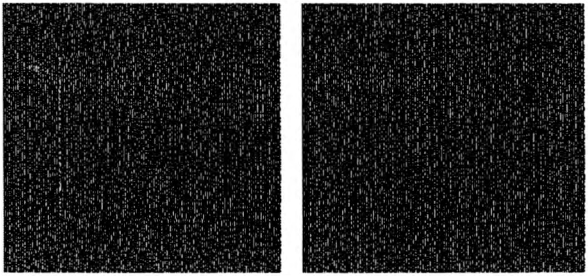


Figure 10.1: *Random Dot Stereogram showing a four layered surface.*

10.1 Sample Imagery

To highlight the extent of the work in this thesis, we present a small sample of stereoscopic intensity surfaces to a small selection of artificial and real data. Unless otherwise stated, all images were compressed under the Compact Pyramid code. The figure captions also include a reference to the proposed algorithm within the context of this work.

10.2 Random Dot Stereograms

We present a single example from a RDS stimulus. The stimulus comprises of four depth layers with disparities of 0 pixels, 5 pixels, 10 pixels and 15 pixels in the central region (fig 10.2). The stereoscopic pair has maximum entropy since 50% dots are black and are randomly distributed in the field. Results are presented for both 1-D and 2-D filtering techniques. Since the epipolar constraint is rigidly adhered to and there is no added noise, the 1-D Phase-locking algorithm has given the most accurate results.

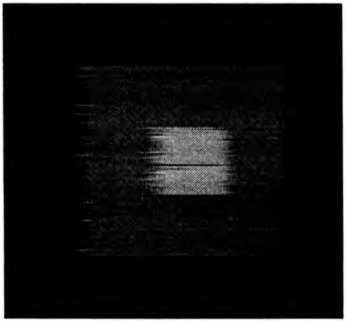


Figure 10.2: *Intensity depth surface obtained from a 1-D filter. Disparity interpretation using the local instantaneous frequency. See section 6.7.4.*

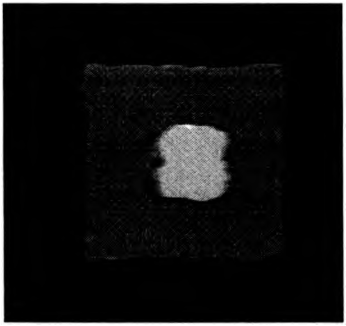


Figure 10.3: *Intensity depth surface from Phase-locking with 12 orientationally selective filters (1/4 cpp) interpreting disparity with the centre frequency of the Gabor function. See section 8.4.*

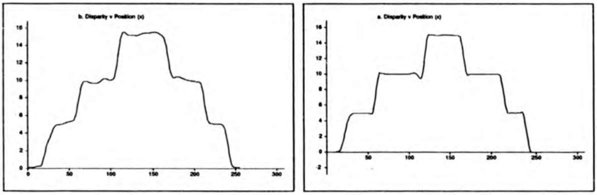


Figure 10.4: *Disparity obtained from the Random Dot Stereogram in figure 10.2. Raster line 128 presented. (a) 1-D filter (1/8 cpp) (b) 12 2-D filters (1/4 cpp) aspect ratio 1:1.*

10.3 The Pentagon

This image pair has received considerable attention over the years as a stereo sequence (fig. 10.5). We haved passed this sequence through several of our proposed algorithms which we show as a series of intensity depth measurements in figures 10.6 to 10.8.

10.4 The Tennis Snack

This image pair highlights a potential problems with the Phase-Locking iteration owing to the sparse image detail. The algorithm is noise sensitive under these conditions. Within this stereo pair, the fixation point lies in the plane of the image with the lunchbox and ball with negative disparity relative to fixation. Two image disparity intensity surfaces are again presented for a 1-D and 2-D Phase-locking algorithm (fig. 10.10 to 10.11).

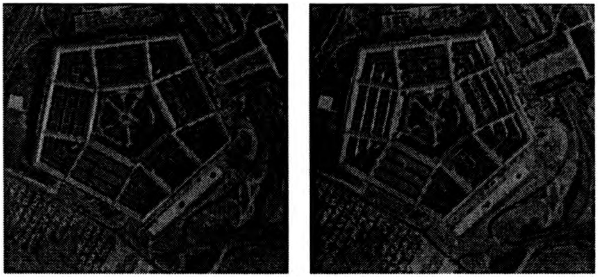


Figure 10.5: *An ariel view of the U.S Pentagon. Disparities were less than 4 pixels in a 256x256 reduction of the original image.*

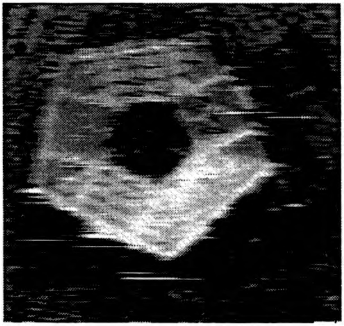


Figure 10.6: *Using a 1-D filter (1/8 cpp) and local instantaneous frequency to intepret disparity. Histogram equalization was used to enhance the stereo features of this image. See section 6.7.4.*

Figure 10.7: *12 orientationally selective filters (1/4 cpp) used to interpret a horizontal component of disparity. See section 8.4.*

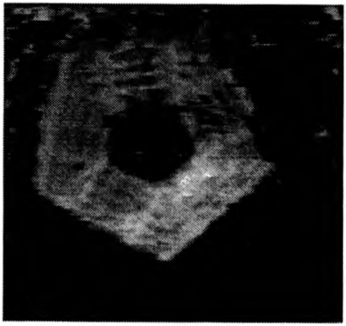


Figure 10.8: *12 orientationally selective filters (1/4 cpp) and their derivatives were used to obtain the instantaneous frequency and interpret disparity. See section 8.7.*

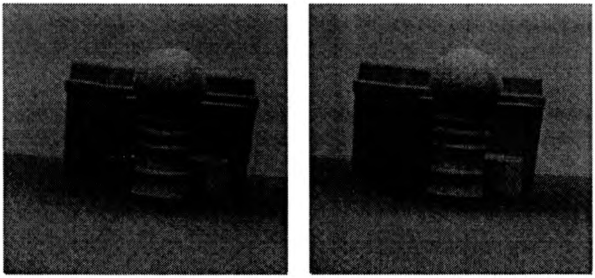


Figure 10.9: *A lunch box and tennis ball standing in a featureless background.*

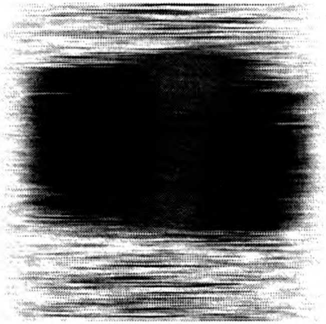


Figure 10.10: *A single dimension filter (1/8 cpp) to interpret disparity from the local instantaneous frequency. See section 6.7.4.*

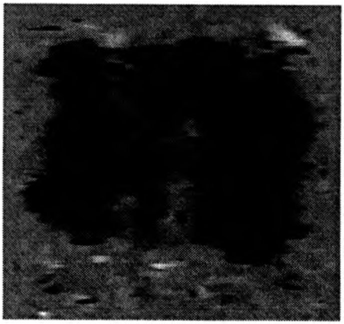


Figure 10.11: *12 orientationally selective (1/4 cpp) filters used to interpret horizontal disparity. See section 8.4.*

Chapter 11

SUMMARY AND CONTRIBUTIONS TO THE FIELD

11.1 Contribution to knowledge

We have proposed a technique that can in principle obtain both vertical and horizontal components of translational disparity. By interpreting the findings from neurophysiological data, we have proposed a weighted solution, by which vertical components of disparity are only obtained in image regions that contain orthogonal signals, while horizontal disparity is obtained from a weighted probabilistic measure to resolve a horizontal component of disparity based upon the presence and absence of edge information alone. Indeed, the corner probabilistic measure that we have found is also new, which followed directly from the recognition of the stereoscopic aperture problem and the work of Wilson et al [48]. To a lesser degree, we have proposed a different interpretation of the disparity gradient in frequency space. From a theoretical standpoint, we have obtained a formalised proof which indicates the conditions in which the Gabor function can be approximated as a quadrature filter.

More importantly, we have shown that the method of phase differences can be applied to real image data to extract depth information from a stereoscopic pair of images. The work has yielded two related interpretations of disparity. Firstly, the phase difference equation may be interpreted as a distance by the center frequency of the applied band-pass filter pairs. Even though this expression requires considerable damping with a 1-D filter, its least-squares extension into the 2-D image plane yields particularly pleasing results. This is because the mathematical operations are linear and final depth surfaces are obtained without any form of thresholding process. In the second place, we have used the instantaneous frequency obtained from the derivative of phase to directly interpet disparity. This technique does, however, retain difficulties in its implementation in 2-D images because of stationary phase.

While the latter method retains an improved description of the image and can improve the interpretation of disparity, it inherently holds instability problems associated with stationary phase, which we prove is due to interference. We see little alternative at present but to apply a series of thresholds to restrict the analysis from these image regions, although we have inferred that it is possible to weight out these regions by filtering between and across scale of filtering. Fundamental to the method of phase differences lies the notion of disparity smoothing at each resolution of filtering which is the dominant operation required to obtain stability. Finally, we should gratefully acknowledge that the original notion for the method of phase differences was independently conceived by both Larcombe [58] and Wilson [100] who are both at present lecturers in Computer Science at the University of Warwick. These notions have, however, been independently extended into a plausible theory for Computer Vision.

11.2 Further Considerations

There are a number of areas that still require further work. We have already inferred, that the method of phase differences is suitable for the extraction of translational and compression/expansion deformations between corresponding image pairs. Within this context, we suggest that the methods proposed in chapter 8 can be used to obtain approximate correspondence or the translation part of the disparity field.

Phase wrap-around is a problem. We have inferred two conditions by which this problem may be reduced by either a vergence mechanism or phase-locking iteration. From the geometry of stereopsis, we could equally enforce a left to right matching paradigm (which would require truncating to zero disparity estimates below a given minimum) or take an ensemble of averages from several views of the same scene, but with vergence initiated from different starting positions. A Phase-locking iteration is sensitive to noise and is particularly concerning with sparse image detail. There are two avenues that might be persued to reduce this problem: Firstly, the Phase-locking iteration may be perceived as an attractive force. Stability would be expected to improve, if a suitable counteracting mechanism could be found. Secondly, recent work has involved the notion of applying a *Phase-pass* iteration, which in terms of a Phase-locking iteration in phase space is analogous to bandpass filtering the frequency domain. Improved noise sensitivity is the main aim with this technique.

Future work must also combine further feature based stereopsis, perhaps with the methods proposed in this thesis. We know from the extensive psychophysical studies and intuition, that stereopsis is neither wholly described by low or higher levels image descriptions. Ultimately, we should consider both the advantages from feature based analysis and those obtained from the methods presented in this thesis. In particular, we could begin to from monocular depth cues from spatial pattern recognition. Indeed, this is an immediate extension of our present work on feature analysis, where we improve on our corner detector by extracting orientation from a *spatially* separated band of filters that are applied at a constant radius a predefined origin. In this way, we hope to disambiguate the problems in the separation of "Y", "X" and "L" junctions, which we cannot hope to extract from filter pairs that are all located at a single point in space. At the same time we increase the noise sensitivity in the detection of corners, since we are now integrating over a greater region of the edges present at image corners. Such a scheme also would hold particular advantages in stereopsis, because we would not generally wish to bring different, but locally indistinguishable features into correspondence based upon local operators.

Appendix A

CORRELATION DATA

1-D CORRELATION DATA

Algorithm Description	Figure	Correlation (r)	Mean (x)	s.d (σ_n)
1-D local energy maxima	B.2	4.59	-6.3	0.96
	B.3	0.83	-1.70	0.284
1-D local energy maxima with median filtering.	B.2	0.599	-7.08	1.03
1-D Wilson and Knuttson Algorithm.	B.2	0.39	3.98	1.28
	B.4	0.04	-1.77	1.97
1-D Recursive filtering using the Gabor centre frequency to interpret disparity.	B.2	0.88	-3.03	0.52
	B.3	0.95	-1.59	0.83
	B.4	0.93	-1.75	0.34
1-D continuity constraint using 1st derivative of Gaussian Hilbert pairs, Compact Pyramid and instantaneous frequency to interpret disparity.	B.2	0.93	-1.53	0.33
	B.3	0.92	0.45	0.15
	B.4	0.93	-3.0	0.46
	B.5	0.96	-0.65	0.20
1-D continuity constraint using Gabor filter pairs, Compact Pyramid and instantaneous frequency to interpret disparity.	B.2	0.95	-1.41	0.28
	B.3	0.89	-0.41	0.18
	B.4	0.93	-2.6	0.41
	B.5	0.96	-2.23	0.35

Table A.1: *A comparison of 1-D techniques used in this thesis to the four control raster line sequences in figures B.3, B.2, B.4, and B.5. Data represents the statistical error between the manually determined disparity and techniques applied in this thesis.*

2-D CORRELATION DATA

Algorithm Description	Figure	Correlation (r)	Mean (x)	s.d (σ_n)
Phase-locking based upon the horizontal component of disparity from equation (8.14).	B.2	0.96	-0.99	0.23
	B.3	0.88	-0.51	0.20
	B.4	0.96	0.17	0.20
	B.5	0.96	0.17	0.20
Phase-locking based upon the horizontally resolved disparity estimate using edge probability measure.	B.2	0.96	-1.60	0.29
	B.3	0.86	-0.42	0.21
	B.4	0.89	-2.11	0.41
	B.5	0.95	-0.58	0.22
Phase-locking using a single oriented 2-D Gabor filter to interpret disparity.	B.2	0.94	-0.95	0.29
	B.3	0.90	1.03	0.21
	B.4	0.91	0.31	0.29
	B.5	0.95	0.27	0.23
Phase-locking using the local directional instantaneous frequency to intepret disparity. Similar thresholds were applied than in the 1-D algorithm.	B.2	0.95	-1.5	0.30
	B.3	0.82	-0.98	0.23
	B.4	0.91	-2.4	0.40
	B.5	0.95	-0.88	0.24

Table A.2: *A comparison of 2-D techniques used in this thesis to the four control raster line sequences in figures B.3, B.2, B.4, and B.5.*

Appendix B

TEST DATA

Figure B.1: *Images taken from a room with a 512x512 CCD camera. Data was compressed to 256x256 using the Compact Pyramid. (a) Left image (b) Right image.*

B.1 Test Data

The following stereo pairs of images were taken with a 512x512 CCD camera and compressed to 256x256 using the Compact Pyramid[15]. Epipolar geometry is assumed with parallel camera vergence. The following raster sequences have been consistently studied throughout the thesis as a control for comparative purposes.

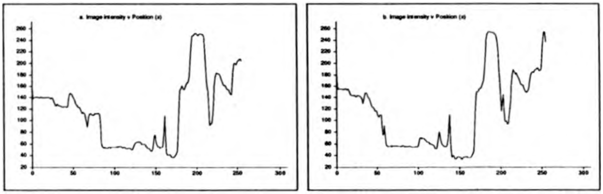


Figure B.2: *Raster line 110. (a) Left image raster (b) Right image raster.*

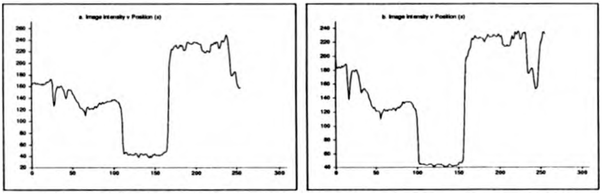


Figure B.3: *Raster line 10 (a) left image (b) right image.*

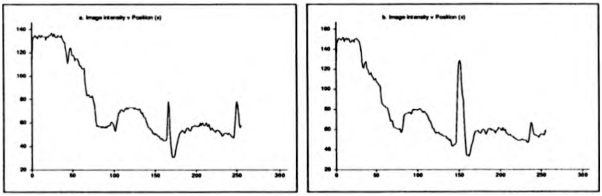


Figure B.4: *Raster line 138. (a) Left image (b) Right image.*

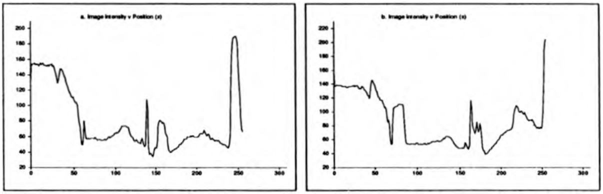


Figure B.5: *Raster line 118 from figure B.1. (a) Left image (b) Right image. This stereo pair sequence proved highly unstable.*

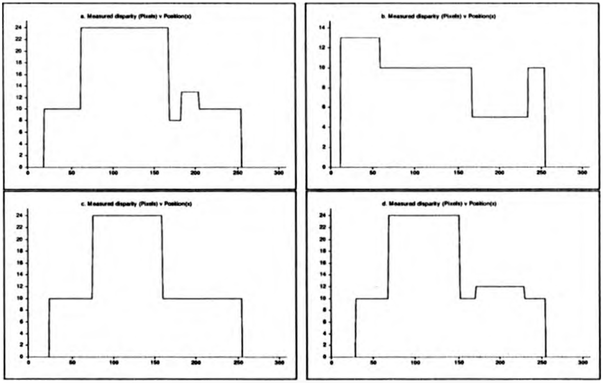


Figure B.6: *Measured edge based disparity estimates for control 1-D stereo raster lines. (a) Raster line 110. (b) Raster line 10. (c) raster line 138. (d) raster line 118.*

Bibliography

[1] Adelson.E.H. and Movshon.J.A. Phenominal coherence of moving visual patterns. *Nature*, 300:523–525, 1982.

[2] Atherton.T.J. The detection and measurement of visual motion. *Pattern Recognition Letters*, 5,2:157–163, 1987.

[3] Baker.H.H. *Depth from Edge and Intensity based stereo.* PhD thesis, Standforth University, 1982.

[4] Bastiaans.M.J. Gabor's expansion of a signal into gaussian elementary signals. *Proc. IEEE*, 68:538–539, 1980.

[5] Bishop.P.O. Binocular vision. *Adler's Physiology of the eye.*, pages 558–614, 1975.

[6] Bishop.P.O. and Pettigrew.J.D. Neural mechanisms of binocular vision. *Vis.Res.*, 1986.

[7] Blakemore.C. A new kind of stereoscopic vision. *Vis. Res*, 10:1181–1200, 1970.

[8] Blakemore.C., Fiorentini.A., and Maffei.L. A second neural mechanism of binocular depth discrimination. *J.Physiol.*, 226:725–749, 1972.

[9] Bolles.R.C., Baker.H.H., and Marimont.D.H. Epipolar-plane image analysis: An approach to determining structure from motion. *Int. J. Computer Vision*, 1:7–55, 1987.

[10] Bracewell.R.N. *The Fourier Transform and its applications.* McGraw-Hill Book Co.,London, 1986.

[11] Braddick.O. Binocular fusion and perceptual analysis. *Ph.D Thesis*, 1968.

[12] Brint.A.T. and Brady.M. Stereo matching of curves. *Proc. 5th ALVEY Vision Conference: Reading*, pages 187–193, 1989.

[13] Brown.J and Glazer.E.V.D. *Telecommunications*. Arrowsmith Ltd., 1974.

[14] Burt.P. and Julesz.B. Modifications of the classical notion of panum's fusional areas. *Perception*, 9:671–682, 1980.

[15] Burt.P.J. and Adelson.E.H. The Laplacian pyramid as a compact image code. *IEEE Trans.Comm.*, COM-31.4:532-540, 1983.

[16] Canny.J.F. A computational approach to edge detection. *IEEE Trans. PAMI*, 8:679–698, 1986.

[17] Clark.J. and Lawrence.P.D. A theoretical basis for diffrequency stereo. *Comp. Vis. Imag. Proc.*, pages 1–19, 1986.

[18] Daugman.J.G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimised by two-dimensional visual cortical filters. *J.Opt.Soc.Am*, 2(7):1160-1168, 1985.

[19] DeValois.R.L., Albrecht.D.G., and Thorell.L.G. Spatial frequency selectivity of of cells in the macasque visual cortex. *Vis.Res.*, 22:545–599, 1982.

[20] Field.D.J and Tolhurst.D.J. The structure and symmetry of simple cell receptive fields profiles in the cat's visual cortex. *Proc.Soc.R.Soc.Lond.B*, 228:379 400, 1986.

[21] Foschini.G.J., Gitlin.R.D., and Weinstein.S.B. On the selection of a two-dimensional signal constellation in the presence of phase jitter and Gaussian noise. *Bell Sys. Tech. J.*, pages 927 – 965, 1973.

[22] Franks.L. *Signal Analysis*. 1969.

[23] Freeman.W.T and Adelson.E.H. Steerable filters for image analysis. *submitted for publication*, 1990.

[24] Frisby.J.P and Mayhew.J.E.W. Global processes in stereopsis:some comments on ramachandran and nelson(1976). *Perception*, 6:195–206, 1977.

[25] Gabor.D. Theory of communication. *J.Inst.Elec.Eng*, 93:429–459, 1946.

[26] Gennery.D.B. *Modelling the environment of an exploring vehicle by means of stereovision*. PhD thesis, Stanford A.I lab., 1980.

[27] Georgeson.M. Human vision combines oriented filters to compute edges. *AVA modelling in Vision meeting*, 1990.

[28] Gibson.J.J. *The perception of the visual world.* The riverside press,Houghton Mufflin C., 1950.

[29] Gillan.B. and Lawergen.B. The induced effect, vertical disparity and stereoscopic theory. *Percept. and Psychophys.*, 34:121–130, 1983.

[30] Gimel'farb.G.L., Marchenko.V.B., and Rybak.V.I. An algorithm for automatic identification of identical sections on stereopair photographs. *Kybernetica*, 2:311–322, 1972.

[31] Granlund.G.H. In search of a general picture processing operator. *Comp. Graphics,Image Processing*, pages 155–173, 1978.

[32] Grimson.E. *From Images to Surfaces: A computational study of the Human early visual processing system.* MIT press, 1981.

[33] Haglund.L. *Hierarchical Scale Analysis of Images using Phase Description.* PhD thesis, Linkoping University: Thesis No. 168, 1989.

[34] Hannah.A.J. *Computer matching of areas in stereo images.* PhD thesis, Stanford A.I. Lab., 1974.

[35] Helmholtz.H.Von. *Handbook of Physiological optics, 3rd Ed.* Rochester, N.Y, 1924.

[36] Helstrom.C.W. *Probability and Stochastic processes for Engineers.* Macmillan Pub. Co., 1984.

[37] Henderson.R.L., Robert.L, Walter.J., Miller.J, and Grosch.C.B. Automatic stereo recognition of man-made targets. *Soc. Photo-Opt. Instr. Eng.*, 186, 1979.

[38] Hubel.D.H. *Eye, Brain and Vision.* Scientific American Library, 1988.

[39] Hubel.D.H. and Wiesel.T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J.Physiol.*, 160:106–154, 1962.

[40] Jeffries.A. *Mathematics for Scientists and Engineers.* 1980.

[41] Jepson.A.D. and Fleet.D.J. *Scale-Space Singularities.* (Ed.) Faugeras.O, Proc. ECCV,Antibes , Springer-Verlag, 1990.

[42] Jepson.A.D and Jenkin.R.M. The fast compuation of disparity from phase differences. *IEEE proc. CVPR:California*, pages 398–403, 1989.

[43] Jones.J.P., Palmer.L.A., and Daugman.J.G. Information management in the visual cortex. *Science*, 1986.

[44] Julesz.B. *Foundations of cyclopean perception*. Chicago, Univ. Chicago Press., 1971.

[45] Julesz.B. and Miller.J.E. Independent spatial-frequency-tuned channels in binocular fusion and rivalry. 4:125–143, 1975.

[46] Knutsson.H. *Filtering and Reconstruction in Image Processing*. PhD thesis, Linkoping University, 1982.

[47] Knutsson.H. Representing local structure using tensors. *6th Scandinavian Conference on Image Analysis*, pages 244–251, 1989.

[48] Knutsson.H., Wilson.R.G., and Granlund.G. Estimating the orientation of local anistropic 2-d signals. *IEEE.A.S.S.P.Spectral Estimation Workshop.Florida*, pages 234–239, 1983.

[49] Koenderink.J.J. and Van Doorn.A.J. Geometry of binocular vision and stereopsis. *Biological Cybernetics*, 21:29–35, 1976.

[50] Koenderink.J.J and Richards.W. Two-dimensional curvature operators. *J.Opt.Soc.Am.*, 5:1136–1141, 1988.

[51] Lange.F.H. *Correlation Techniques*. Iliffe Books Ltd.(ed) Johns.P.B, 1967.

[52] Langley.K. Phase from Gabor filters. Technical report, Univ.Warwick.R119., 1988.

[53] Langley.K. and Atherton.T.J. A computational theory of stereopsis in the mammalian striate cortex. *submitted to IVC*, 1988.

[54] Langley.K. and Atherton.T.J. A confidence measure for the aperture problem. *Submitted to IVC*, 1988.

[55] Langley.K. and Atherton.T.J. Phase relationships in stereoscopic computation. *RARDE: Applied Vision Association on Modelling in Vision*, 1990.

[56] Langley.K., Atherton.T.J., Wilson.R.G., and M.H.E.Larcombe. *Vertical and Horizontal Disparities from Phase.* (Ed.) Faugeras.O, Proc. ECCV,Antibes , Springer-Verlag, 1990.

[57] Langley.K, Rogers.B.J, and Brady.J.M. The Computation of Deformation and Rotation in Stereopsis. *accepted in BMVC90, University of Oxford*, 1990.

[58] Larcombe.M.H.E. Personal communication of unpublished work. 1984.

[59] Levine.M.D., O'Handley.D.A., and Yagi.G.M. Computer determination of depth maps. *Computer Graphics and Image Processing*, 2:131–150, 1973.

[60] Longuet-Higgins.H.C. The role of the vertical dimension in stereoscopic viewing. *Perception*, 11:377–386, 1982.

[61] Longuet-Higgins.H.C. Double and triple ambiguities of a scene viewed from two different camera locations. *3rd Alvey Vision conference*, pages 237–243, 1987.

[62] Lunneberg.R. *Mathematical theory of Binocular vision.* Princeton Univ. Press., 1947.

[63] Maffei.L. and Fiorentini.A. The visual cortex as a spatial frequency analyser. *Vis.Res.*, 13:1255–1267, 1973.

[64] Marcelja.S. Mathematical description of the responses of simple cortical cells. *J.Opt.Soc.Amer*, 70:1297–1300, 1980.

[65] Marr.D. *Vision.* W.H.Freeman & Co., 1982.

[66] Marr.D. and Hildreth.E. Theory of edge detection. *Proc.R.Soc.Lond*, 207:187–217, 1980.

[67] Marr.D. and Poggio.T. A computational theory of human stereo vision. *Proc.R.Soc.Lond.*, 204:301–328, 1979.

[68] Marr.D. and Ullman.S. Directional selectivity and its use in early visual processing. *Proc.R.Soc.Lond.*, pages 151–180, 1981.

[69] Mase.G.E. *Continuum Mechanics.* McGraw-Hill Book Co., 1970.

[70] Maske.R., Yamane.S., and Bishop.P.O. End-stopped cells and binocular depth discrimination in the strite cortex of cats. *Proc.Soc.Lond.B.*, 229:257–276, 1986.

[71] Mayhew.J.E. and Frisby.J.P. Convergent discriminations in narrow-band filtered random-dot stereograms. *Vision Research*, 19:63–71, 1979.

[72] Mayhew.J.E.W and Frisby.J.P. *Psychophysical and Computational Studies towards a Theory of Human Stereopsis. In* Computer Vision, Ed. Brady.J.M, North-Holland Book Co., 1981.

[73] Miller.W.M. Video image stereo matching using phase-locked loop techniques. *IEEE Proc. Int. Conf. on Robotics and Automation.*, 1:112–117, 1986.

[74] Montgomery.H.C. Transistor noise in circuit applications. *Proc. I.R.E*, 40(11):1162–1471, 1952.

[75] Moravec.H.P. *Obstacle avoidance and navigation in the real world by a seeing robot rover.* PhD thesis, Stanford A.I Lab., 1980.

[76] Mori.K.I., Kidode.M., and Asada.H. An iterative prediction and correction method for automatic stereocomparison. *Computer Graphics and Image Processing*, 2:393–401, 1973.

[77] Movshon.J.A. Two-dimensional spatial frequency tuning of cat striate cortical neurons. *Society for Neuroscience, Abstracts*, 9th Annual meeting:799, 1979.

[78] Noble.A. Finding corners. *3rd Alvey Vision Conference*, pages 267–274, 1987.

[79] Ogle.K.N. *Binocular Vision*. Saunders, 1964.

[80] Oppenheim.A.V and Lim.J.S. The importance of phase in signals. *Proc.IEEE.*, 69,5:529–541, 1981.

[81] Papoulis.A. *Signal Analysis*. McGraw-Hill Int. Editions, 1987.

[82] Papoulis.A. *Systems and Transforms with Applications in Optics*. McGraw-Hill, New York, 1968.

[83] Pollard.S.B., Mayhew.J.E.W., and Frisby.J.P. PMF: A stereo correspondence algorithm based using a disparity gradient limit. *Perception*, 14:449–470, 1985.

[84] Porrill.J., Pollard.B., Pridmore.T.P., Bowen.J.B., Mayhew.J.E.W., and Frisby.J.P. Tina: A 3d vision system for pick and place. *Proc. 3rd Alvey Vision Conf.*, pages 65–72, 1987.

[85] Richards.W. Anomalous stereoscopic depth perception. *J.Opt.Soc.Am*, 61:410–414, 1971.

[86] Richter.J. and Ullman.S. Non-linearities in cortical simple cells and the possible detection of zero-crossings. *Biol.Cyb.*, 53:195–202, 1986.

[87] Rogers.B and Koenderink.J.J. Monocular aniseikonia: A motion parallax analogue of the disparity induced effect. *Nature*, 322:62–63, 1986.

[88] Rogers.B.J. and Graham.M. Anisotropies in the perception of three-dimensional surfaces. *Science*, 221:1409–1411, 1983.

[89] Rozenfield.A and Kak.A.C. *Digital Picture Processing 2nd.Ed.* Academic Press, 1982.

[90] Sanger.T.D. Stereo disparity computation using Gabor filters. *Biol. Cybernetics*, pages 405–418, 1988.

[91] Schwartz.M. *Information Transmission, Modulation and Noise.* McGraw-Hill, 3rd Ed., 1987.

[92] Srinivasan.R. The significance of the phase synthesis. *Proc. Indian Acad. Sci.*, 53:252–261, 1961.

[93] Terzopolous.D., Witkin.D., and Kass.M. Energy constraints on deformable models: Recovering shape and non-rigid motion. *Proc. AAAI-87 conf. Seattle*, 1987.

[94] Trivedi.H.P. and Lloyd.S.A. The role of disparity gradient in stereo vision. *Perception*, 14:685–690, 1985.

[95] Viterbi.A.J. Phase-locked loop dynamics in the presence of noise by fokker-plank techniques. *Proc.IEEE*, pages 1737–1753, 1963.

[96] Wheatstone.C. On some remarkable and hitherto unresolved phenomina of binocular vision. *Phil.Trans.R.Soc.Lond*, 2:371–394, 1838.

[97] Willsky.A.S. Fourier series and estimation on the circle with applications to synchronous communication. part 1:analysis. *IEEE Trans.Info.Theory*, IT-20,5, 1974.

[98] Wilson.R.G. and Granlund.G. The uncertainty principle in image processing. *I.E.E.E trans. on Pattern Analysis and Machine Intelligence*, pages 758-767, 1984.

[99] Wilson.R.G. and Knutsson.H. Uncertainty and inference in the visual system. *IEEE trans. Sys., Man and Cyb.*, 18,2:157-163, 1988.

[100] Wilson.R.G. and Knutsson.H. A multiresolution stereopsis algorithm based on the Gabor representation. *Proc. 3rd IEE Conf. on Image Processing.*, pages 19-22, 1989.

TITLE Phase Relationships in Stereoscopic Computation

AUTHOR Keith Langley

INSTITUTION and DATE University of Warwick 1990

CAMERA 8