COMPUTING SHAPE
USING A THEORY OF HUMAN STEREO VISION

by

William Eric Leifur Grimson

B.Sc., University of Regina
(1975)

Submitted in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology
June 1980

Signature of Author _____
Department of Mathematics
June 2, 1980

Certified by _____
David C. Marr
Thesis Supervisor

Accepted by _____
Michael Artin
Chairman, Departmental Committee on Graduate Students

COMPUTING SHAPE
USING A THEORY OF HUMAN STEREO VISION

by
William Eric Leifur Grimson

Submitted to the Department of Mathematics
on June 2, 1980 in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

Abstract

This thesis is concerned with the construction of a complete specification of the shapes of the visible surfaces of a scene, given a stereoscopic pair of images of that scene as input. This problem is investigated within the framework of a computational theory. There are two parts to the problem.

A theory of human stereo vision was recently developed by Marr and Poggio (1979). Part I of this thesis deals with a computer implementation of that theory. The implementation served both as a source of feedback for the theory itself, outlining problems and illuminating previously overlooked difficulties, and as a method of testing the adequacy of the theory. By running the program on a series of images from the human stereopsis literature and comparing the results with the perceptions of stereo observers, it was shown that the program, as a reflection of the Marr-Poggio theory, adequately accounted for a wide range of stereoscopic perception.

Part II deals with a computational theory of surface interpolation. One of the consequences of the Marr-Poggio theory, and of most early visual processes, is that explicit surface information is available only at a certain restricted set of positions in the image. In order to construct a complete specification of the surface shapes, it is necessary to account for implicit information in the formation of the images. A theory of this process was developed and implemented. In particular, a series of results relating the shapes of surfaces and the initial representation of the images were proven. From these results, a set of mathematical constraints on the interpolation problem were derived. An algorithm for interpolating the surfaces was developed, using results from nonlinear programming, and that algorithm was implemented in a computer program and tested on a range of images.

Thesis Supervisor: David C. Marr
Title: Professor of Psychology

Acknowledgements

Many people have contributed to this work and provided advice and encouragement to its author. I would like to express my gratitude to:

David Marr, whose constant and timely encouragement and enthusiasm and whose insightful advice were invaluable, whose high standards are a constant inspiration, and without whom, this work would not have happened.

Tommy Poggio, who, together with David Marr, developed the stereo theory upon which this thesis relies, who is never stumped for an answer, and whose excitement for scientific discovery is contagious.

Mike Brady, who carefully read drafts of the thesis, and who provided valuable criticism on many points; Berthold Horn, who also provided valuable criticism on the thesis, and who frequently enlightened me on the details of image formation; and Whitman Richards, who enlightened me on many aspects of psychophysics.

Patrick Winston, who exhibited unending enthusiasm about this project, who provided financial support, and who carefully maintains an atmosphere within the AI Lab that is both stimulating and supportive.

Shimon Ullman, Keith Nishihara and the other "Visionaries", who provided many useful comments and criticisms, and who are excellent "sounding boards" for ideas, both good and bad.

Denis Hanson, who convinced me that even a "prairie boy" could make it at MIT.

Matt, Mary, Tomas, Lorraine, Chuck, Candy, Bob Sj., Bob W., Brian, Ken and John, who help make the AI Lab an exciting place to be; with whom I've had many discussions, and without whom, the time spent in completing this work would have been far less enjoyable.

Ellen Hildreth, who read more drafts of this thesis than I did, who provided many valuable comments, and who helped me keep my sanity throughout it all.

and my parents, who taught me long ago that few things are more valuable and lasting than a thirst for knowledge.

Table of Contents

CHAPTER 1

INTRODUCTION

Although the world in which we exist is three-dimensional, the projection of light rays onto the surface of the retina presents our visual system with an image of the world which is inherently two-dimensional. Yet we must be able to navigate within this three-dimensional world, and in fact we do so without difficulty. One of the requirements of visual processing is then to reconstruct a three-dimensional representation of the world, from its two-dimensional projection onto our eyes.

There are several sources of information in the retinal images which can be used for this three-dimensional reconstruction. For example, one can use shading information (Horn, 1970, 1975; Ikeuchi, 1979), the motion of objects over time (Ullman, 1979a), surface contours (Stevens, 1979), texture gradients (Stevens, 1979; Kender, 1978, 1979; Bajcsy and Leiberman, 1976; Ikeuchi, 1980), focusing (Horn, 1968), or stereo vision (Marr and Poggio, 1979). All of these processes can be viewed as transforming representations of the images into representations of the surface shapes.

Most of these processes, especially as performed by the human visual system, consist of two stages. In the first stage, explicit surface information is computed at particular points in the image. In the second stage, the surface is interpolated between these known points to obtain a complete specification of its shape. In the first part of this thesis, the particular method of stereo vision is examined as a method for computing surface information at a sparse set of points in the image. In the second part of the thesis, the general problem of surface interpolation is considered. This problem is investigated independent of the particular process used to compute the initial surface information,

and is valid for surface interpolation as applied to many visual processes, such as stereo, structure from motion, and structure from surface contours.

The initial motivation for this study of visual processes arose from a desire to understand and model the human visual system. As a result, as much as possible, our theory is designed to be consistent with known evidence about that system, concerning its structure and its input-output behavior. Aspects of the method by which the human visual system analyzes information can be characterized as belonging to one of two categories:

(1) aspects relevant to the processing of visual information by any general visual processor, and

(2) aspects specific to the demands of implementation in a biological system (for example, the fact that information must be carried by neurons which can, in general, interconnect to a number of other neurons).

Although the human system forms the basis for our study, it is the first category which is of primary interest to us. Certainly, if an accurate model of the human visual system can be formed, it will provide a method for solving the visual problem in general situations. At the same time, insight may be gained into the processing of the human visual system — a goal of considerable value and interest in its own right. However the human system should serve as a tool for understanding the general processing involved in computing surface descriptions, without digressing into details of a specific neural model for such processing in the human system. Over the past few years, an approach to the study of vision has emerged, which makes a clear distinction between aspects of visual processing relevant to any visual processor, and those specific to the hardware in which the process has been implemented. This is the computational approach of Marr (1976a, 1976b, 1980, also Marr and Poggio 1977a). It is this approach which will be examined in the next section, and which will be utilized in this thesis.

1.1 The Computational Approach

The computational approach views the human visual system as performing computations over internal symbolic representations of visual information. Marr (1978) and Marr and Nishihara (1978) argue for at least three such representations in the course of visual processing:

(1) The Primal Sketch, in which properties of the intensity changes in the image are made explicit,

(2) The $2\frac{1}{2}$-D Sketch, which describes properties of the visible surfaces for every point in the image, and

(3) The 3D Model, which now makes explicit the three-dimensional shape of objects in the scene, in object-centered coordinates.

An important property of these representations is that each makes explicit information which it is possible to obtain from the previous description of the image, and which is useful for the construction of the next representation. That is, directly from an image, one can describe places where intensity changes which will be useful for processes such as stereo and texture analysis for obtaining descriptions of surfaces, and so on. Two criteria for judging our representation of surfaces obtained through stereo vision will be: first, the computability of this description, and second, its suitability for higher level processing (see Marr and Nishihara for application of these criteria for judging shape representations).

Critical to the computational approach is the distinction between several levels of description of a process. Since we are dealing with the manipulation of symbolic descriptions, we can distinguish between the meaning of the symbols and the physical embodiment of those symbols. In other words, one can study the computation performed by the system (almost) independent of the mechanisms which actually perform the computation. In the computational approach, we study the visual system at three different levels: the computational theory, the algorithm, and the underlying implementation of the computation (Marr and Poggio, 1977a).

Important at the level of the computational theory are the physical constraints that restrict the problem sufficiently to allow the process to do what it does. In general, the problems faced by modules of early visual processing appear to be insoluble if one attempts to solve them from the image alone. Ullman's (1979a) rigidity assumption in the interpretation of three-dimensional structure from motion, Marr and Hildreth's (1979) condition of linear variation and spatial coincidence assumption in the analysis of intensity changes and Marr and Poggio's (1979) assumptions of uniqueness and continuity are examples of such physical constraints restricting the problem at hand. The critical step in the formulation of the computational theory is the enunciation of these additional constraints on the process, that are imposed naturally as a consequence of the way the world is made, and which constrain the result sufficiently to allow a unique solution to be found.

Once the additional information has been isolated, one can incorporate it into the design of a process. There are a number of ways in which the process may utilize a constraint. It may be transformed into an assumption, which is taken always to be true, with or without verification (optical illusions often illustrate situations in which these assumptions break down). An example of this is the case of linear variation (Marr and Hildreth, 1979). The process might explicitly "look for" the satisfaction of the constraint, and if it is consistent with visual input, then assume the constraint to be true. An example of this is the case of rigidity (Ullman, 1979a). Alternatively, the constraint may be explicitly embedded in the process, such as in the continuity and uniqueness assumptions in stereo (Marr and Poggio, 1979).

I stated earlier that visual processing will be viewed as transforming visual information from one representation to another. All the processes of early vision take as their input properties of the image and produce as their output properties of surfaces - either relating to their geometry or their reflectance. It will be seen that in the stereo process, it is important to determine what kind of representation of the image will form the input to the process, by what means will the process transform this information into a representation of surfaces, and what is the nature of this surface representation.

Although these questions can be addressed for visual processing in general, it is desirable to have the theory be consistent with processing in the human system. Thus, psychophysical evidence concerning the nature of these representations, and the processes by which they are transformed, will be crucial for answering these questions.

A major assumption is being made at the level of the computational theory: that the human visual system is an inherently modular system, allowing us, for example, to study the process of stereo vision in isolation. At first glance, it is not at all clear how separate stereoscopic processing is from the monocular analysis of each image. If stereo processing were an isolated module, then one could study stereo independent of other visual processing, such as texture analysis, shading, and so on. One method for testing whether a process can be studied in isolation is to present the visual system with images in which, as far as possible, all kinds of information except one have been removed, and then seeing whether we can make use of just that one. For stereo this can be demonstrated by the random-dot stereogram, invented by Julesz (1960). Each of the images in Figure 1.1 is a collection of black and white squares, which are identical except for the fact that a centrally located square-shaped region is

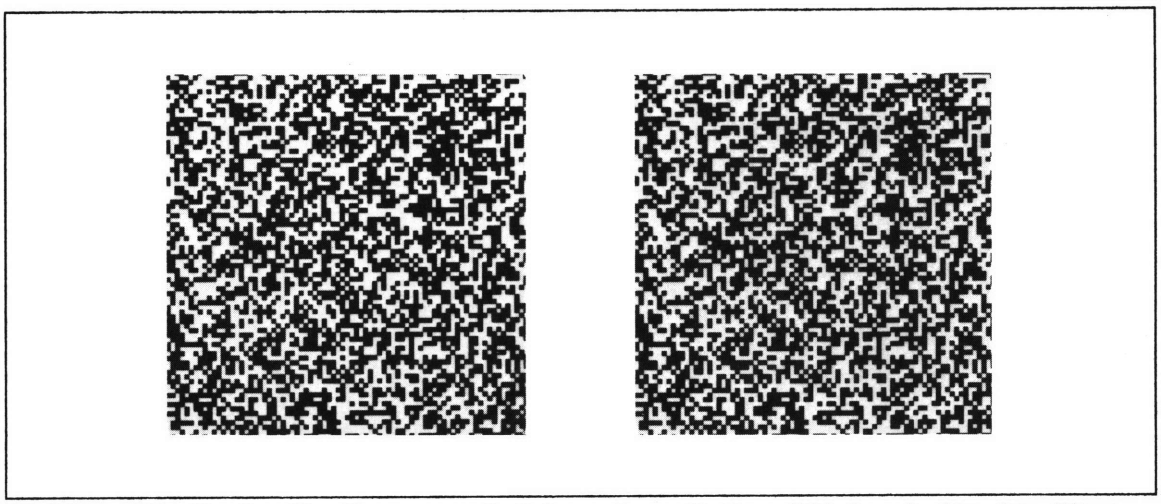


Figure 1.1. Random Dot Stereogram. Each image is a collection of black and white dots. When viewed stereoscopically, a central square is perceived as separated in depth from the rest of the pattern, although each monocular image contains no cue to this effect.

shifted horizontally in one image relative to the other. Other than this disparity, the images contain no information whatsoever about the visible surfaces. Yet, when the pair is viewed stereoscopically and fused, one clearly and vividly perceives a square floating in space above the plane of the background. This illustrates that disparity alone can cause the sensation of depth. The fact that neither image contains any recognizable monocular organization implies, as well, that the stereo process may be studied in relative isolation from other visual processes.

The idea that a large computation should be split up and implemented as a collection of small sub-parts that are as nearly independent of one another as the overall task allows, is what Marr (1976b, 1980) calls the *principle of modular design* and forms a cornerstone to the approach. Its importance lies in the fact that otherwise, a small change in one place in a process will have consequences in many other places. This means that the process as a whole becomes extremely cumbersome, and difficult to debug and analyze.

Having considered a computational theory of the processing involved in a visual task, one can then turn to the design of a particular algorithm to achieve the task. One is ultimately interested in the algorithm used by the human visual system. However, a second purpose for studying an algorithm is that it serves as an excellent source of feedback for the computational theory. Any implementation of a theory by a particular algorithm helps to point out otherwise unnoticed difficulties with the task, as

well as helping to indicate the sufficiency of the theory. Furthermore, any assumptions made by the theory can frequently be tested by an implementation of the theory. Chapter 4 of this thesis will detail how this particular implementation of a theory of human stereo vision helped refine the theory both through the act of implementation and through the use of the implementation on trial data.

Marr (1976b, 1980) outlines two principles which one can use to guide the design of algorithms, and which probably ought to be satisfied by any serious candidate for an early visual process in the human visual system. The first one is the *principle of graceful degradation* and says that whenever possible, degrading the data will not prevent one from delivering at least some of the answer. The second principle is the *principle of least commitment* and says that one should never do something that may later have to be undone.

It is important to note that for any computational theory, there may be several possible algorithms for solving it. In many cases, one can distinguish between the acceptability of different algorithms. For models of the human visual system, we shall take a set of algorithmic criteria, outlined in Chapter 7, as one set of criteria for algorithmic acceptability. That is, we shall seek algorithms which seem biologically feasible, (Ullman, 1979b).

The third level of description is that of the implementation. We are ultimately interested in understanding the neural implementation used by the human system. It would be nice to be able to give general rules about processes at the level of the neural implementation; unfortunately, with only a few theories developed to the point where specific neural implementations have been proposed (for example, Marr 1969), and none having been confirmed experimentally in every detail, it is not yet possible to formulate such rules.

1.2 Synopsis

The thesis is divided into two major parts, corresponding to the two stages involved in transforming images into surface representations.

In the first part, a specific method for computing surface information at particular points in the image is considered. In particular, an implementation of the Marr-Poggio theory of human stereo vision (1979) is described.

In Chapter 2, we consider what is involved in the stereo computation. The essence of the stereo problem is to select an element from one image, find the corresponding element from the other

image, measure the disparity between them, and use this disparity to compute the distance from the viewer to the element in the scene. The two basic questions are what are the elements to be matched, and how are these elements matched. It has been argued that the image elements to be matched must be in one-to-one correspondence with well-defined locations on a physical surface in the scene (Marr, 1974; Marr and Poggio, 1979). The Marr-Hildreth theory of edge detection (1979) proposes that such elements are obtained by filtering the image with a collection of different sized operators. The form of these operators is $\nabla^2 G$, the Laplacian of a Gaussian, and the zero-crossings of their output correspond to those locations in an image which can be matched by the stereo process. This determines the transformation of the image intensities into the input representation required by the stereo process.

The second problem to solve is the question of how to match these elements. The Marr-Poggio stereo theory (1979) applies two constraints — uniqueness and continuity — to the problem. These constraints, together with evidence from psychophysics and neurophysiology, led to the development of a specific algorithm which is discussed in Chapter 2.

At this point, we have dealt with the computational theory of the first two steps in our process. One of the important steps in the design of a computational theory is the implementation of that theory. In Chapter 3, the implementation of the Marr-Poggio theory of stereo vision is outlined. This consists of explicitly specifying the implementation of each of the steps in the theory.

In Chapter 4, the implementation is used to test the adequacy of the theory. To do this, the performance of the implementation is compared to human perception over a wide range of test cases from the human stereopsis literature. The performance of the algorithm on natural images is also illustrated.

The matching component of the Marr-Poggio theory is based in part on an analysis of the statistical distribution of zero-crossings. The assumptions used by Marr and Poggio in performing this analysis are not entirely valid, and a correct reworking of the statistical arguments is performed. These statistics have been evaluated for actual images and are compared with the predictions made by the analysis. This allows us to consider the relevance of the consequent assumptions made by the theory.

During the development of the stereo implementation, a number of observations were made concerning aspects of the theory. These are outlined and their effect on refinements of the theory

are discussed. Included are discussions on the type of representation used by the $2\frac{1}{2}$-D sketch, the accuracy of the disparity values output by the stereo program, the symmetry of the matching function, the use of oriented versus non-oriented filters, and others.

The second part of the thesis considers the problem of surface interpolation — of transforming the sparse surface representation into a complete one. This problem is relevant to many visual processes other than stereo, for example, structure from motion and structure from surface contours. Hence, the problem is considered in the general case, independent of the method used to obtain the input representation. The only assumption applied to the problem is that the points at which explicit surface values are known correspond to the zero-crossings obtained from the convolved images.

In Chapter 5, the implicit information about surface shape contained in the zero-crossing descriptions of the images is considered. Horn's image formation equation (1970, 1975) is developed and used to prove a set of theorems relating the shape of a surface to the zero-crossings detected by the Marr-Hildreth operators. The major point about these theorems is that they describe the probability of a surface being inconsistent with what is known about the form of the image intensities, specifically, about the positions of local changes in image intensity.

In Chapter 6, this set of theorems is used to develop a computational theory of surface interpolation. Since the information available about the image formation process is insufficient to exactly reconstruct the original surface, the best that can be done is assign some measure of the probable inconsistency of a surface to each possible surface. In this sense, one would like to compare surfaces, in order to determine the least consistent one. The standard method of doing this is to assign some real number to each surface, by applying a functional to the surface. Then, to compare two surfaces, one need only compare the corresponding real numbers obtained by the functional. In this way, a measure of the inconsistency of any particular surface can be obtained. Thus, to find the least consistent surface, one need only derive a method for choosing the surface which minimizes this functional which measures surface inconsistency.

There are two aspects to this problem. The first is to determine under what conditions the problem is well-defined, in the sense of having a unique solution. A set of simple theorems is used to determine the conditions on the functional under which this will be true. The second is to determine the actual form of the functional.

Several possibilities for the form of this functional are considered, seeking one which satisfies the conditions developed above. One candidate for choosing the least inconsistent surface is given by:

$$\Theta(f) = \left\{ \int \int f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2 \, dx dy \right\}^{\frac{1}{2}}.$$

This functional satisfies all the conditions necessary to guarantee a unique solution, and satisfies the conditions imposed by the theorems of Chapter 5.

This provides a computational theory of the second stage in our process, the creation of complete surface specifications.

In Chapter 7, an algorithm for solving this computational problem, independent of the particular form of the functional, is developed. Although there are several possible algorithms for solving the computational problem, the methods of nonlinear programming are used, because such methods are more biologically feasible, (Ullman 1979b). In particular, the Kuhn-Tucker theorem relates the solution of a constrained optimization problem to the saddle-point of an associated Lagrangian. To find the saddle-point, the Arrow-Hurwicz system of difference equations can be applied. These results are used to create an algorithm for interpolating the surfaces.

In Chapter 8, the implementation of the interpolation algorithm is considered. The biological feasibility of the algorithm is discussed and the performance of the algorithm on both synthetic and real depth maps is examined.

Finally, in Chapter 9, the interpolation algorithm is analyzed and possible refinements to the theory are discussed. These include the detection of surface discontinuities, and the use of such discontinuities to refine the surface approximation; the removal of noisy disparity values from the depth map; the role of acuity in determining accurate depth values, and the effect of acuity on the interpolated surfaces; and the role of retinal mappings in the stereo matching problem.

CHAPTER 2

THE STEREO THEORY

2.1 What Is Stereo

If two objects are separated in depth from a viewer, then the relative positions of their images will differ in the two eyes. This is illustrated in Figure 2.1. The process of stereo vision, in essence, measures this difference in the relative positions and uses it to compute depth information about the objects in the scene.

To avoid confusion, three terms relating to stereopsis are defined here. The term *disparity* will refer to the angular difference in position of the image element in the two eyes. *Distance* will refer to the objective physical distance from the viewer to the object, usually measured from one of the two eyes. Finally, the term *depth* will refer to the subjective distance to the object as perceived by the viewer.

It can be seen from Figure 2.1 that disparity varies as the relative depth positions of objects varies. Suppose the eyes are fixating at a particular point, such as that indicated by the circle in the figure. Disparities such as that associated with the crosses in the figure (denoted by d_3 in the figure) are usually referred to as crossed or convergent disparities. Those, like the one associated with the ellipse (denoted by d_1 in the figure) are referred to as uncrossed or divergent disparities. Objects at the same depth as the fixation point (such as with disparity d_2 in the figure) are said to have zero disparity. It can also be seen that as an object moves closer to the viewer than the distance to d_2, the disparity

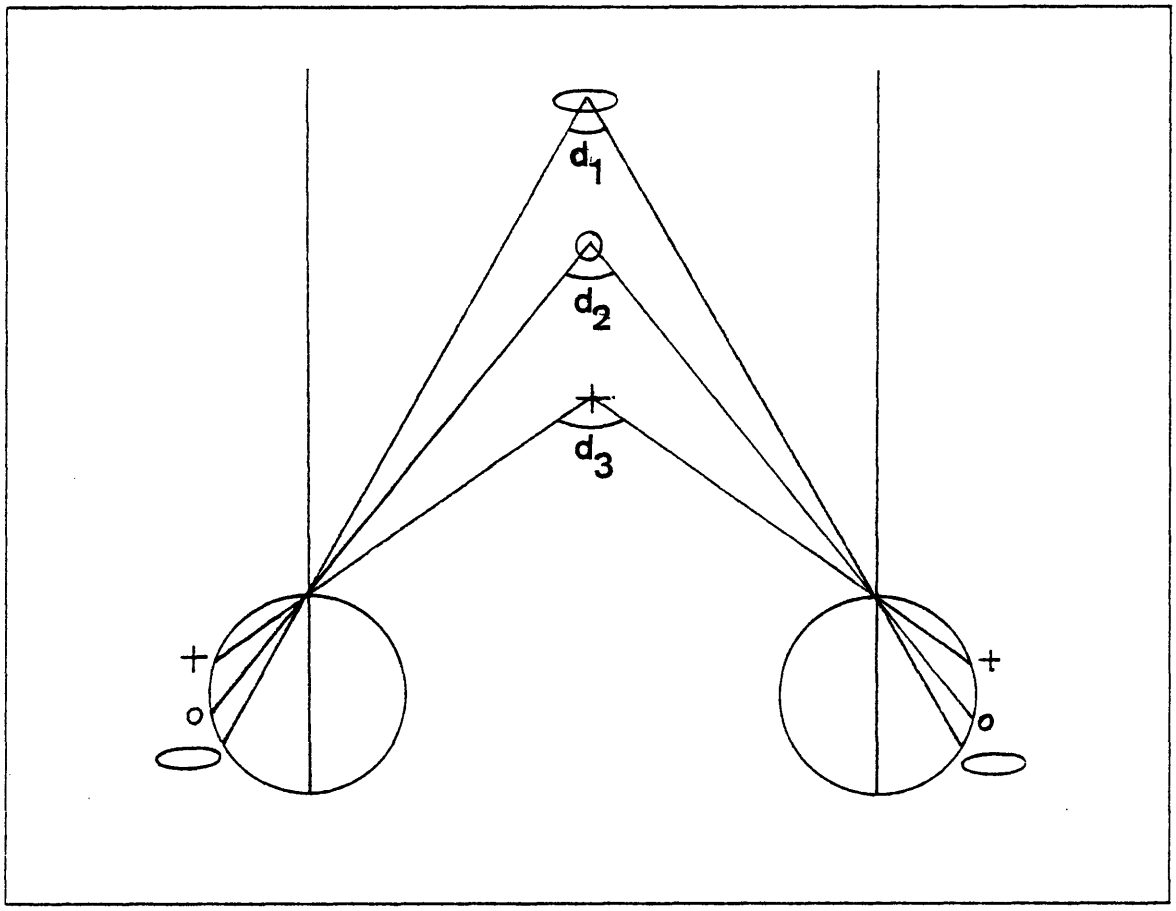


Figure 2.1. Definition of Disparity. The disparity of a point refers to the angular difference in position of the image element in the two eyes. Given a fixation point, such as the circle in the figure, disparities may be crossed (d_3), uncrossed (d_1), or zero (d_2).

increases in magnitude.

This relationship between distance and disparity suggests that one could determine the distance to objects in a scene, by measuring their disparity. The stereo system is concerned with exactly this process.

How does one measure this disparity? Marr (1974) notes that there are three steps involved:

(1) A particular location on a surface in the scene is selected from one image;

(2) The same location is found in the other image; and

(3) The disparity between the two corresponding image points is measured.

The actual computation of distance, once the two image points are identified, involves a simple

geometric transformation, and hence does not pose any major difficulties. If a location could be identified beyond doubt in the two images, then steps (1) and (2) could be avoided and the whole process would be simple. For instance, if a location in the scene could be illuminated with a spot of light, the identification of the corresponding image locations would be simple. That is, by appropriately choosing what locations in an image are to be matched, the actual matching or correspondence can be greatly simplified. Unfortunately, we are not equipped with lasers in our foreheads, and hence cannot perform depth computations based on the matching of spots of light So we need a more passive method for sensing the environment.

Thus, the problem of the identification of corresponding image points cannot be avoided, and this problem forms the heart of the stereo computation. There are two aspects to this problem: what locations in the image are selected to be matched; and by what process are corresponding locations from the two views actually matched.

A major problem for theories of stereo vision is the question of what elements are matched in the two views. Are they intensities? objects? something else? The answer to this question will have a strong influence on the second problem of how to match these elements. In fact, there is a trade-off between the two aspects of the problem: between the complexity of the monocular analysis used to extract the elements to be matched, and the complexity of the process which matches them.

At one end of the spectrum, the problem of finding the corresponding locations can be greatly simplified, at the expense of determining which locations to match in the first place. If the number of potentially matching elements is kept small in each view, the problem of finding corresponding elements becomes simple. One method for accomplishing this is to perform object recognition before stereo. Thus one would first process each monocular view, identifying the various objects contained therein. Since in most scenes, the probability of having two or more "identical" objects is small, the matching problem is simple. To determine the disparity, a particular object, say a desk, would be located in one image, the same desk would be found in the other image, and the disparity would be computed. The solution of the correspondence problem — which element in one view matches which element in the other — is, however, performed here at the expense of identifying and locating objects in a scene, a non-trivial task. Under such a scheme, the stereo process would be a later visual process, following earlier processes which would analyze each monocular image. Of course, we have already

seen that this is not the case for the human stereo system. The example of Julesz random-dot patterns showed that the process occurs very early in the visual process and is relatively independent of other forms of visual processing.

At the other end of the spectrum, the correspondence problem is more difficult, while the problem of determining which elements to match is simplified. For example, if the human stereo process is an early one, then much more primitive matching elements could be used; for example, local descriptors of sudden changes in image intensities. If the tokens to be matched arise at a very early stage in the analysis of an image, then the density of such objects in the images is significant. In this case, the number of potentially matching elements will increase. Furthermore, the difference between the possible tokens to be matched is very small. Thus, for any single descriptor of the left image, there is likely to be a large number of possible matching descriptors in the right image. All but one of these will be incorrect, and thus the correspondence problem is non-trivial, although the identification of elements to be matched is fairly simple.

This introduces the major problem of stereopsis. The task of identifying corresponding locations in the two images is a difficult one because of what is called the *false targets problem.* An example of it is shown in Figure 2.2 (after Marr and Poggio, 1979). The question is, to which dot viewed from the left eye do dots viewed from the right eye correspond? Each eye sees four dots, but which are the correct matches? *A priori,* any of the 16 possible matches is a plausible candidate, but in fact when we observe such a stereo pair, we make the correspondence shown with the filled circles, and not any of the correspondences shown with open circles. These alternate candidates are therefore called false targets.

This is somewhat surprising, since there seems to be little reason to distinguish one match as more favorable than another on the basis of the elements being matched. Moreover, as Marr observes, there is another solution to this particular correspondence problem which seems just as "valid". This is the four central vertical matches, in which R1 is paired with L4, R2 with L3, R3 with L2, and R4 with L1. But we never "see" this match perceptually, as a set of circles in a line receding from us. Why?

To answer this question, additional information is needed to help to decide which matchings are correct, by constraining them in some manner. The only way to do this is to examine the basis in the

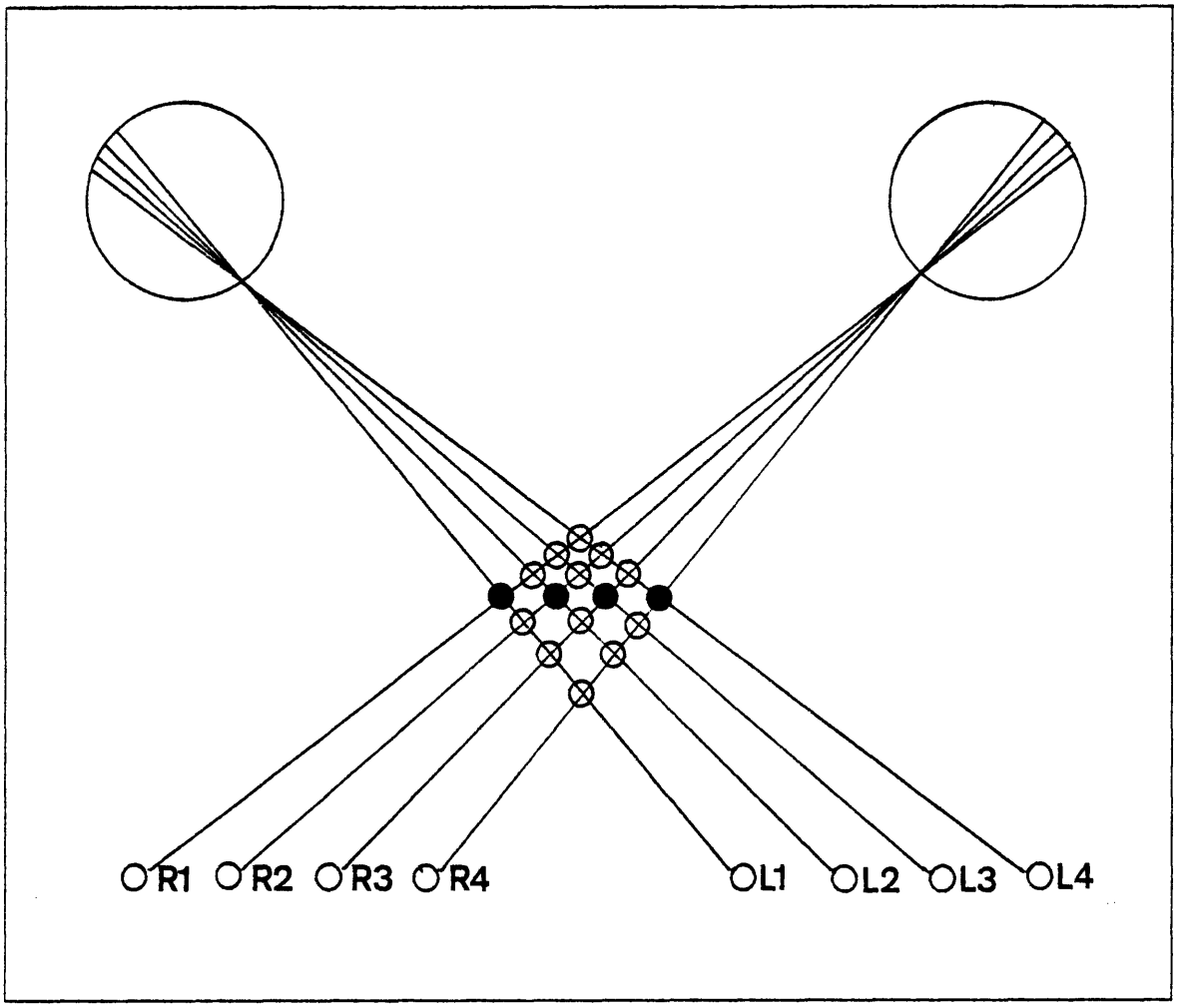


Figure 2.2. False targets problem. Of the 16 possible matches indicated by the circles, only those indicated by the filled circles are actually perceived.

physical world for making a correspondence between two images. Marr and Poggio (1979) (also Marr 1974) describe two physical constraints which are relevant to the stereo process:

(1) A given point on a physical surface has a unique position in space at any one time.

(2) Matter is cohesive, it is separated into objects, and the surfaces of objects are generally such that the changes in the surface are very small compared with their distance from the viewer.

Let us briefly return to the example of Figure 2.2. Since, in principle, any one of the 16 possible candidates could be a match, there are 2^{16} or 65, 536 possible matches for this situation. The first constraint essentially states that in general, excluding accidental alignment of objects, there should not be more than one match along any line of sight from either eye. The number of such matches

is greatly reduced; from 65,536 to 24. The second constraint essentially states that, barring other information, the most likely of these 24 matches is the one actually chosen in Figure 2.2. This is because this match corresponds to the "smoothest" object (relative to the other 23 matches).

The major point about the observations of Marr and Poggio is that they are properties of physical surfaces, and they constrain the way in which positions on the surface behave. If they are to be used to deal with the correspondence between two images, we must ensure that the items to which they apply are in one-to-one correspondence with well-defined locations on a physical surface. We shall turn to this point in the next section.

Thus, the computational problem for the stereo process may be summarized as follows. First, for each view of the scene, construct a description of the elements that are to be matched. We have seen that for the human system, the descriptors must be extracted at an early stage of visual processing. As a consequence, each description may contain a large number of identical elements, which makes the correspondence problem, and the removal of false targets, a significant problem. The second step consists of solving this correspondence problem, determining which descriptor from one view matches which descriptor from the other view.

2.2 The Marr-Hildreth Theory of Edge Detection

We have argued from the evidence of random dot patterns that the extraction of a description of the scene, from which the matching may take place, must be, at least in part, an early visual process, occurring prior to object recognition. That is, while the evidence of random dot stereograms does not imply that only early analysis is done, it does imply that stereo vision must be capable of operating on primitive descriptions. The creation of a description of surface locations suitable for matching between the images is somewhat difficult. As we have seen, if a surface location could be chosen in some absolute way — such as shining a narrow beam of light onto it — the problem would be simple. However, we are not able to do this. Thus the problem of identifying surface locations from their image projections must be addressed. The difficulty is precisely that assertions about surface locations must be made based solely on information in the image.

It will not be possible to unambiguously identify surface locations from the images themselves at all points in the images. A smooth featureless surface will be of no avail to this process, because the image intensities will be indistinguishable over this surface. Any surface containing scratches, texture

or other markings will, however, give rise to locally sharp changes in reflectance which can be used to define surface points. Since only certain features of a surface will be well defined in the images, it is claimed (Marr 1974, Marr and Poggio 1979) that the computation of disparity takes place by comparing symbolic descriptions of those features in the two images.

The solution to the first problem, selecting a surface location, will be solved by creating a symbolic description of those features of the surface which give rise to identifiable physical locations. The design of a module to match the symbolic descriptions, by determining which descriptors correspond to the same physical location, will solve the correspondence problem.

It has been remarked that the features of an object of likely interest to the matcher which solves the correspondence problem are associated with locally sharp changes in reflectance, and hence with locally sharp changes in the image intensities. It can be argued (Marr, 1976b) that such a early symbolic description of intensity changes in an image is useful for other early visual processes besides stereopsis. This description has been called the *Primal Sketch* (Marr, 1976b) and was used, for example, in the solution of the motion correspondence problem (Ullman 1979). Recently, (Marr and Hildreth, 1979) a theory of the process of extracting these low-level primal sketch descriptions has been developed. This theory has developed in part due to findings concerning the human visual system, and in part due to the development of the stereo theory dealt with in this thesis. The relevant points of the Marr-Hildreth theory of edge detection are outlined below.

It would be pleasant if one could design a single filter to extract all the features of interest from a scene. Unfortunately, such intensity changes take place over a wide range of scales (Marr, 1976b). For example, if we look at individual picture elements (pixels), we find intensity changing from pixel to pixel. However, such changes are occuring at too small a scale to be of interest to us. Most edges in the real world are sharp edges, and the associated intensity function will be composed of a few steep changes over a small number of pixels. At the same time, other edges, such as shading edges, are spatially extended, and the intensity function will increase slowly over a large number of pixels. Moreover, these different types of intensity changes are not distinct in the image; one can have high contrast edges superimposed on a spatially extended shading edge.

As a consequence of this wide range of intensity changes, one cannot hope to find a filter which will be simultaneously optimal at all scales. The findings of Campbell and Robson (1968),

concerning the existence of separate spatial-frequency channels in the human visual system, suggest that one should seek a method of dealing separately with the changes occurring at different scales. As a consequence, Marr and Hildreth suggest that one first take some local average of intensity at several resolutions, and then detect the changes in intensity that occur at each resolution. They determine both the optimal smoothing filter and a method for detecting intensity changes at a given scale.

The optimal smoothing filter must satisfy two physical constraints. First, the filter is intended to reduce the range of scales over which intensity changes take place. This suggests that the filter's spectrum should be band-limited. Second, the aspects of an object which give rise to intensity changes are all spatially localized. This suggests that the filter should perform an average over a small localized portion of the image. These two requirements are in conflict and are related by a type of uncertainty principle. It has long been known that under these circumstances, the optimal filter for minimizing band-width in space and frequency is the Gaussian, (see for example Leipnik, 1960). Thus, the image I is initially convolved with a two dimensional Gaussian operator G. For each channel, the size of the operator G will vary. This essentially smoothes the image.

Having determined a suitable smoothing filter, Marr and Hildreth then addressed the issue of detecting intensity changes at a particular scale. Any intensity change along a particular orientation will cause an extremum in the first directional derivative, and a zero-crossing in the second directional derivative, perpendicular to the orientation of the change (Marr 1976, Marr and Poggio 1979). Thus, the task of detecting intensity changes becomes equivalent to that of detecting the zero-crossings of the second derivative of intensity in the appropriate direction. That is, intensity changes in $G * I$ are characterized by the zero-crossings in the second directional derivative $D^2(G * I)$. This operator is roughly band pass, and so it only examines a portion of the image's spectrum. By the derivative rule of convolutions, the above operator becomes $(D^2G) * I$.

The final thing left to do is to determine the direction in which the directional derivative must be taken. Although this is possible (Marr and Hildreth 1979), it may not be necessary. In fact, a number of practical considerations led Marr and Hildreth to argue that the initial operators not be directional in nature. If this is the case, then the operator to be used is the Laplacian, since it is the only non-directional linear second derivative operator. It was then shown (Appendix A, Marr and Hildreth 1979) that provided two simple conditions on the intensity function in the neighbourhood of

an edge are satisfied, the zero-crossings of the second directional derivative taken perpendicular to an edge will coincide with the zero-crossings of the Laplacian along that edge. As a consequence, Marr and Hildreth propose that at each given scale, intensity changes may be detected by searching for the zero-crossings in the convolution $\nabla^2 G * I$. This will give us precisely the symbolic descriptions of changes in the image that we required. Note that I use the term zero-crossing to refer to a non-trivial zero point in the convolved image. Thus, those positions such that the value of convolved image at this point is zero and its gradient is non-zero are taken as the primitive descriptions of the image. In this manner, the convolution values actually pass through zero, rather than simply remaining at zero. This distinguishes between an actual edge in the scene, and the zero response obtained by the operator over a region of uniform intensity.

It is interesting to note that although the use of these operators was motivated by computational and practical arguments, the use of non-oriented filters was arrived at independently on psychophysical grounds by Mayhew and Frisby (1978).

The form of the operator is given by:

$$\nabla^2 G(r,\theta) = \left[\frac{r^2 - 2\sigma^2}{\sigma^4}\right] \exp\left\{\frac{-r^2}{2\sigma^2}\right\}.$$

This is a rotationally symmetric function, and its cross-section is shown in Figure 2.3. Note that its central panel width, denoted by w, and defined as the width of the central negative region, is given by

$$w_{2-d} = 2\sqrt{2}\sigma.$$

If the visual input to the operator is a one-dimensional grating, then the response of the operator is equivalent to that obtained by applying the equivalent one-dimensional operator to the one-dimensional input. This equivalent one-dimensional operator is obtained by projecting $\nabla^2 G$ onto a line, and is given by

$$D_{xx}G = \sqrt{2\pi}\left[\frac{x^2 - \sigma^2}{\sigma^3}\right] \exp\left\{\frac{-x^2}{2\sigma^2}\right\}.$$

The central panel width of this operator is

$$w_{1-d} = 2\sigma.$$

It is illustrated in Figure 2.3.

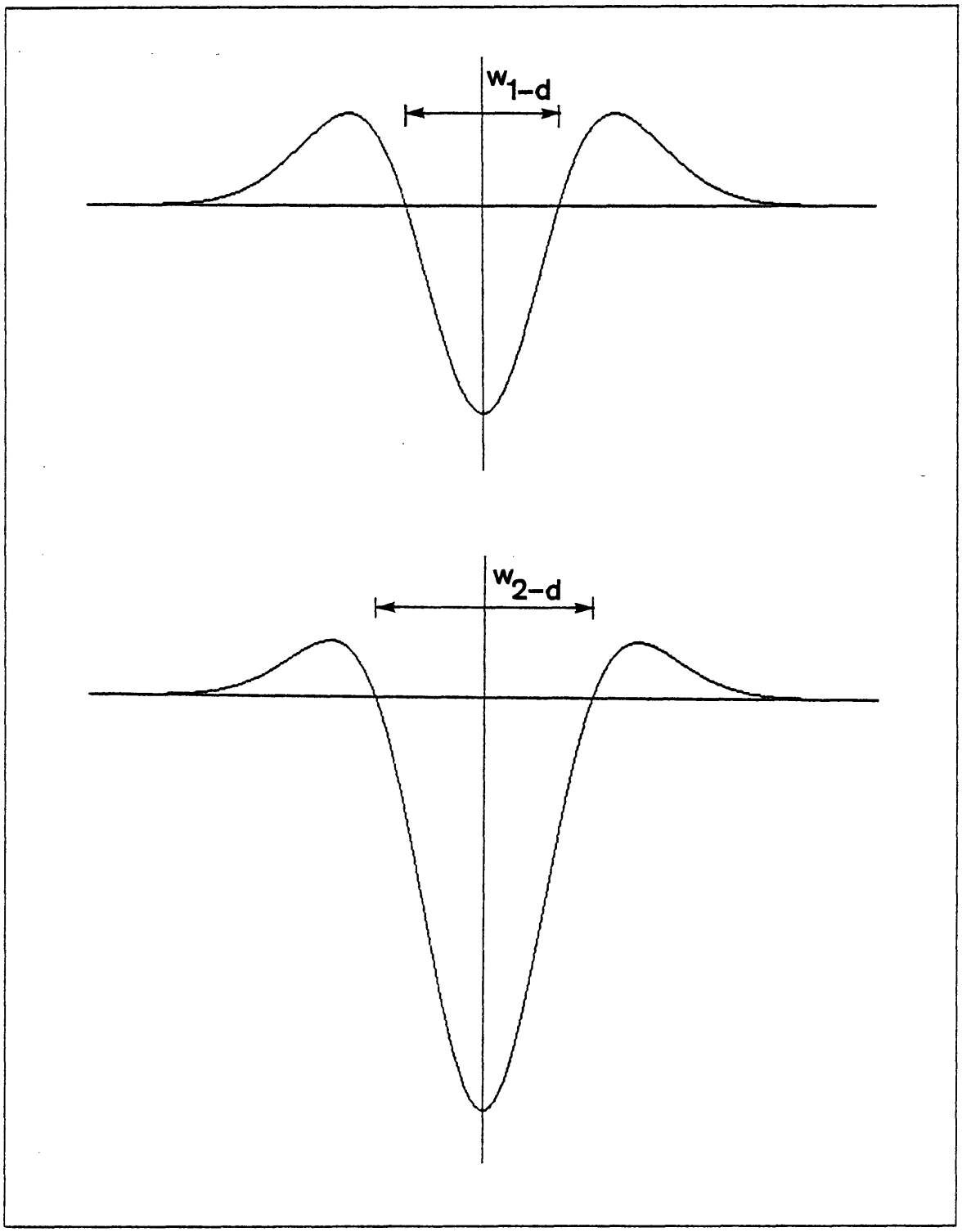


Figure 2.3. The Primal Sketch Operator. The top figure shows the one-dimensional operator. This operator applies in the case of images whose intensity values are constant along each vertical slice of the image. The bottom figure shows a cross-section of the equivalent, rotationally-symmetric, two-dimensional operator. The sizes of the operators are determined by the values of w_{1-d} and w_{2-d}.

At this point, it is interesting to note the evidence of Wilson and Giese (1977), and Wilson and Bergen (1979), concerning the existence and characteristics of such operators in the human visual system. They have found strong evidence that at each point in the visual field, there exists at least four and possibly five independent channels in the human early visual system. (This is consistent with the earlier evidence of Campbell and Robson 1968, and other investigators.) The point of interest in this context is that the form of the channels, as analyzed by Wilson and collaborators, very closely fits the shape of a difference of two Gaussians. Further, a difference of Gaussians is a close approximation to a Laplacian applied to a Gaussian (Appendix B, Marr and Hildreth, 1979). Thus, the operator for detecting features of a visual scene which was developed on basic information processing grounds is seen to be closely related to the operator which appears to be used by the human visual system. Wilson and collaborators also observe that the peak sensitivity wavelength of these channels increases linearly with retinal eccentricity, from some initial value in the fovea.

Given the form of the operators, it only remains to determine the size of these filters. Wilson and Bergen's data indicated difference of Gaussian filters whose sizes — specified by the width w of the filter's central region — range from 3.1' to 21' of visual arc. The variable w is related to the constant σ of $\nabla^2 G$ by the relation:

$$\sigma = \frac{w}{2\sqrt{2}}.$$

Wilson and Bergen's values were obtained by using oriented line stimuli. To obtain the diameter of the corresponding circularly symmetric center-surround receptive field, the values of w must be multiplied by $\sqrt{2}$. Finally, we want the resolution of the initial images to roughly represent the resolution of processing by the cones of the retina, and the size of the filters to represent the size of the retinal operators. In the most densely packed region of the human fovea, the center-to-center spacing of the cones is 2.0 to 2.3μm, corresponding to an angular spacing of 25 to 29 arc seconds (O'Brien, 1951). Accounting for the conversion of Wilson and Bergen's data, and using the figure of 27 arc seconds for the separation of the cones in the fovea, one arrives at values of w in the range 9 to 63 image elements, and hence, values of σ in the range 3 to 23 image elements.

Although Wilson and collaborators have found definite evidence only for four different sized channels, it has recently been proposed (Marr, Poggio and Hildreth, 1979) that a further, smaller channel may be present. In the human system, this channel would consist of a single retinal receptor,

although because of the diffraction in the eye, the actual size of the equivalent operator would be larger. In the implementation, since diffraction is not a factor, this channel would have a central width of $w = 1.5'$, roughly corresponding to 4 image elements.

Two other features of zero-crossings can be computed, and will be of use in the stereo matching problem. One is the sign, or contrast, of the zero-crossing. This is computed simply by noting whether the convolution values change from positive to negative or negative to positive while scanning across the zero-crossing. The second feature is the local orientation of a set of zero-crossings on the image plane. This is found by computing the gradient of the convolution values across the zero-crossing and taking the orientation of the projection of the gradient onto the image as the orientation of the zero-crossing.

Thus, the problem of what is to be matched in determining disparity in a scene has been solved. Specifically, the image will be filtered with operators whose form is a Laplacian of a Gaussian and the convolution values will be searched for zero-crossings. This will give rise to a series of zero-crossing contours, and it is this symbolic description of changes in an image which will be matched to obtain disparity information.

Some examples of the use of these operators on images are shown in Figures 2.4 and 2.5.

Figure 2.4. Examples of Convolutions. A natural image is indicated at the top. Below are examples of the convolved image, after application of different sized $\nabla^2 G$ operators, with central panel widths of 36, 18, 9 and 4 picture elements. The original image was 480 picture elements on a side.

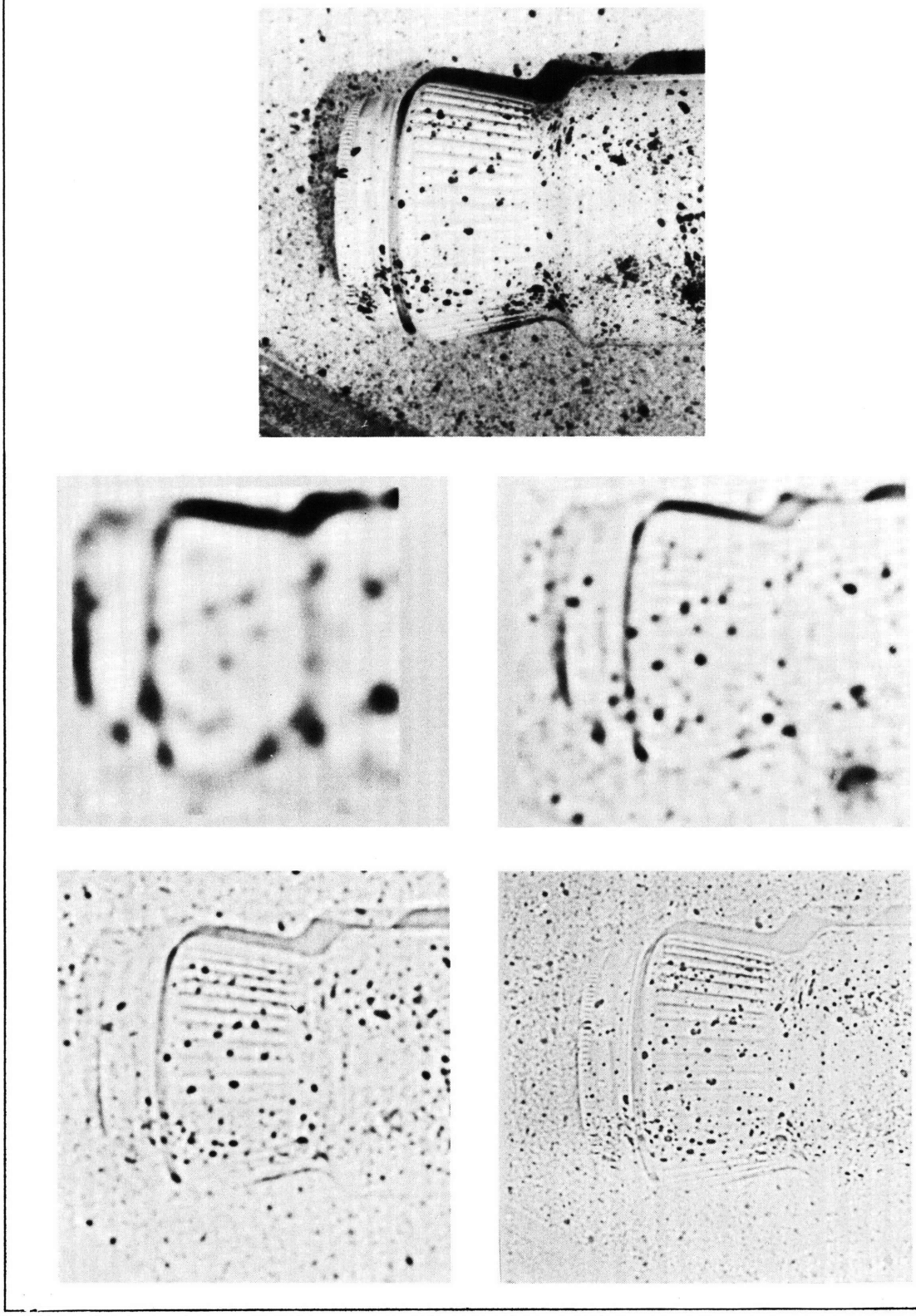
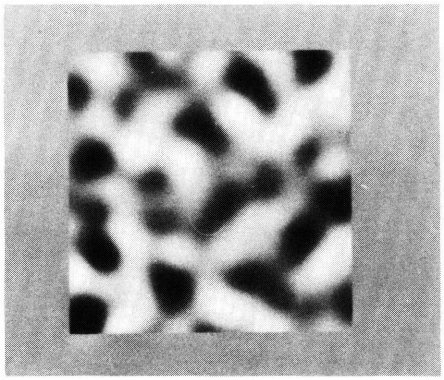
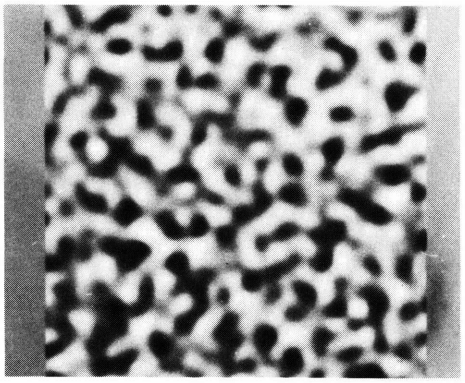


Figure 2.5. Examples of Convolutions. A random dot pattern is indicated at the top. Below are examples of the convolved image, after application of different sized $\nabla^2 G$ operators, with central panel widths of 36, 18, 9 and 4 picture elements. The original image was 320 picture elements on a side.





2.3 The Marr-Poggio Theory of Stereopsis

2.3.1 Outline of the Marr-Poggio Theory

We saw earlier that there were two major aspects to the stereo process: the extraction from the images of symbolic elements to be matched; and the actual process of matching those elements.

In the previous section, we determined the basic descriptors to be matched by the stereo process and how to create them from the image. We now turn to the solution of the correspondence problem. Marr and Poggio (1979) have developed a theory of human stereopsis which directly addresses this problem. In this section, the theory is outlined outline and some of the evidence in support of it is indicated.

Put succinctly, the problem is how to determine which of the descriptors in the left image corresponds to a particular descriptor in the right image and vice versa.

The two constraints outlined previously are physical constraints, applying to locations on a surface. When they are translated into computational constraints, it is necessary to ensure that the items to which they apply in the image are in one-to-one correspondence with well-defined locations on a physical surface. In the previous section, we have seen how to obtain descriptions of the images which satisfy this condition.

The stereo problem is then reduced to that of matching two primitive symbolic descriptions. The two physical constraints can now be translated into two rules on how the left and right descriptions are combined (Marr and Poggio, 1979):

(1) *Uniqueness.* Except in rare cases, each item from either image may be assigned at most one disparity value. This condition relies on the assumption that an item corresponds to something that has a unique physical position. The exceptions can occur when two features lie along the line of sight from one eye, but are separately visible in the other eye.

(2) *Continuity.* Disparity varies smoothly almost everywhere. This condition is a consequence of the cohesiveness of matter, and it states that only a small fraction of the area of an image is composed of boundaries that are discontinuous in depth.

We must now consider how actually to apply these rules. As has been stated, the major problem for stereo vision is the solution of the correspondence problem, which requires the elimination of false

targets. Marr and Poggio note that the difficulty of this problem is directly proportional to the range and resolution of disparities considered, and to the density of matchable features in an image. In other words, the greater the range over which a match is sought, the greater the number of false targets. At the same time, the finer the resolution at which features are detected, the greater the number of false targets. Thus, one way to avoid the false targets problem is to make the "features" of the image rare. This could be done this by making them very complex or specific, so that even if the total density of features were very high in an image, the density of individual types of features would be very small. However, we have already seen from the evidence of Julesz random-dot stereograms, that this is very unlikely to be a feature of the human system. The other way of making "features" rare is to reduce drastically the density of all features in the image, by decreasing the spatial resolution at which it is examined.

In light of this observation, the existence of independent spatial-frequency-tuned channels, which we discussed in the previous section, suggests that successively finer filtered copies of the image can be used during fusion.

What does this imply computationally? Consider the symbolic descriptions generated by the largest of the channels described in the previous section. Since this channel is tuned to very low spatial frequencies, this means that the density of "features" in the symbolic description is low. As a consequence, the range over which a match may be sought without encountering too many false targets is large. However, this is at the price of reduced resolution of the disparities associated with each match. For a smaller channel, there is a trade-off of these two effects. In particular, the density of false targets will increase, causing the range over which one can safely search for a match to decrease. However, although disparity range is lost, disparity resolution is gained. Thus, for the smallest channel, very high resolution disparity information would be obtained, but the range over which one could safely search for this information will be severely curtailed.

Thus, Marr and Poggio suggest a scheme for solving the fusion problem which has as a first stage, the following operation:

(1) Each image is analyzed through channels of various coarsenesses, and matching takes place between corresponding channels from the two eyes, for disparity values of the order of the channel resolution.

To take advantage of this range of trade-off over the set of channels, some method of combining the information from the different channels is needed. In the end, one would like to be able to obtain fine resolution disparity information. But since the range over which the smallest channel can safely operate is small, it would seem that one must accurately align the two images in order to allow the matching to take place for the smallest channel. Fortunately, the information needed to align the images is available. In particular, the coarse resolution disparity information obtained by matching the larger channel descriptions can be used to align the two images. This will bring the images into a range of alignment where the smaller channels can safely operate, thereby allowing a refinement of the disparity values. In the human visual system, this alignment of the images is accomplished by changing the positions of the eyes, through the use of vergence movements. Note that since the human system performs a global vergence of the two eyes, only the region of the fovea will be likely to be brought into alignment in this manner. In order to obtain fine disparity values throughout the image, each region must be separately brought into the fovea and matched.

Thus, Marr and Poggio suggest that as a second step in the process, the following operation is performed:

(2) Coarse channels control eye vergence movements, thus causing finer channels to come into correspondence.

The disparity information obtained from stereo will be needed by other processes of the visual system. For example, Marr and Nishihara (1978) have argued on computational grounds that the visual system requires an "orientation/depth map" of the visible surfaces of a scene. This is necessary since descriptions of the shapes of objects must be derived via a description of their visible surfaces, and information about these is obtained from a number of different and independent sources. This representation of depth, called the $2\frac{1}{2}$-D sketch, is essentially a memory, into which information from various sources is combined and maintained. Thus, the next stage in the Marr-Poggio scheme is:

(3) When a correspondence is achieved, it is held and written down somewhere.

Finally, once the correspondence for a section of the image has been found, one would like to be able to reuse that information. That is, once the disparity for a region of the image has been computed, one would like to be able to use that information again to fuse that region, without the expense of recomputing the correspondence. Thus, Marr and Poggio suggest as a final stage:

(4) There is a backwards relation between the memory and the filters, that allows one to fuse any piece of a surface easily once its depth map has been established in the memory.

The scheme proposed by Marr and Poggio can be split into two sections, corresponding to the two major steps to be solved in the stereo problem: the extraction of a symbolic description of the images which forms the input to the stereo process; and the determination of the correspondence between these descriptions.

We have indicated that the extraction of a symbolic description may be accomplished by the method of Marr and Hildreth (1979). Note that the Marr-Poggio theory requires the extraction not only of zero-crossings, labelled by contrast and rough orientation, but also of terminations — places of sharp changes in the orientation of zero-crossing contours — also labelled by contrast. Having determined the symbolic descriptions which must be matched, we now turn to the development of the matcher.

2.3.2 The Matcher

The heart of the matching problem concerns the difficulty of false targets. We have noted that computational simplicity can be preserved only if false targets are rare. The use of several independent spatial-frequency-tuned channels provides a way of accomplishing this, over a large range of disparity. In other words, by performing the matching for different resolutions, and using the results of the larger channels to drive the alignment of the smaller channels, fine disparity information can be achieved over a wide range of disparity while avoiding the problem of false targets.

In general terms, the matching process consists of matching symbols of the same type, for each set of filters of a given size. The type of a zero-crossing is determined by its sign and orientation. We have already seen how to compute these attributes. The sign of the zero-crossing is important as a matching attribute, as can be seen from the experiments of Julesz (1963) concerning the impossibility of fusing of a pattern and its negative image.

In designing the matching process, one would like to be able to ensure that false targets are rare, in order to avoid problems associated with choosing between several possible matches. To do this, Marr and Poggio performed a statistical analysis of the zero-crossings to determine the probability distribution of the interval between adjacent zero-crossings of the same sign in the filtered image. Based on this probability distribution, one can design a matching process which will in essence avoid

the false target problem.

According to this analysis, inherently valid only for oriented filters, if one wishes to avoid false targets altogether, the disparity range over which a match is sought must be restricted to $\pm\frac{w}{2}$, where w is the width of the central region of the filter. Suppose that a zero-crossing L in the left image matches some zero-crossing R in the right image. The probability distribution derived by Marr and Poggio states that the probability of another zero-crossing of the same sign within $\frac{w}{2}$ image elements of R is less than 0.05. This means that if the disparity between the two images for this region of the image is less than $\frac{w}{2}$, a matcher which searches for possible matches in the range $\pm\frac{w}{2}$ will find only the correct match with probability 0.95, and we will almost always get the correct match.

To enlarge the range of disparity accepted, the following modification is possible. Suppose the size of the search region is expanded to $\pm w$. Then, if the disparity between the two images is less than $|w|$, Marr and Poggio show that 50% of all matches will be correct and unambiguous. The remainder will be ambiguous, mostly with two alternatives, one convergent (in the range $(0, w)$) and one divergent (in the range $(-w, 0)$), one of which is always correct. For these ambiguous cases, the correct alternative can still be found. Consider the sign of the disparity of the neighbouring unambiguous matches. (By the sign of the disparity, I mean the sign of the direction of the matching zero-crossing: crossed or convergent, uncrossed or divergent, and zero.) One may choose one of the ambiguous alternatives by simply selecting that which has the same sign as the dominant sign of the neighbouring unambiguous matches. Note that this implicitly uses the constraint of continuity.

Finally, the above cases deal with the situation in which the disparity of the elements is less than $|w|$. If the disparity exceeds the operating range, one wants to ensure that the algorithm is capable of detecting this fact. Marr and Poggio show that given that the images are outside the acceptable disparity range $|w|$, the probability of a random match is about 0.7. Thus the probability of a zero-crossing having no candidate match is 0.3 if the images are outside the acceptable disparity range, and 0.0 if the images are within the acceptable disparity range. This can easily be detected and used to remove random matches. In this way one can ensure not only that correct disparities are computed where available, but also that incorrect disparities are avoided.

Thus, on basic computational grounds, a method for matching the symbolic descriptions of the images has been developed. Marr and Poggio (1979) also show strong psychophysical and

neurophysiological grounds for this particular form of matcher. For details, the reader is referred to their paper.

2.4 Summary

We have shown that the stereo process consists of two major problems, the extraction of a description of the elements of an image corresponding to physically identifiable locations in the scene, and the determination of the corresponding descriptors from each processed image. The problem of extracting matchable descriptions of the images is solved by the Marr-Hildreth theory of edge detection. This consists of convolving each image with a set of filters of the form $\nabla^2 G$, where ∇ is the Laplacian and G is a Gaussian. For each size mask, the zero-crossings of the convolved image are localized, and form the descriptions of the image to be matched.

For each mask size, the Marr-Poggio theory of stereo vision states that matching takes place between zero-crossing segments of the same sign and roughly the same orientation in the two images, for a range of disparities up to about the width of the mask's central region. Within this disparity range, false targets pose only a simple problem, because of the roughly bandpass nature of the filters, and the matching can proceed successfully.

Note that a major consequence of this approach is that not every point in an image will receive an explicit match. I shall return to this consequence in the latter half of the thesis.

CHAPTER 3

THE STEREO IMPLEMENTATION

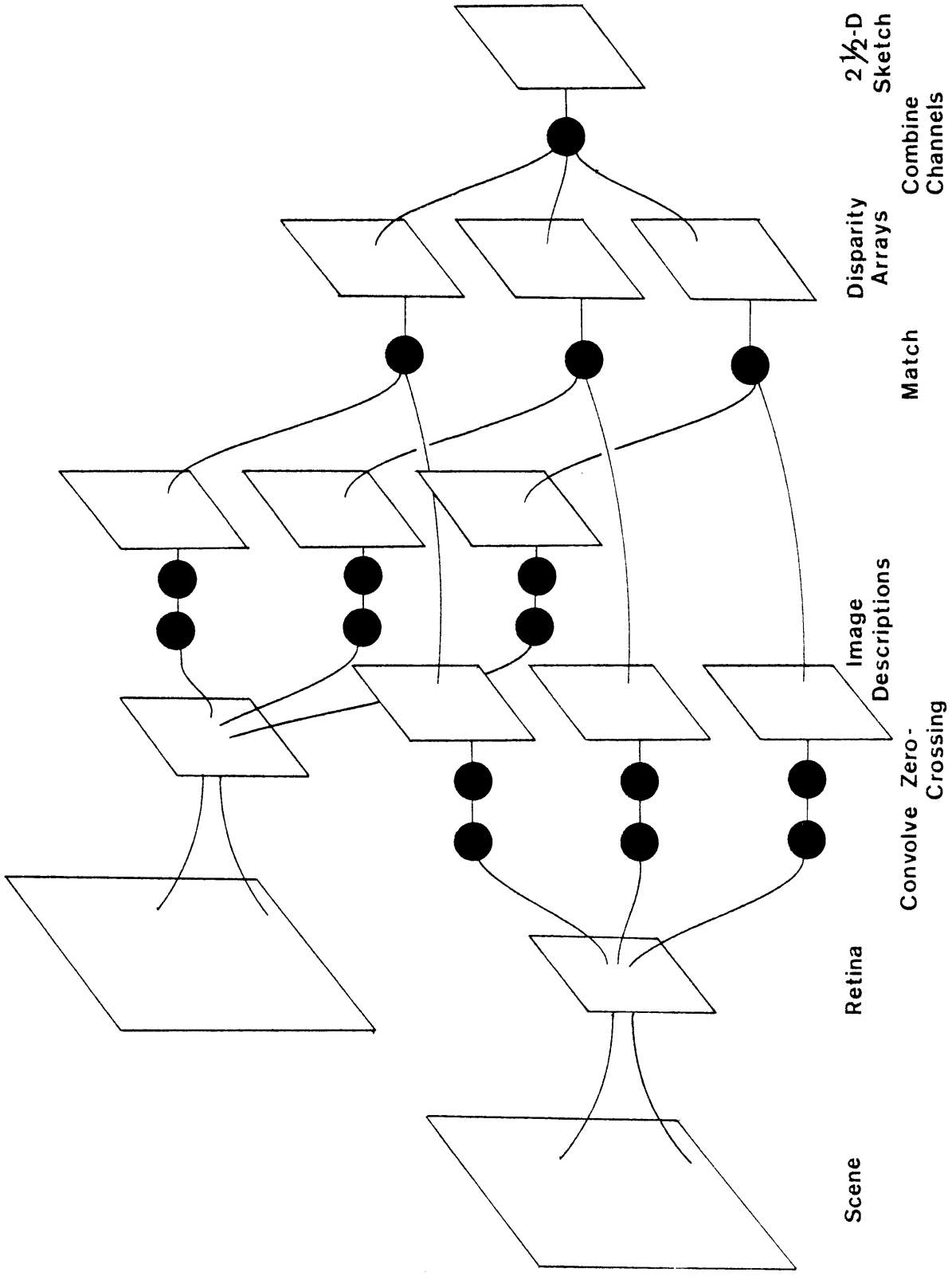
An important aspect in the development of any computational theory is the design and implementation of an explicit algorithm for that theory. There are several benefits from such an implementation. One concerns the act of implementation itself, which forces one to make all details of the theory explicit. This often uncovers previously overlooked difficulties which guide further refinement of the theory.

A second benefit concerns the performance of the implementation. Any proposed model of a system must be testable. In this case, by testing on pairs of stereo images, one can examine the performance of the implementation, and hence of the theory itself, provided, of course, that the implementation is an accurate representation of that theory. In this manner, the performance of the implementation can be compared with human performance. If the algorithm differs strongly from known human performance, its suitability as a biological model is quickly brought into question (c.f. the analysis in Marr and Poggio, 1979 of the cooperative algorithm of Marr and Poggio, 1977b).

This chapter describes an implementation of the Marr-Poggio stereo theory, written with particular emphasis on the matching process (Grimson, 1980). Later chapters will discuss the effect of the implementation on refining the theory, and the analysis of the performance of the implementation on several test cases.

The implementation is divided into five modules, roughly corresponding to the five steps of the Marr-Poggio theory. These modules are in Figure 3.1, and each is described in turn.

Figure 3.1. Diagram of the Implementation. From the initial arrays of the scene, a retinal pair of images are extracted. These subimages reflect the current eye positions relative to the scene. Each retinal image is convolved with a set of different sized filters of the form $\nabla^2 G$, and the zero-crossings of the output are located. For each filter size, the zero-crossing descriptions are matched, based on the sign and orientation of each zero-crossing point. The disparity arrays created for each filter are combined into a single disparity description, and information from the larger filters can be used to verge the eyes, thereby bring the smaller filters into a range of operation.


Scene
Retina
Convolve
Zero-Crossing
Image Descriptions
Match
Disparity Arrays
Combine Channels
2½-D Sketch

3.1 Input

There are two aspects of the human stereo system, embedded in the theory, which must be made explicit in the input to the algorithm. The first is the position of the eyes with respect to the scene, as eye movements will be critical for obtaining fine disparity information. The second is the change in resolution of analysis of the image with increasing eccentricity, (this was indicated by the Wilson and Bergen data).

To account for these effects, the algorithm maintains as its initial input a stereo pair of arrays, representing the entire scene visible to the viewer. This pair of arrays corresponds to the environment around the visual system, rather than some integral part of the system itself. To create this representation of the scene, photographs of natural images were digitized on an Optronix Photoscan System P1000. The sizes of these images are indicated in the legends. Grey-level resolution is 8 bits, providing 256 intensity levels. For the random dot patterns illustrated in this thesis, the images were constructed by computer, rather than digitized from photographs.

For a given position of the eyes relative to the scene, a representation of the images on the two retinas is extracted. The program creates this retinal representation by obtaining a second, smaller pair of images from the arrays representing the whole scene. The mapping from the scene arrays into the retinal images accounts for one of the factors inherent in the way the human visual system is constructed. Different sections of the scenes will be mapped to the center (fovea) of the retinal images as the positions of the eyes are varied. In this way, different sections of the scenes can separately be matched to a very fine level of disparity, by allowing the smallest channels to come into range of correspondence. Since the matching process will take place on the array representing the retinal images, it is important that the coordinate systems of those arrays coincide with the current positions of the eyes. Note that the portion of the scene image which is mapped into the retinal image may differ for the two eyes, depending on the relative positions of the two optical axes. In particular, there may be differences in vertical alignment as well as in horizontal alignment. There is a second factor which should also be taken into account. Wilson and Bergen (1979), and Wilson and Giese (1977) state that the resolution of the earlier stages of the algorithm — the convolution and zero-crossings — scales linearly with eccentricity. However, this aspect has not been implemented and in our situation is not critical, since the images analyzed correspond to small visual angles, on the order of 4^o on a side.

After the completion of this stage, the program has created a representation of the images that has accounted for eye position and if appropriate also for retinal scaling with eccentricity. For each pass of the algorithm, the matching will take place on the representation of the retinal images, thereby implicitly assuming some particular eye positions. Once the matching has been completed, the disparity values obtained may be used to change the positions of the two optic axes, thus causing a new pair of retinal images to be extracted from the representations of the scene, and the matching process may proceed again.

3.2 Convolution

Given the retinal representations of the images, it is necessary to transform them into a representation upon which the matcher may operate. We have already seen arguments in the previous chapter concerning the form and size of the filters required to perform this transformation. The present implementation uses four filters, each of which has the form of $\nabla^2 G$, the Laplacian of a Gaussian, with w values of 4, 9, 17 and 35 image elements. These values are derived from the data of Wilson and Bergen (1979). The coefficients of the filters were represented to a precision of 1 part in 2048. Coefficients of less than $\frac{1}{2048}$'th of the maximum value of the filter were set to zero. Thus, the truncation radius of the filter (the point at which all further filter values were treated as zero) was approximately $1.8w$, or equivalently, 5.08σ.

The actual convolutions were performed on a LISP machine constructed at the MIT Artificial Intelligence Laboratory, using additional hardware specially designed for the purpose (Knight, et al. 1979). Figures 2.4 and 2.5 illustrated some images and their convolutions with various sized filters.

After the completion of this stage of the algorithm, one has four filtered copies of each of the images, each copy having been convolved with a different size filter.

3.3 Detection and Description of Zero-Crossings

The elements that are matched between images are (i) zero-crossings whose orientations are not horizontal, and (ii) terminations. The exact definition and hence the detection of terminations is at present uncertain. Moreover, terminations are much rarer than zero-crossings. As a consequence, only zero-crossings are used as input to the matcher.

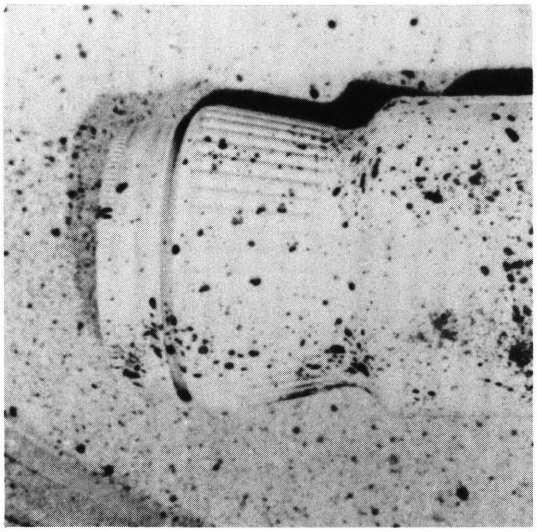
Since, for the purpose of obtaining disparity information, horizontally oriented segments may

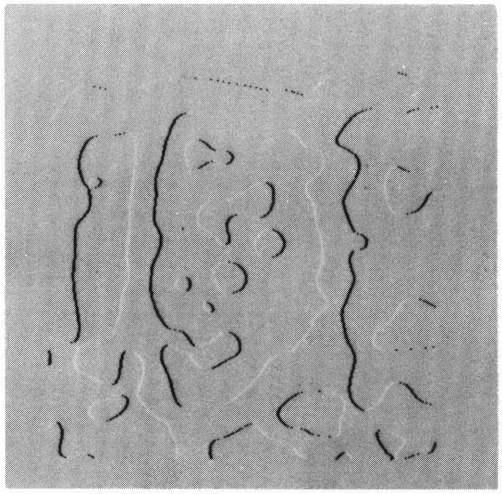
be ignored, the detection of zero-crossings can be accomplished by scanning the convolved image horizontally for adjacent elements of opposite sign, or for three horizontally adjacent elements, the middle one of which is zero, the other two containing convolution values of opposite sign. This gives the position of zero-crossings to within an image element. Note that there is no theoretical limit on the accuracy with which the zero-crossings may be localized. For the purposes of matching, however, a resolution of one pixel suffices.

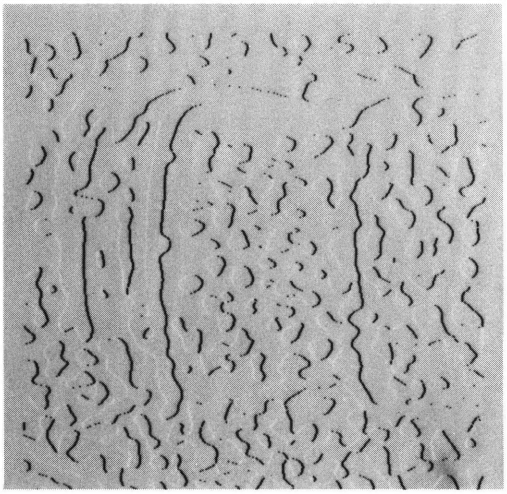
In addition to their location, the sign of the zero-crossings and a rough estimate of the local, two-dimensional orientation of pieces of the zero-crossing contour are recorded. In the present implementation, the orientation at a point on a zero-crossing segment is computed as the direction of the gradient of the convolution values across that segment, and recorded in increments of 30 degrees. Figures 3.2 and 3.3 illustrate zero-crossings obtained in this way from the convolutions of Figures 2.4 and 2.5. Positive zero-crossings are shown white, and negative crossings, black.

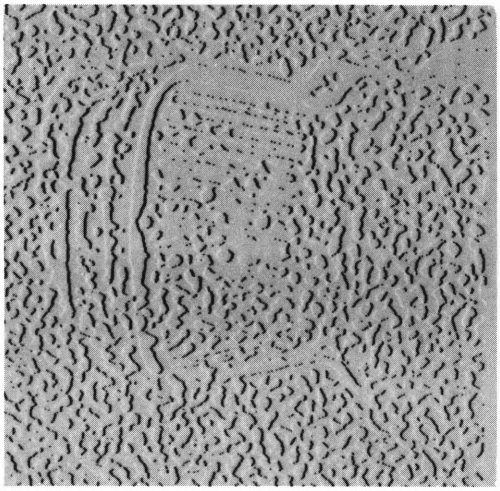
This zero-crossing description is computed for each image and for each size of filter.

Figure 3.2. Examples of Zero-Crossings. A natural image is shown at the top. Below are examples of the zero-crossings, obtained from different sized $\nabla^2 G$ operators, with central panel widths of 36, 18, 9 and 4 picture elements. The positive zero-crossings are shown as white, the negative ones as black.









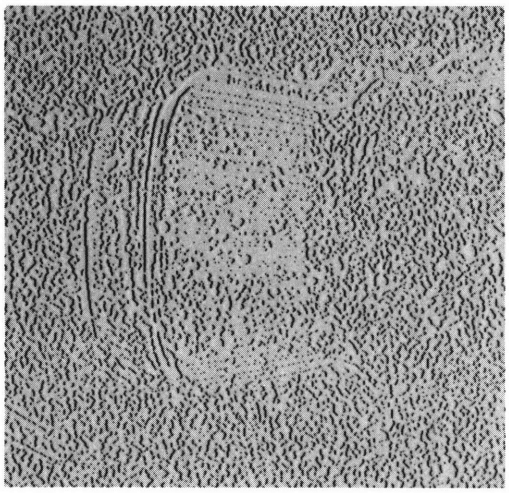
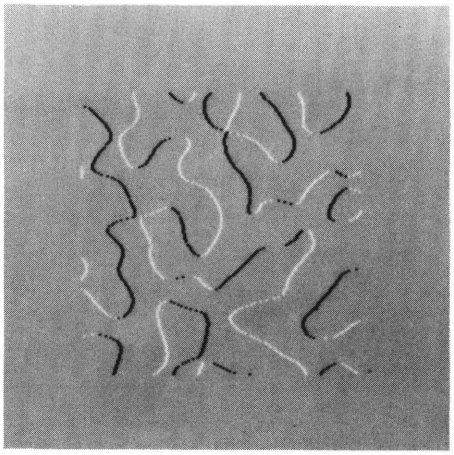
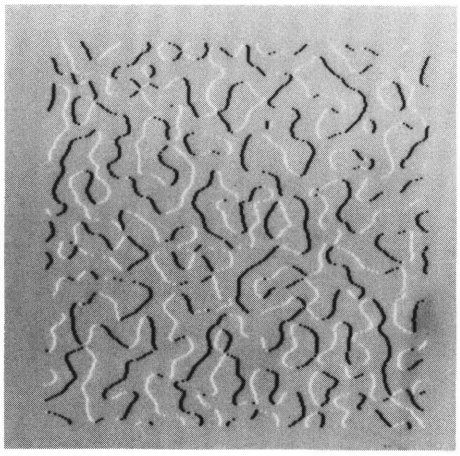
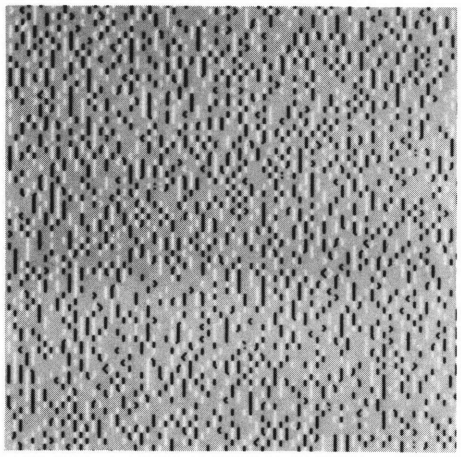


Figure 3.3. Examples of Zero-Crossings. A random dot pattern is shown at the top. Below are examples of the zero-crossings, obtained from different sized $\nabla^2 G$ operators, with central panel widths of 36, 18, 9 and 4 picture elements. The positive zero-crossings are shown as white, the negative ones as black.







3.4 Matching

The matcher implements the second of the matching algorithms described by Marr and Poggio (1979, p.315), and outlined in Chapter 2. For each size of filter, matching consists of 6 steps:

(1) Fix the eye positions.

(2) Locate a zero-crossing in one image.

(3) Divide the region about the corresponding point in the second image into three pools.

(4) Assign a match to the zero-crossing based on the potential matches within the pools.

(5) Disambiguate any ambiguous matches.

(6) Assign the disparity values to a buffer.

These steps may be repeated several times during the fusion of an image. Given positions for the eyes, these matching steps are performed, with the results stored in a buffer. These results may be used to refine the eye positions, causing a new set of retinal images to be extracted from the scene, and the matching steps are performed again.

The first step consists of fixing the two eye positions. The alignment between the two zero-crossing descriptions, corresponding to the positions of the eyes, is determined in two ways. The initial offsets of the descriptions are arbitrarily set to zero. Thereafter, the offsets of the two eyes are determined by accessing the current disparity values for a region and using these values to adjust the vergence of the eyes. In my implementation, this is done by modifying the extraction of the retinal images from the images of the entire scene, as explained earlier.

Once the eye positions have been fixed, and the retinal images extracted, the zero-crossing descriptions are obtained as is Figures 3.2 and 3.3. For a zero-crossing description obtained from a particular filter size, the matching is performed by locating a zero-crossing and performing the following operation. Given the location of a zero-crossing in one image, a region about the same location in the other image is partitioned into three pools. These pools form the region to be searched for a possible matching zero-crossing and consist of two larger convergent and divergent regions, and a smaller one lying centrally between them. Together these pools span a disparity range equal to $2w_{1\_d}$ where w_{1_d} is the width of the central excitatory region of the corresponding one-dimensional convolution filter.

MATCHING

The following criteria are used for matching zero-crossings in the left and right filtered images, for each pool:

(1) The zero-crossings must come from convolutions with the same size filter.

(2) The zero-crossings must have the same sign.

(3) The zero-crossing segments must have roughly the same orientation.

A match is assigned on the basis of the responses of the pools. If exactly one zero-crossing of the appropriate sign and orientation (within 30 degrees) is found within a pool, the location of that crossing is transmitted to the matcher. If two candidate zero-crossings are found within one pool (an unlikely event), the matcher is notified and no attempt is made to assign a match for the point in question. If the matcher finds a single crossing in only one of the three pools, that match is accepted, and the disparity associated with the match is recorded in a buffer. If two or three of the pools contain a candidate match, the algorithm records that information for future disambiguation.

Once all possible unambiguous matches have been identified, an attempt is made to disambiguate double or triple matches. This is done by scanning a neighbourhood about the point in question, recording the sign of the disparity of the unambiguous matches within that neighbourhood. (The sign of the disparity refers to the sign of the pool from which the match comes: divergent, convergent or zero.) If the ambiguous point has a potential match of the same sign as the dominant type within the neighbourhood, then that is chosen as the match (this is the "pulling" effect). Otherwise, the match at that point is left ambiguous.

There is the possibility that the region under consideration does not lie within the $\pm w$ disparity range examined by the matcher. This situation is detected and handled by the following operation. Consider the case in which the region does lie within the disparity range $\pm w$. Excluding the case of occluded points, every zero-crossing in the region will have at least one candidate match (the correct one) in the other filtered image. On the other hand, if the region lies beyond the disparity range $\pm w$, then the probability of a given zero-crossing having at least one candidate match will be less than 1. In fact, the probability of a zero-crossing having at least one candidate match in this case is roughly 0.7. Hence, the following operation can be performed. For a given eye position, the matching algorithm is run for all the zero-crossings. Any crossing for which there is no match is marked as such. If the percentage of matched points in any region is less than a threshold of 0.7 then the region is declared

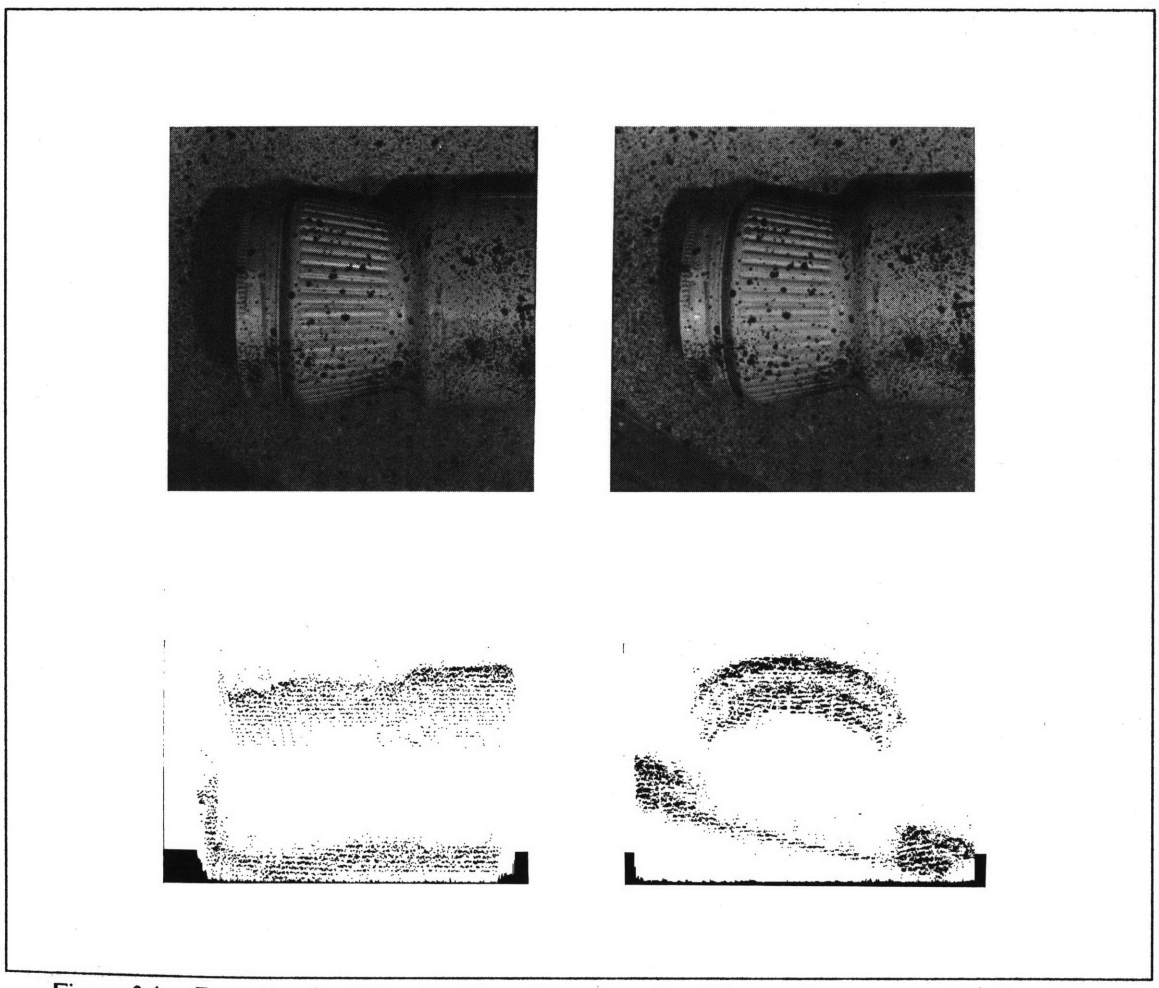


Figure 3.4. Example of a Disparity Map. The top pair of images is a stereo pair of a natural scene. In the bottom two images, two views of the disparity map are shown. A point (x, y) with disparity d is represented in a three-dimensional array as the point (x, y, d). This array is then viewed from two different angles, the left image corresponding to viewing the array from the lower edge of the original image, the right image corresponding to viewing the array from the left edge of the original image. For graphical convenience, the disparity values have been inverted.

to be out of range, and no disparity values are accepted for that region.

The size of the regions used for checking the statistics of matching zero-crossings should be proportional to the density of the zero-crossings, in order to ensure a fixed confidence level. Typically, the regions were roughly 25 picture elements on a side, for a channel with filter size $w = 9$.

The overall effect of the matching process, as driven from the left image, is to assign disparity values to most of the zero-crossings obtained from the left image. An example of the output appears in Figures 3.4 and 3.5. In this array, a zero-crossing at position (x, y) with associated disparity d has

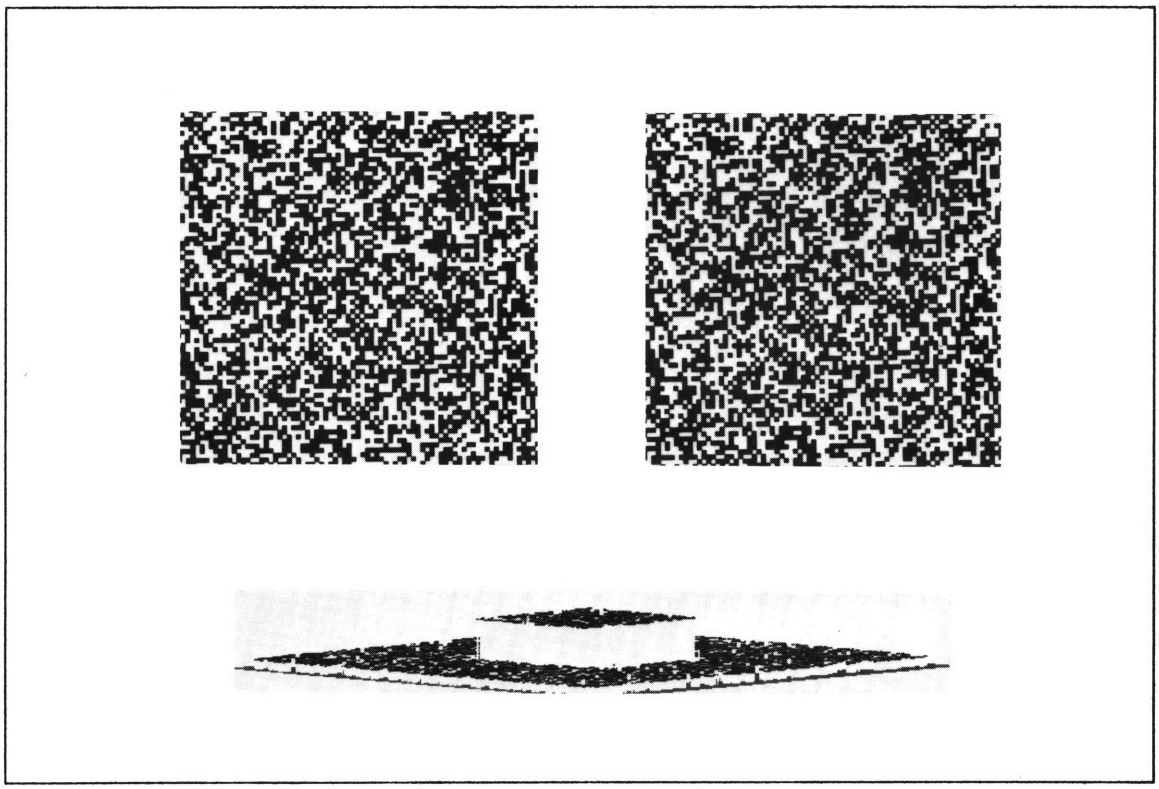


Figure 3.5. Example of a Disparity Map. The top pair of images is a stereo pair of a random dot stereogram. The bottom image shows a view of the disparity array. For graphical convenience, the disparity values have been inverted.

been placed in a three-dimensional array with coordinate (x, y, d). For display purposes, the array is shown in the figures as viewed from a point some distance away. The heights in the figure correspond to the assigned disparities. (For graphical convenience, the disparities have been inverted.)

After completion of this stage of the processing, a disparity array for each filter size has been obtained. The disparity values are located only along the zero-crossing contours obtained from that filter.

3.5 Vergence Control

The Marr-Poggio theory states that in order to obtain fine resolution disparity information, it is necessary that the zero-crossings from the smallest channels be assigned a match. Since the range of disparity over which a channel can obtain a match is directly proportional to the size of the channel, this means that the eyes must move to ensure that the corresponding zero-crossing descriptions from

the two images are within a matchable range. The disparity information required to bring the smallest channels into their matchable range is provided by the larger channels. That is, if a region of the image is declared to be out of range of fusion by the smaller channels, one can frequently obtain a rough disparity value for that region from the larger channels, and use this to verge the eyes. In this way, the smaller channels can be brought into a range of correspondence.

Thus, after the disparities from the different channels are combined, there is a mechanism for controlling vergence movements of the eyes. This operates by searching for regions of the image which have disparity values from the larger channels, but do not have disparity values for the smallest channel. The values from the large channel are used to provide a refinement to the current eye positions, thereby bringing the smaller channels into range of correspondence. Possible mechanisms for extracting the disparity value from a region of the image include using the peak value of a histogram of the disparities in that neighbourhood, using the average of the local disparity values, or using the median of the local disparity values. In the current implementation, the search for such a region proceeds outwards from the fovea.

It should be noted here that although Marr and Poggio state that disparity information from the coarser channels can drive eye movements, they do not rule out that other information can also do this. There may be other modules of the visual system that can initiate eye movements. Kidd et al (1979) for example, found that certain types of texture boundaries can initiate eye movements. However, such effects are somewhat orthogonal to the question of the adequacy of the matching component of the Marr-Poggio theory, since they affect the input to the matcher, but not the actual performance of the matching algorithm itself.

3.6 The $2\frac{1}{2}$-Dimensional Sketch

Once the separate channels have performed their matching, the results are combined and stored in a buffer, called the $2\frac{1}{2}$-D sketch. There are several possible methods for accomplishing this. As far as the Marr-Poggio theory is concerned, the important point is that some type of storage of disparity information occurs, that there is some kind of buffer. Perhaps the strongest argument for this is the fact that up to 2 degrees of disparity can be held fused in the fovea, although the matching range for a single fixation of the eyes is only 30 minutes of arc.

I have considered two different possibilities for the way in which information from the different

channels is combined. The method used in the current implementation will be described below. A more biologically feasible method will be outlined in the discussion.

One of the critical questions concerning the form of the $2\frac{1}{2}$-D sketch is whether its coordinates are consistent with those of the scene or those of the retinal images. For all the cases illustrated, the sketch was constructed by directly relating the coordinates of the sketch to the coordinates of the scene arrays. That is, as disparity information was obtained, it was stored in a buffer at the position corresponding to the position in the original scene from which the underlying zero-crossing came. Since disparity information about the scene is extracted from several eye positions, explicit information about the positions of the eyes is required in order to store this information into a buffer. This is probably inappropriate as a model of the human system, but it suffices for demonstrating the effectiveness of the matching module.

The actual mechanism for storing the disparity values requires some combination of the disparity maps obtained for each of the channels. Currently, the sketch is updated, for each region of the image, by writing in the disparity values from the smallest channel which is within range of fusion. Vergence movements are possible in order to bring smaller channels into a range of matching for some region. Further, for those regions of the image for which none of the channels can find matches, modification of the eye positions over a scale larger than that of the vergence movements is possible. By this method, one can attempt to bring those regions of the image into a range of fusion. There are several possibilities for the actual method of driving the vergence movements. Two of these were outlined in the previous section.

The final output of the algorithm consists of a representation of disparity values in the image, specified along zero-crossings segments from the smallest channel that was used to analyze that part of the scene.

3.7 Summary of the Process

The complete algorithm, as currently implemented, uses four filter sizes. Initially, the two views of the scene are mapped into a pair of working arrays. These arrays are convolved with each filter. The zero-crossings and their orientation are computed, for each channel. The initial alignments of the eyes determine the initial registration of the images. The matching of the descriptions from each channel is performed for this alignment. Any points with either ambiguous matchings or with no match are

marked as such.

Next, the percentage of unmatched points is checked, for all square neighbourhoods of a particular size. This size is chosen so as to ensure that the measurement of the statistics of matching within that neighbourhood is statistically sound. Only the disparity points of those regions whose percentage of unmatched points is below a certain threshold, are allowed to remain. All other points are removed. These values are stored in a buffer. At this stage, vergence movements may take place, using information from the larger channels to bring the smaller channels into a range where matching is possible. Further, if there are regions of the image which do not have disparity values for any channel, an eye movement may take place in an attempt to bring those portions of the image into a range where at least the largest filter can perform its matching.

Note that the matching process takes place independently for each of the four channels. Once the matching of each channel is complete, the results are combined into a single representation of the disparities.

The final output is thus a disparity map, with disparities assigned along most portions of the zero-crossing contours obtained from the smallest filters used. The accuracy of the disparities thus obtained depends on how accurately the zero-crossings have been localized, which may, of course, be to a resolution much finer than the initial array of intensity values that constitutes the image.

CHAPTER 4

ANALYSIS AND DEVELOPMENT

In the previous chapter, I described an explicit algorithm derived from the Marr-Poggio theory, and its implementation in a computer program. Since Marr and Poggio's explicit concern was to develop a theory of the human stereo system, the implementation can be used to test the adequacy of their theory. This was done by comparing the results of running the program with the preception of human subjects on a range of carefully chosen stereoscopic images.

Besides helping to examine the adequacy of the theory, the performance of the algorithm can also bring out aspects of the visual processing system which had not previously been noticed. In particular, it can uncover difficulties with the theory, or bring out its advantages. Thus, the program itself plays an essential role in the overall task of deciding whether, and how closely, the algorithm it implements mirors the human analysis of stereoscopic images.

4.1 Performance on Random Dot Patterns

Since random dot stereograms (Julesz 1960, 1971) contain no visual cues other than the stereoscopic ones, they are a useful tool for studying the stereo component of the human visual system in isolation. One test of the adequacy of the algorithm as representative of human stereo vision is to compare human perception and the performance of the algorithm on such patterns. Since random dot stereograms have known disparity values, these patterns can also be used to assess the correctness of the algorithm's performance.

PERFORMANCE ON RANDOM DOT PATTERNS

Table 1 lists some of the matching statistics for various random dot patterns. These are illustrated in the figures and discussed below.

TABLE OF MATCHES						
Pattern	Density	Total	Exact	One Pixel	Wrong	%Wrong
Square	50%	11847	11830	14	3	.03
Square	25%	9661	9632	22	7	.07
Square	10%	5286	5264	20	2	.04
Square	5%	3500	3498	0	2	.06
Wedding	50%	11162	11095	61	6	.06
Noise-w4	50%	2270	1909	346	15	.7
Noise-w9	50%	8683	6621	1868	194	2.
Noise-w4-1	50%	63	28	24	11	17.
Noise-w9-1	50%	8543	5194	2864	485	6.
90%Corr	50%	9545	9091	263	191	2.
80%Corr	50%	4343	4120	143	80	2.
70%Corr	50%	134	127	2	5	4.
Diag Corr	50%	6753	6325	271	157	2.

Table 1.

The first pattern consisted of a central square separated in depth from a second plane. The statistics of matching are labelled by the 50% Square row in Table 1 . The pattern had a dot density of 50% and its analysis was shown in Figure 3.4. Each dot was a square four image elements on a side. For the algorithm, this corresponds to a dot of approximately two minutes of visual arc. The total pattern was 320 image elements on a side. The central plane of the figure was shifted 12 image elements in one image relative to the other. The final disparity map assigned after the matching of the smallest channel had the following statistics. The number of zero-crossing points in the left description which were assigned a disparity was 11847. Of these 11847, 11830 were disparity values which were exactly correct, and an additional 14 deviated by one image element from the correct value. Approximately 0.03% of the matched points, or roughly 3 points in 10000 were incorrectly matched.

A similar test was run on patterns with a dot density of 25%, 10% and 5%. These are shown in Table 1 in the rows labelled 25% Square, 10% Square and 5% Square. The results are illustrated in Figures 4.1 and 4.2. For each of these cases, the number of incorrectly matched points was extremely

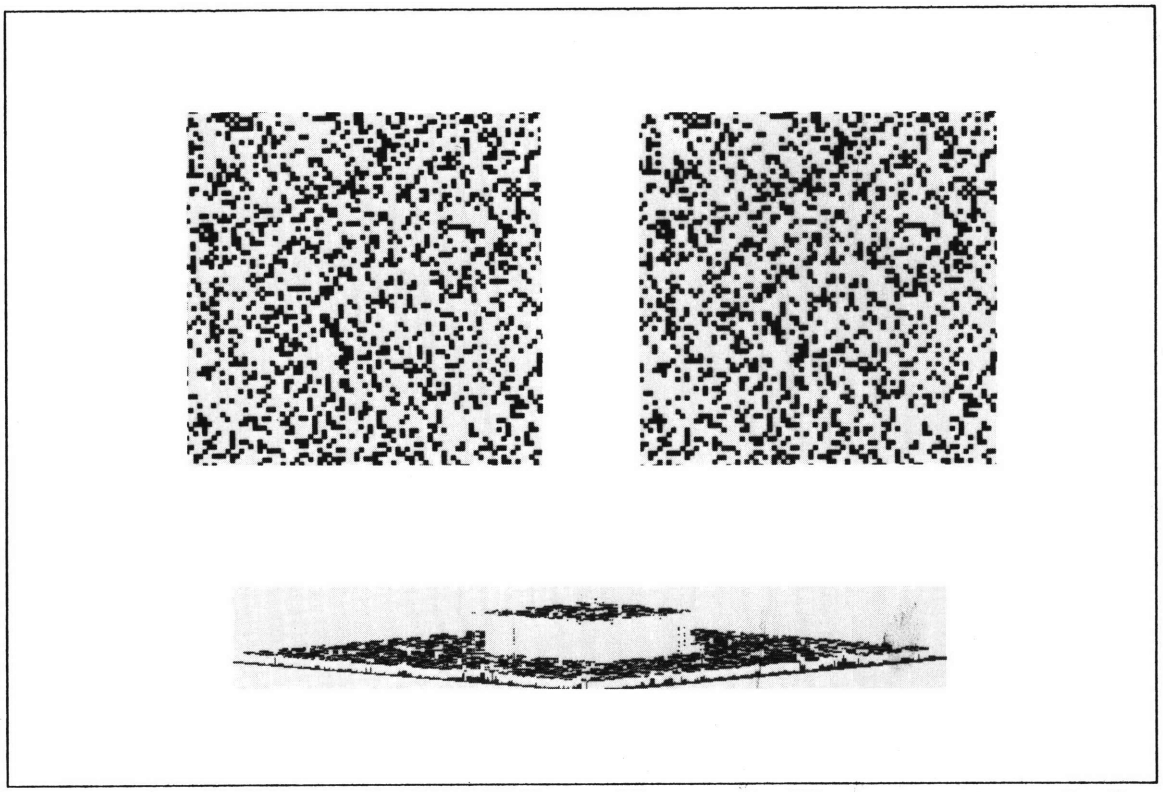


Figure 4.1. 25% Density Random Dot Pattern. The top pair of images are a stereo pair of random dot patterns. The bottom image is a view of the disparity array. A point (x, y) with disparity d has been represented as the point (x, y, d). For graphical clarity, the disparities have been inverted. This three-dimensional array is viewed from some distance away in order to illustrate the disparities.

low, their error rates lying around 0.05%. Those points which were assigned incorrect disparities all occurred at the border between the two planes, that is, along the discontinuity in disparity. This was also true for the 50% density case.

Figure 4.3 shows a more complex random dot pattern, consisting of a wedding cake, built from four different planar layers, each separated by 8 image elements, or 2 dot widths. The matching statistics are shonw in Table 1 in the row labelled Wedding. In this case, the number of zero-crossing points assigned a disparity was 11162. Of these points, 11095 were assigned a disparity value which was exactly correct, and an additional 61 deviated from the correct value by one image element. Approximately 0.06% of the points were incorrectly matched. Again, these incorrect points all occured at the boundaries between the planes. A second complex pattern is illustrated in Figure 4.4. The object is a spiral staircase with a range of continuously varying disparities.

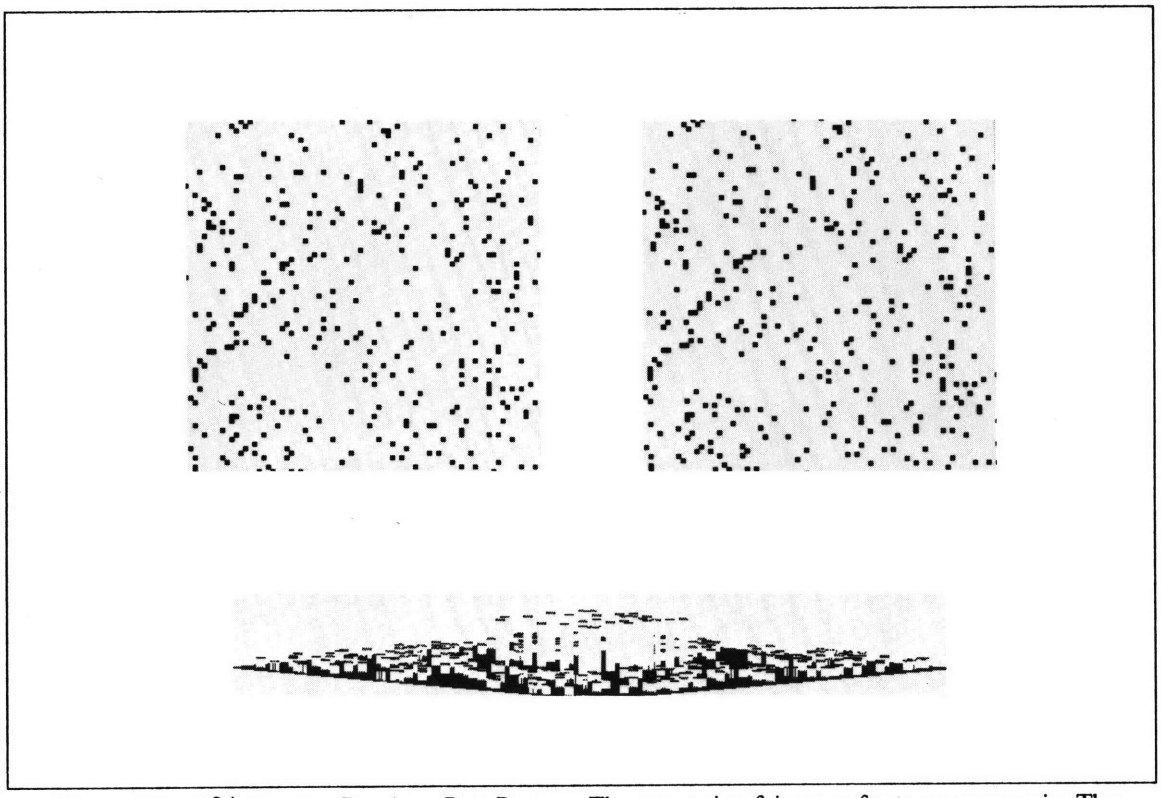


Figure 4.2. 5% Density Random Dot Pattern. The top pair of images form a stereo pair. The bottom image is a view of the disparity array.

There are a number of special cases of random dot patterns which have been used to test various aspects of the human visual system. The algorithm was also tested on several of these stereograms. They are outlined below and a comparison is given between the performances of the algorithm and of humans with good stereo vision.

It is known that if one or both of the images of a random dot stereogram are blurred, fusion of the stereogram is still possible (Julesz 1971, p.96). To test the algorithm in this case, the left half of a 50% density pattern was blurred by convolution with a Gaussian filter. This is illustrated in Figure 4.5. The disparity values obtained in this case were not as exact as in the case of no blurring. Rather, there was a distribution of disparities about the known correct values. As a result, the percentage of points that might be considered incorrect (more than one image element deviation from the correct value) rose to 6%. The qualitative performance of the algorithm was still correct, however, representing two planes separated in depth. It is interesting to note that slight distribution of disparity values about those corresponding to the original planes is consistent with the human perception of a

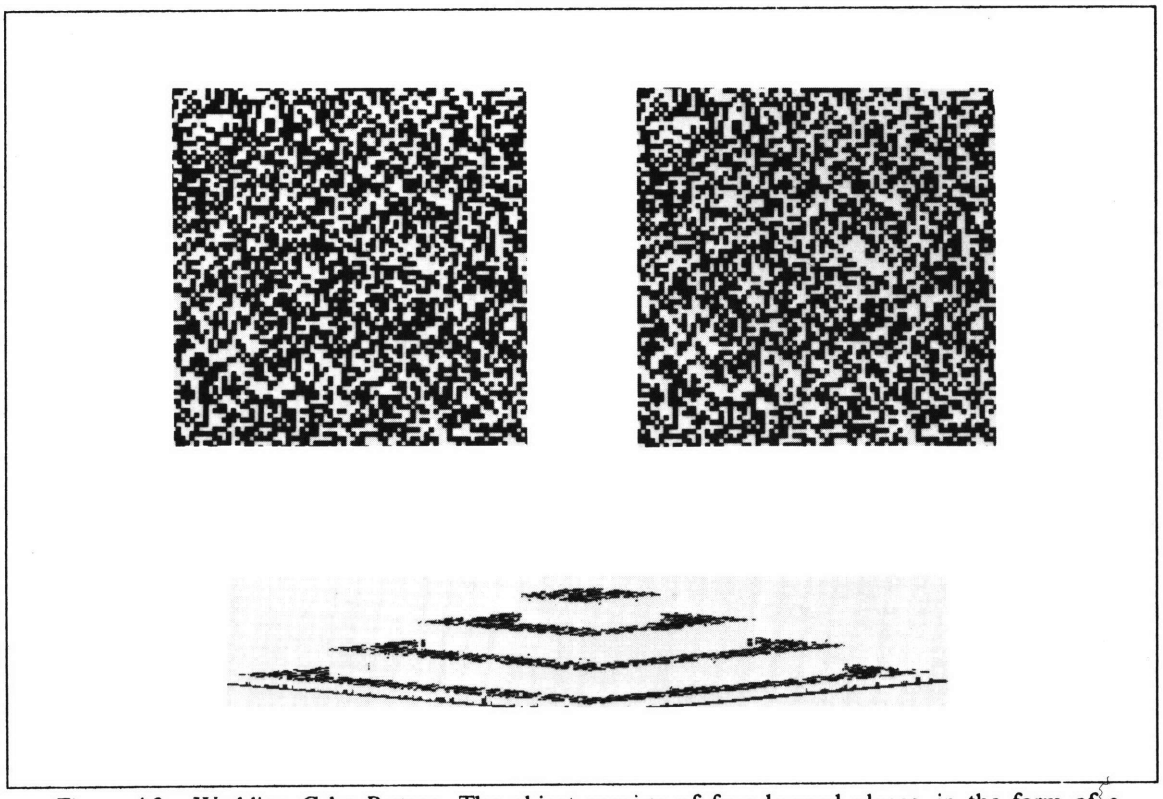


Figure 4.3. Wedding Cake Pattern. The object consists of four layered planes, in the form of a wedding cake.

pair of slightly warped planes. For larger filters, there was little difference between the performance of the algorithm on this stereogram and its performance on stereograms which have not been blurred.

Julesz and Miller (1975) showed that fusion is also possible in the presence of some types of masking noise. In particular, if the spectrum of the noise is sufficiently far from the spectrum of the pattern, fusion of the pattern is still possible. Within the framework of the Marr-Poggio theory, this is equivalent to stating that if one introduces noise of such a spectrum as to interfere with one of the stereo channels, fusion is still possible among the other channels, provided that the noise does not have a substantial spectral component overlapping other channels as well. This was tested on the algorithm by high pass filtering a second random dot pattern, to create the noise, and adding the noise to one image. In the cases illustrated in the Figures 4.6 and 4.7, the spectrum of the noise was designed to interfere maximally with the smallest channel. In the case of the patterns labelled in table 1 by Noise-w4 and Noise-w9, the noise was added such that the maximum magnitude of the noise was equal to the maximum magnitude of the original image. Noise-w4 illustrates the performance of the

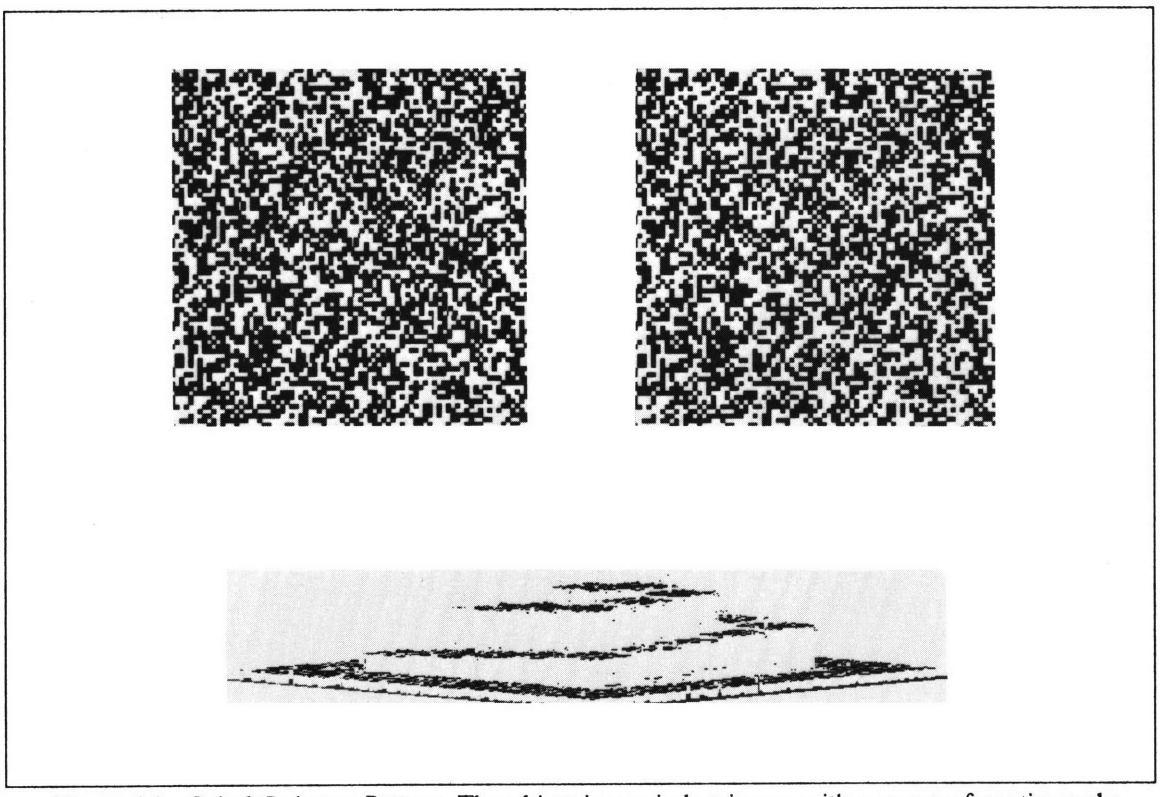


Figure 4.4. Spiral Staircase Pattern. The object is a spiral staircase, with a range of continuously varying disparity.

smallest channel. Noise-w9 illustrates the performance of the next larger channel. It can be seen that for this case, some fusion is still possible in the smallest channel, although it is patchy. The next larger channel also obtains fusion. In both cases, the accuracy of the disparity values is reduced from the normal case. This is to be expected, since the introduction of noise tends to displace the positions of the zero-crossings. In the cases labelled in table 1 by Noise-w4-1 and Noise-w9-1, the noise was added such that the maximum magnitude was twice that of the maximum magnitude of the original image. Here, matching in the smallest channel is almost completely eliminated (Noise-w4-1). Yet matching in the next larger channel is only marginally affected (Noise-w9-1).

The implementation was also tested on the case of adding low pass filtered noise to a random dot pattern, with results similar to that of adding high pass filtered noise. Here, the larger channels are unable to obtain a good matching, while the smaller channels are relatively unaffected.

If one of the images of a random dot pattern is compressed in the horizontal direction, the

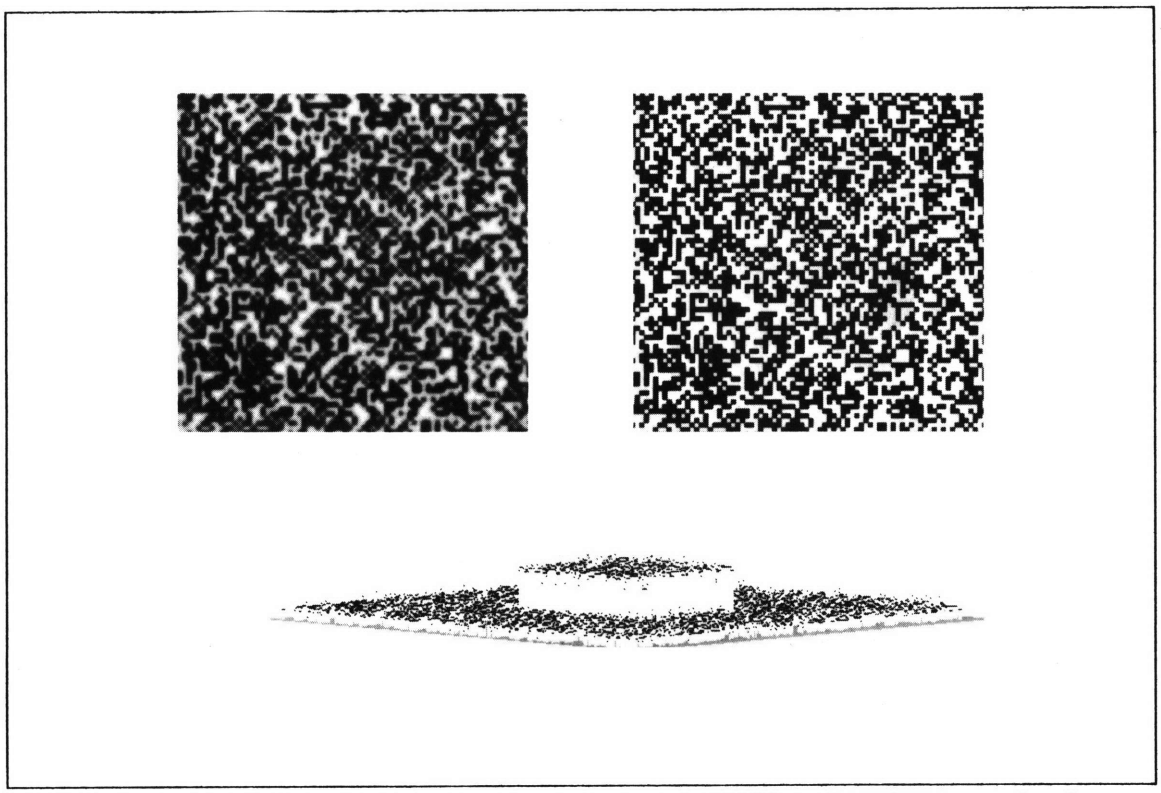


Figure 4.5. Blurred Stereogram. One of the images of a 50% random dot pattern has been altered by convolution with a Gaussian. Fusion is still obtained, although the disparity values are not as sharp as in the original case.

human stereo system is still able to achieve fusion (Julesz 1971, p.213). The algorithm was tested on this case, and the results are shown in Figure 4.8. It can be seen that the program still obtains a reasonably good match. The planes are now slightly slanted, which agrees with human perception.

If some of the dots of a pattern are decorrelated, it is still possible for a human observer to achieve some kind of fusion (Julesz 1971, p.88). Two different types of decorrelation were tested. In the first type, increasing percentages of the dots were decorrelated in the left image at random. In particular, the cases of 10%, 20% and 30% were tried, and are illustrated in Figures 4.9 and 4.10. For the 10% case, (table entry 90% Corr) it can be seen that the algorithm was still able to obtain a good matching of the two planes, although the total number of zero-crossings assigned a disparity decreased, and the percentage of incorrectly matched points increased. When the percentage of decorrelated dots was increased to 20% (table entry 80% Corr), the number of matched points decreased again, although the percentage of those incorrectly matched remained about the same. Finally, when

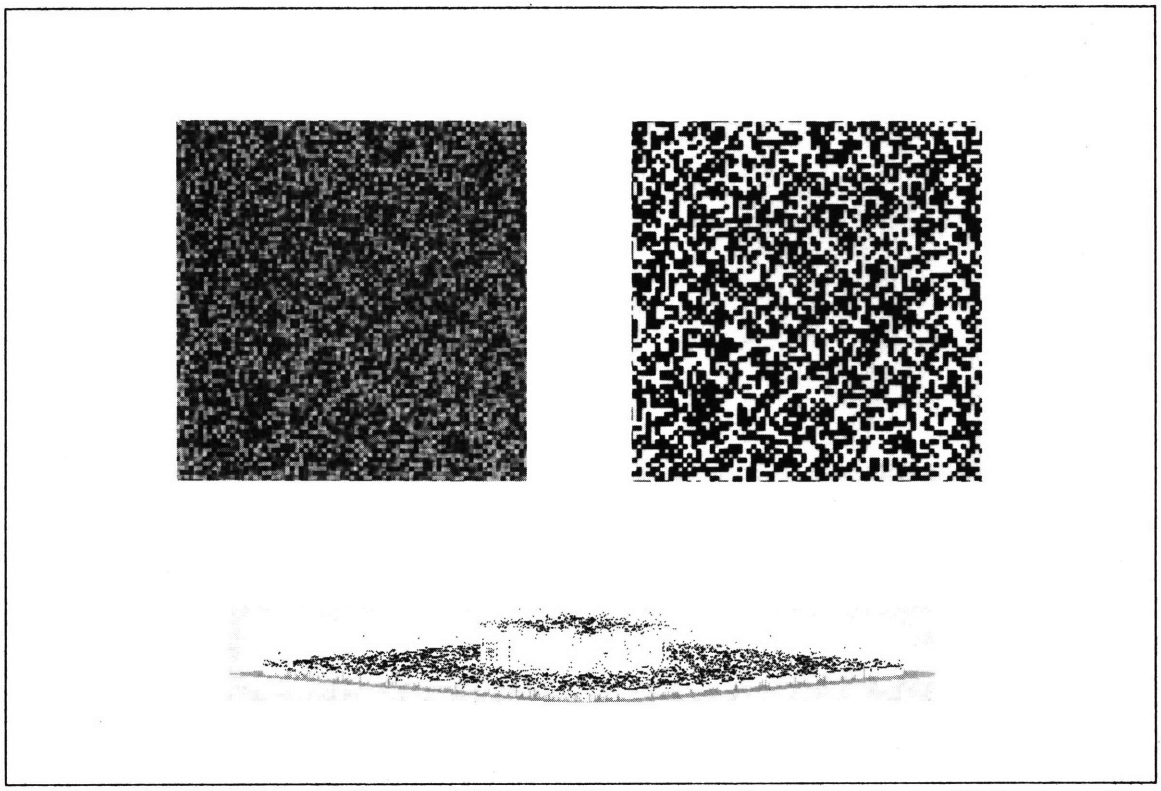


Figure 4.6. Stereogram with Filtered Noise. High pass filtered noise has been added to one of the images. Virtually no disparity values are obtained for the smallest channel. The disparity map in the figure is that of the next largest channel.

the percentage of decorrelated dots was increased to 30% (table entry 70% Corr), the algorithm found virtually no section of the image which could be fused.

The failure of the algorithm to match the 30% decorrelated pattern is caused by the component of the algorithm which checks that each region of the image is within range of correspondence. Recall that in order to distinguish between the case of two images beyond range of fusion (for the current eye positions) which will have only randomly matching zero-crossings, and the case of two images within range of fusion, the theory requires that the percentage of unmatched points be less than approximately 0.3. For the case of the pattern with 30% decorrelation, each region of the image will, on the average, have roughly 30% of its zero-crossings with no match and the algorithm will decide that the region is out of matching range. Thus, the algorithm cannot distinguish a correctly matched region of a degraded pattern from the matches that would be made between two random patterns. Hence, no disparities will be accepted for the region. It is interesting to note that many

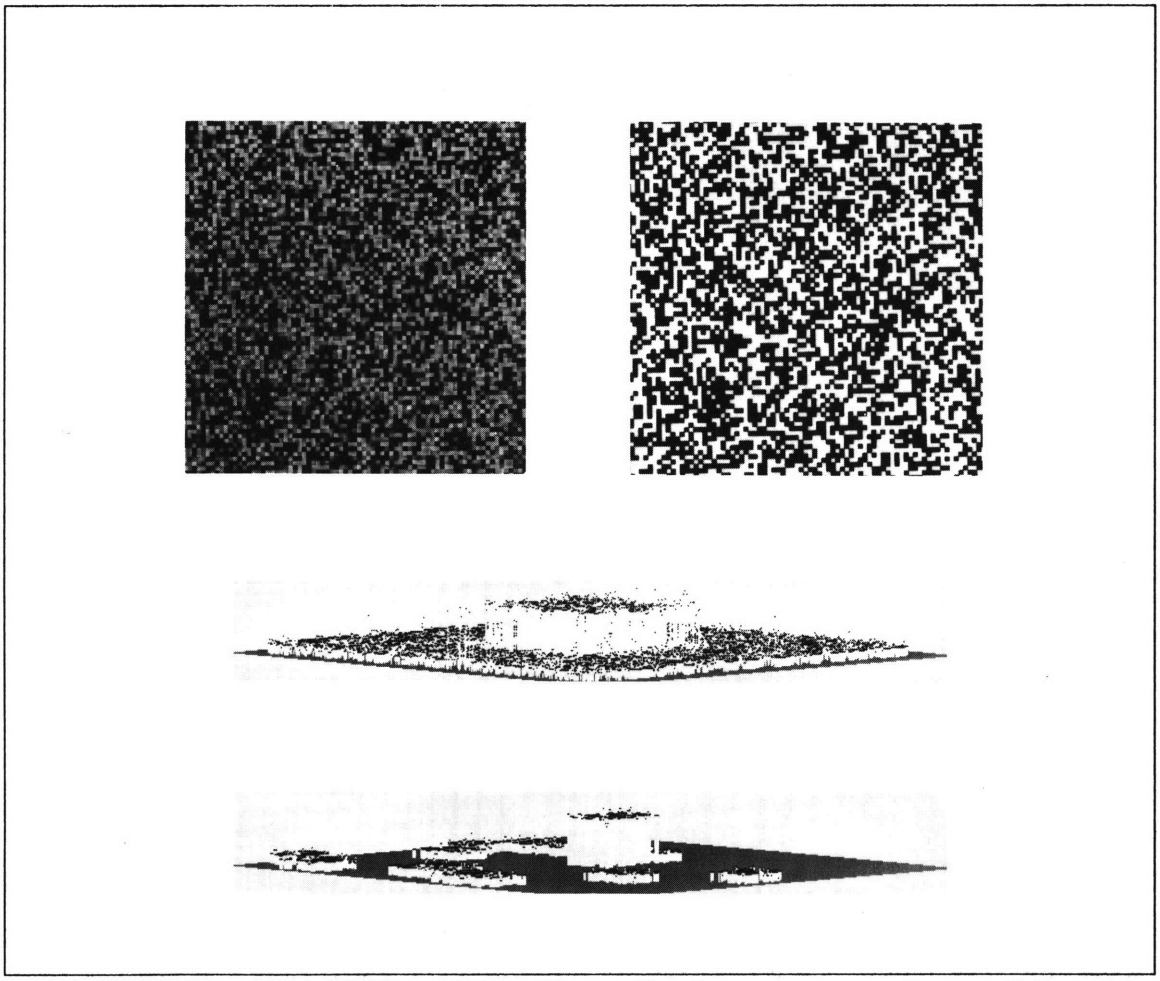


Figure 4.7. Stereogram with Filtered Noise. The top disparity map represents the $w = 9$ channel. The lower disparity map represents the $w = 4$ channel.

human subjects can achieve some kind of fusion up to about 20% decorrelation, the fusion becoming weaker as the decorrelation increases, being eliminated for patterns with 30% decorrelation.

Fascinatingly, one can also decorrelate the pattern by breaking up all white triplets along one set of diagonals, and all black triplets along the other set of diagonals (Julesz 1971, p.87). The table entry Diag Corr indicates the matching statistics for this case. Again, it can be seen that the program still obtains a good match, as do human observers. The performance of the algorithm is illustrated in Figure 4.11. This is a particularly fascinating example, since at first glance it would appear extremely unlikely that the two patterns could be fused. Yet the program is quite consistent with human perception on this case, obtaining a good matching of the two images.

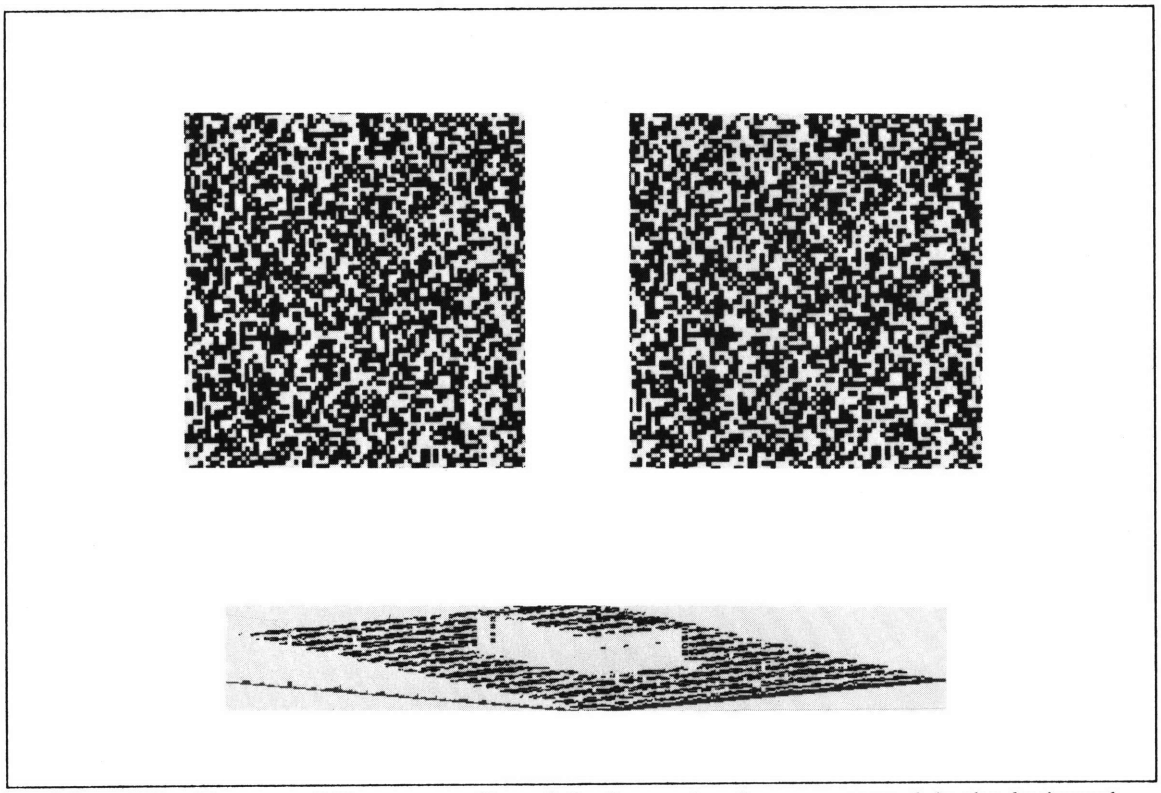


Figure 4.8. Compressed Stereogram. One of the images has been compressed in the horizontal direction. Fusion is still obtained, although now the planes appear to be slightly slanted.

4.2 Natural Images

The algorithm was also tested on some natural images. In such cases, an exact evaluation of the performance of the algorithm is difficult. A qualitative comparison is however possible, and the results of the algorithm may be seen in the Figures 4.12 and 4.13.

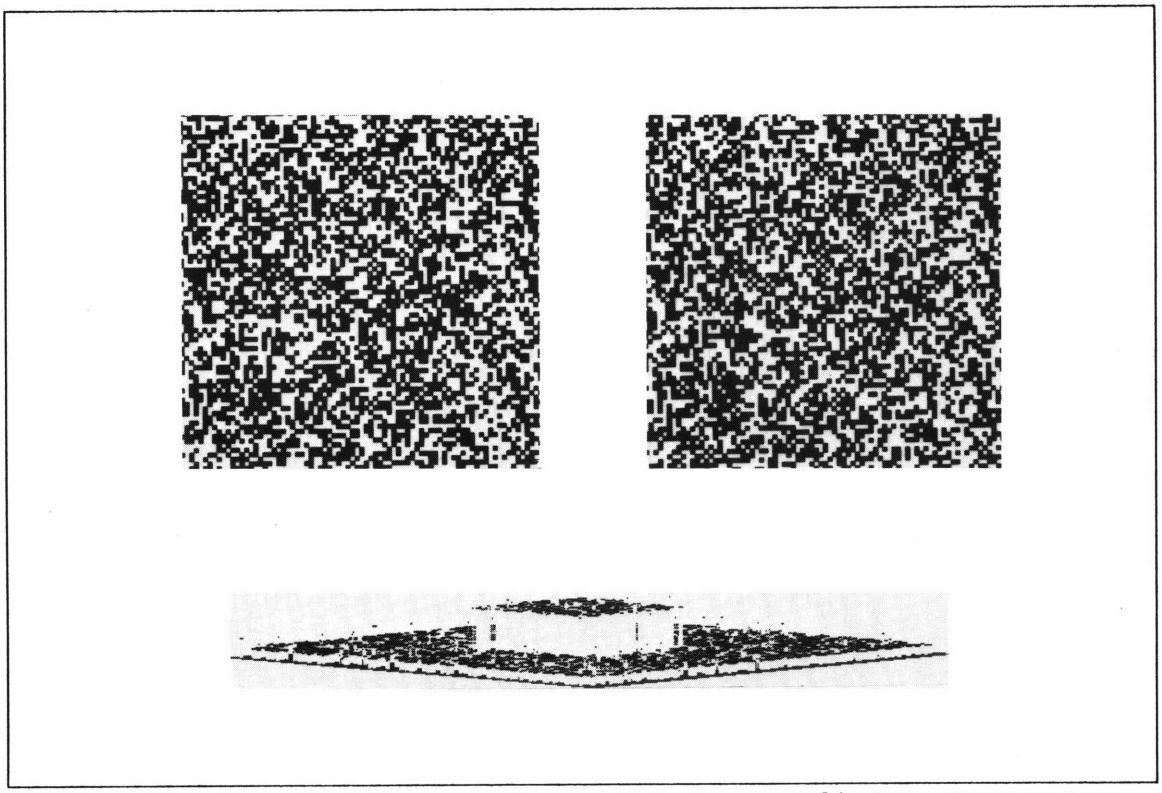


Figure 4.9. Decorrelated Stereogram. In one of the images, 10% of the dots have been decorrelated at random. Fusion is still obtained.

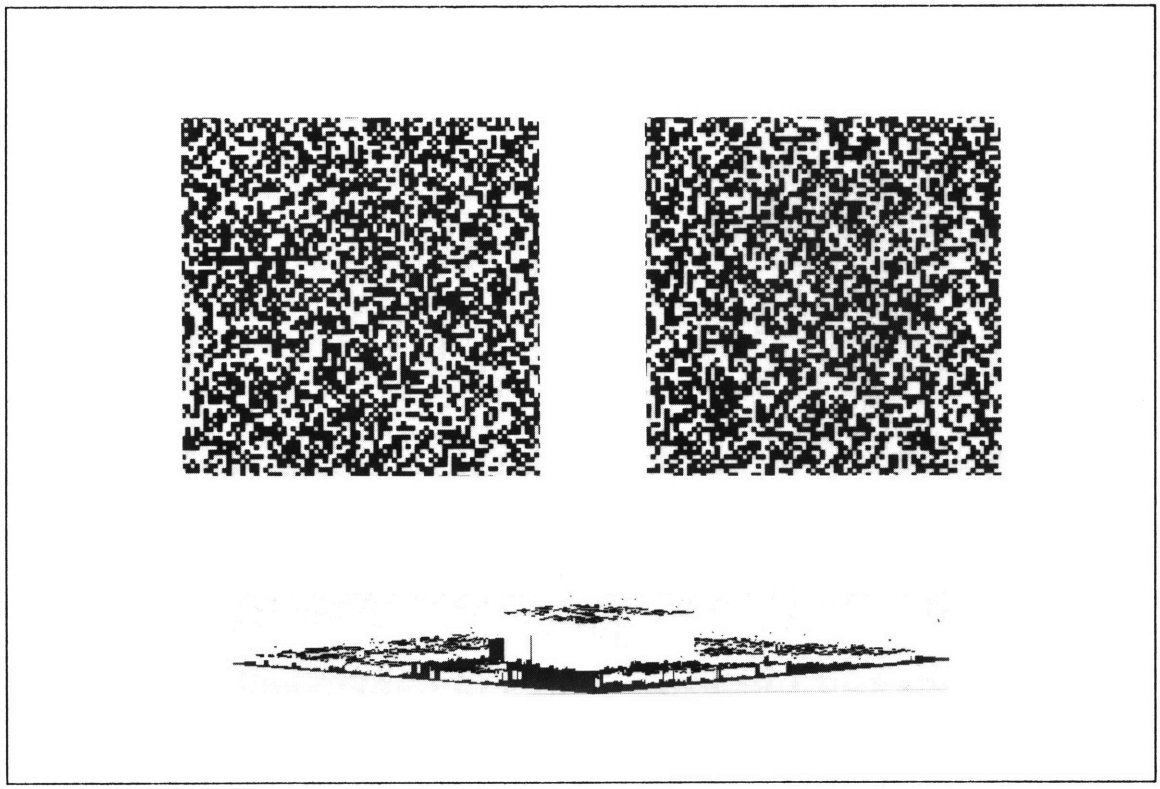


Figure 4.10. Decorrelated Stereogram. In one of the images, 20% of the dots have been decorrelated at random. Fusion is still obtained, although in this case, there are many regions of the image where no disparity values are assigned.

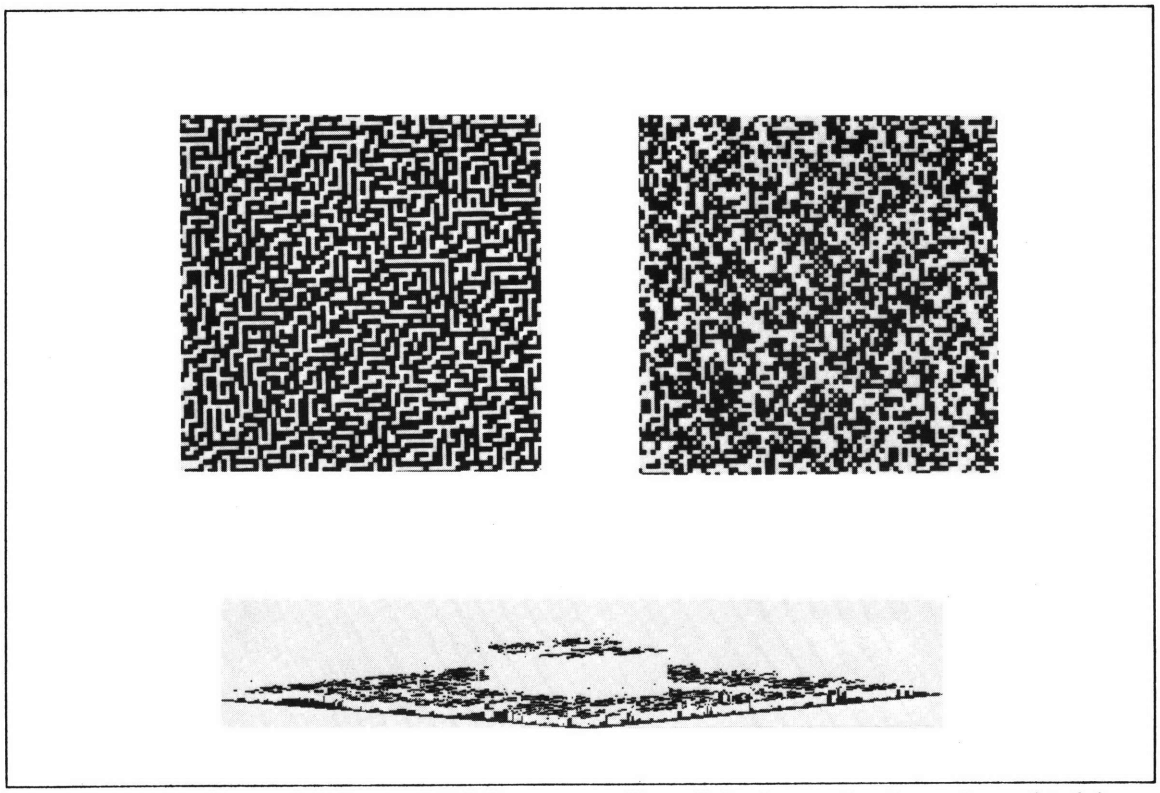


Figure 4.11. Diagonally Decorrelated Stereogram. One of the images has been decorrelated in the following manner. Along all the diagonals in one direction, any white triples have been broken by the insertion of a black dot. Along the diagonals in the other direction, any black triples have been broken by the insertion of a white dot. Fusion is still obtained.

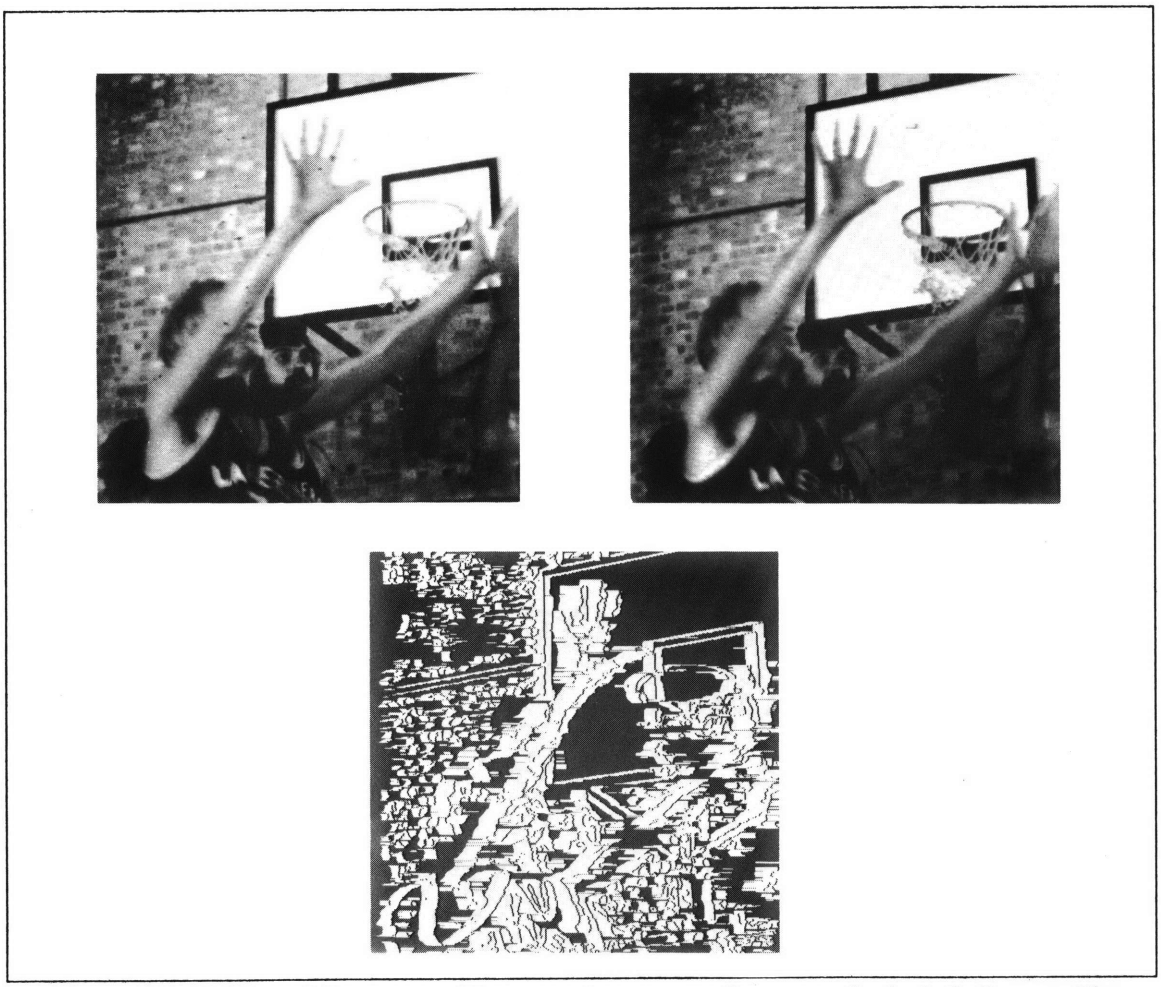


Figure 4.12. A Natural Stereogram. The stereo images are of a scene of a basketball game. The disparity map is represented in such a manner that the width of the black bars, terminated by a white dot, correspond to the disparity of the point. It can be seen that the disparity values are all qualitatively correct, with the arm of the foremost player emerging from the background of the basket and the wall. The images were 480 pixels on a side.

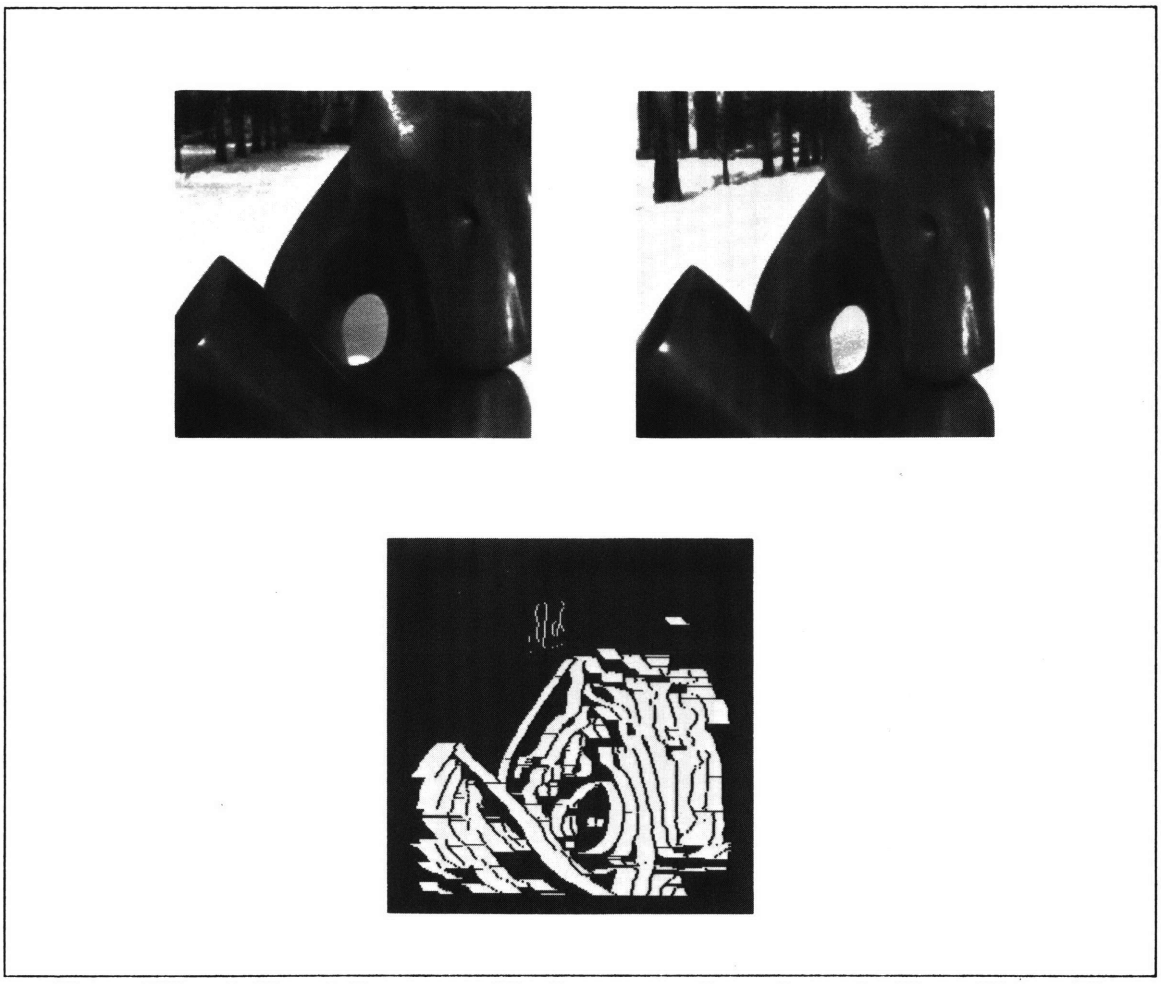


Figure 4.13. A Natural Stereogram. The images are of a sculpture by Henry Moore. The disparity array is represented as in Figure 4.12. It can be seen that the disparity values obtained by the program roughly correspond to the shape of the surface. The images were 320 pixels on a side.

TABLE OF STATISTICS			
parameter	worst case behavior, expected	without orientation, empirical	with orientation, empirical
average distance between zero-crossings of same sign	$2w$	$1.85w$	$6.56w$
probability of candidates in at most one pool	$> .50$	.38	.81
probability of candidates in two pools	$< .45$	.60	.19
probability of candidates in all three pools	$< .05$	.02	.001
given a candidate near zero, probability of no other candidates	$> .9$	.75	.93

Table 2.

4.3 Statistics

The theory of the Marr-Poggio matcher is based on an assumption concerning the distribution of intervals between zero-crossings. This leads to assumptions concerning the worst possible occurrences of false targets. The empirical occurrence of false targets has been measured in random dot patterns and the worst occurrences of false targets are indicated in Table 2. The theoretical worst case bounds used by Marr and Poggio appear for comparison.

From the table, it can be seen that the assumptions of the Marr-Poggio are not very reliable, compared with the empirical statistics found for random dot patterns. This is in part due to the fact that the analysis performed by Marr and Poggio was based on their assumption of oriented filters. The current implementation uses non-oriented filters to obtain the zero-crossings and uses orientation as a matching criterion after the descriptions have been obtained. It then becomes of interest to check whether a proper accounting of the use of non-oriented filters will make statistical predictions more consistent with the empirically observed statistics.

For the case of non-oriented filters, the derivation is very similar to that used by Marr and Poggio (1979). Assume that $f(x, y) = \nabla^2 G * I(x, y)$ is a white Gaussian process, where $I(x, y)$ is the image intensity. The problem is to find the distribution of intervals between alternate zero-crossings, taken along a horizontal slice of the image.

Assume that there is a zero-crossing at the origin, and let $P_1(\tau), P_2(\tau)$ be the probability densities of the distances to the first and second zero-crossings. P_1 and P_2 are approximated by the following formulae (Rice 1945, section 3.4; Longuet-Higgins 1962, equations 1.2.1 and 1.2.3; Leadbetter 1969).

$$P_1(\tau) = \frac{1}{2\pi}\sqrt{\frac{\psi(0)}{-\psi''(0)}}\frac{M_{23}(\tau)}{H(\tau)}\Big(\psi^2(0) - \psi^2(\tau)\Big)\Big[1 + H(\tau)\cot^{-1}(-H(\tau))\Big],$$

$$P_2(\tau) = \frac{1}{2\pi}\sqrt{\frac{\psi(0)}{-\psi''(0)}}\frac{M_{23}(\tau)}{H(\tau)}\Big(\psi^2(0) - \psi^2(\tau)\Big)\Big[1 - H(\tau)\cot^{-1}(H(\tau))\Big],$$

where $\psi(\tau)$ is the autocorrelation of the filter $\nabla^2 G$, a prime denotes differentiation with respect to τ, and

$$H(\tau) = \frac{M_{23}(\tau)}{\sqrt{M_{22}(\tau) - M_{23}(\tau)}},$$

$$M_{22}(\tau) = -\psi''(0)\Big(\psi^2(0) - \psi^2(\tau)\Big) - \psi(0)\psi'^2(\tau),$$

$$M_{23}(\tau) = \psi''(\tau)\Big(\psi^2(0) - \psi^2(\tau)\Big) + \psi(\tau)\psi'^2(\tau).$$

It is now necessary to compute the autocorrelation $\psi(\tau)$. The filter is given by

$$\nabla^2 G(r) = \left[\frac{r^2 - 2\sigma^2}{\sigma^4}\right]\exp^{-\frac{r^2}{2\sigma^2}}.$$

The two-dimensional Fourier transform of this filter is given by

$$F(w) = -2\pi\sigma^2\omega^2\exp^{-\frac{\sigma^2\omega^2}{2}}.$$

Since the interest is in the distribution of zero-crossing intervals along a horizontal slice of the image, it is necessary to obtain the portion of the spectrum corresponding to that slice. This is done by projecting the two-dimensional spectrum onto a line of similar orientation through the origin (Mersereau and Oppenheim, 1974). This compression of the spectrum yields

$$F_1(u) = -(2\pi)^{\frac{3}{2}}\sigma\left(\frac{1 + \sigma^2u^2}{\sigma}\right)\exp^{-\frac{\sigma^2u^2}{2}}.$$

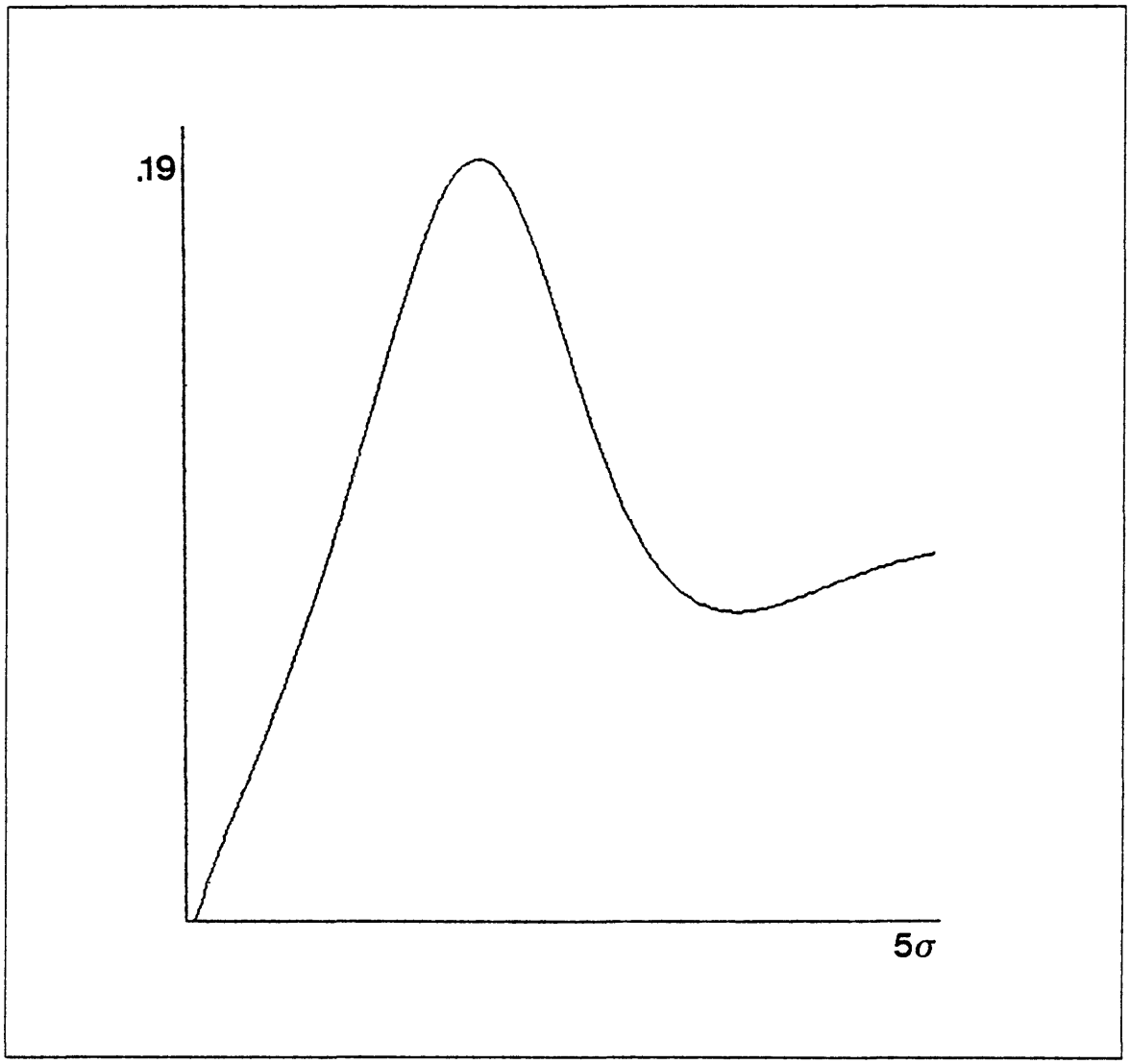


Figure 4.14. Probability Distribution of First Zero-Crossing. This is a graph of the probability of reaching a zero-crossing within a certain distance, given a zero-crossing at the origin.

The power spectrum is the given by

$$F_1^2(u) = (2\pi)^3\left(\frac{1+2\sigma^2u^2+\sigma^4u^4}{\sigma^2}\right)\exp^{-\sigma^2u^2}.$$

Taking the inverse transform of the power spectrum yields the autocorrelation function

$$\psi(\tau) = \pi^{\frac{5}{2}}\sigma\left\{\frac{11}{\sigma} - 5\frac{\tau^2}{\sigma^3} + \frac{1}{4}\frac{\tau^4}{\sigma^5}\right\}\exp^{-\frac{\tau^2}{4\sigma^2}}.$$

The formulae of Rice may then be applied to this autocorrelation function, and the probability distributions obtained in this way are shown in Figures 4.14 and 4.15.

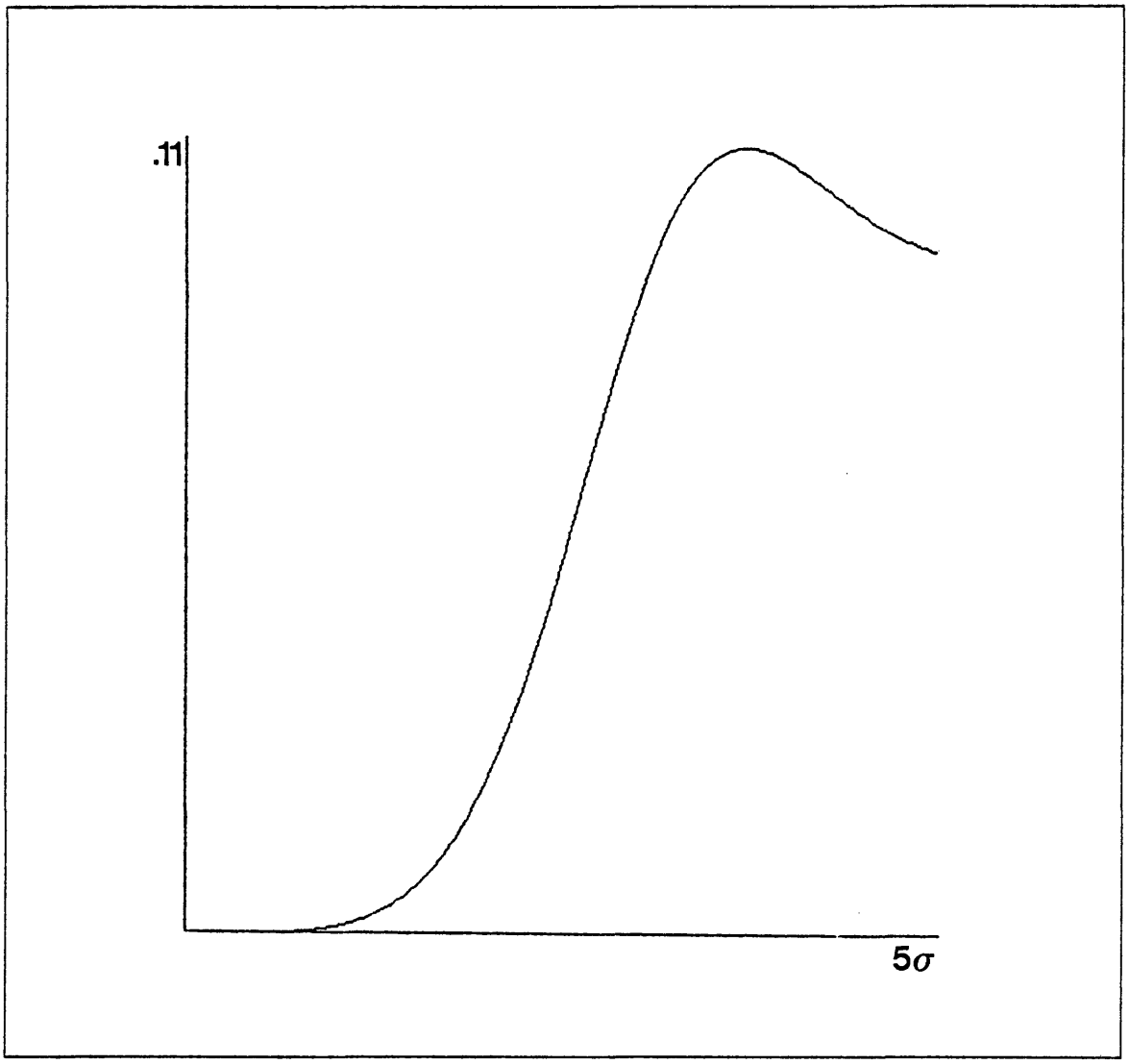


Figure 4.15. Probability Distribution of Second Zero-Crossing. This is a graph of the probability of reaching a second zero-crossing within a certain distance, given a zero-crossing at the origin.

In this case, the expected worst case behavior of multiple zero-crossings within a particular range of matching is somewhat different. In particular, the predicted density of zero-crossings is somewhat higher than in the Marr-Poggio case. At first sight, this would seem to suggest that the situation is worse than that given by the Marr-Poggio analysis. The use of orientation as a matching criterion, however, has yet to be included. To do this, some estimate of the distribution of orientation of zero-crossings is needed. The matching algorithm segments the orientation distribution into blocks of 30 degrees. The simplest estimate is given by assuming that the orientations are uniformly distributed,

in which case, the probabilities given by the statistical analysis should be adjusted by a factor of $\frac{1}{6}$. However, there is no real justification for assuming that the orientations will be uniformly distributed. In fact, the distribution tends to be more strongly weighted towards vertical orientations. Hence, rather than adjusting by a factor of $\frac{1}{6}$, a more pessimistic and factor of $\frac{1}{3}$ will be used.

TABLE OF STATISTICS				
parameter	without orientation, expected	without orientation, empirical	with orientation, expected	with orientation, empirical
average distance between zero-crossings of same sign	5.29σ	5.24σ	15.87σ	18.56σ
probability of candidates in at most one pool	$> .35$	.38	$> .78$	.81
probability of candidates in two pools	$< .64$	.60	$< .21$	.19
probability of candidates in all three pools	$< .01$	.02	$< .01$	.001
given a candidate near zero, probability of no other candidates	$> .81$	.75	$> .94$	.93

Table 3.

A table comparing predicted and empirical statistics of the distribution of similar zero-crossings is shown in Table 3. A number of interesting comparisons can be made from this. First, consider the case in which orientation is not used as a matching criterion. The number of multiple targets predicted by the formula of Rice agrees well with the empirical statistics found in practice. For example, if the range of matching neighbourhood is taken as $\pm 2.044\sigma$, Rice's formula predicts a worst case probability of 0.36 double targets. The empirical statistics in this case are 0.33. If the range is extended to $\pm 2.36\sigma$, Rice's formula predicts a worst case probability of 0.49 and the empirical statistics are 0.44. For a range of $\pm 2.84\sigma$, the prediction probability is 0.65 and the empiricial statistic is 0.62.

Second, the use of orientation as a matching criteria can greatly improve the problem of false targets. For example, let w_{1_d} be the central panel width of the one-dimensional projection of the $\nabla^2 G$ operator. This is related to the central panel width w_{2_d} of the two-dimensional operator by

$$w_{2\_d} = \sqrt{2} \cdot w_{1\_d}$$

and these are related to the space constant σ of the operator by

$$w_{2\_d} = 2\sqrt{2}\sigma.$$

If the matching range has the value $\pm w_{1\_d}$, then using orientation to within 30 degrees as a matching criterion reduces the percentage of false targets to 0.091. Even for a matching range of $\pm w_{2\_d}$, the percentage of false targets rises only to 0.20.

This raises an interesting possibility concerning the range of the matching neighbourhood. In the original theory, Marr and Poggio (1979) considered the possiblity of avoiding the false targets problem almost entirely by reducing the probability of its occurence, while maintaining a range of matching consistent with estimates of the size of Panum's area. According to their analysis, however, if the matching range is so restricted as to reduce the probability of false targets to less than 0.05, the range is too small by a factor of 2. If the range of matching is adjusted to account for the size of Panum's area, then the probability of false targets rises to 0.50 and it is necessary to introduce a disambiguation mechanism to resolve the false targets problem.

Given the statistical analysis derived above, one can propose a matching mechanism in which the zero-crossings are obtained from non-oriented filters and the orientation of the zero-crossing is used as a criterion for matching. In this case, matching over a range consistent with Panum's area will result in very few false targets (on the order of 0.10), and there is no need to introduce a disambiguation mechanism.

4.4 Discussion

Implementing a computational theory offers the opportunity to test its adequacy. In this case, I have found that the performance of the implementation coincides well with that of human subjects over a broad range of random dot test cases obtained from the literature, including defocussing, compression, and the introduction of various kinds of masking noise to one image of a random dot stereo pair.

DISCUSSION

In running the program, a number of interesting points concerning the form the algorithm have arisen. These are discussed below.

4.4.1 Pool Responses

The neighbourhood over which a search for a matching zero-crossing is conducted is broken into three pools, corresponding to convergent, divergent and zero disparity. In the present implementation, the pools are used to deal with the ambiguous case of two matching zero-crossings, while the disparity values associated with a match are represented to within an image element. A second possibility is to use the pools not only to disambiguate multiple matches, but also to assign a disparity to a match. Thus, a single disparity value, equal to the disparity value of the midpoint of the pool, would be assigned for a matching zero-crossing lying anywhere within the pool. In this scheme, only three possible disparities could be assigned to a zero-crossing: zero, corresponding to the middle pool; or $\frac{+}{2}^{w}$, corresponding to the divergent or convergent pools.

Note that under this type of scheme, the role of a finely tuned zero pool becomes more important. In the current implementation, in which disparity values are assigned exactly, there is no obvious need for a zero-tuned pool — it is not really necessary for the disambiguation of multiple matches. However, if each pool can only assign a constant disparity value, then a finely tuned zero pool is very useful in providing fine disparity values, since the narrowness of the pool, as compared to the convergent and divergent pools, will provide finer information.

Interestingly, computer experiments show that either scheme will work. In the case of a single disparity value for each pool, the disparities assigned by the smallest channel are within an image element of those obtained using exact disparities for each match. This modification was tried on both natural images and random dot patterns, and suggests that the accuracy with which the pools represent the match is not a critical factor for the actual matching process.

4.4.2 Matching Errors

The points that were incorrectly matched in the test cases all lay along depth discontinuities. The major reason for this is connected with the occlusion of a region. Note that at any depth discontinuity, there will be an occluded region which is present in one image, but not the other. Any zero-crossings within that region cannot, of course, have a correct matching zero-crossing in the other region. There is, however, a certain probability of such a zero-crossing being matched incorrectly to a random

zero-crossing in the other image. In principle, the algorithm detects regions which are occluded, by checking the statistics of the number of unmatched zero-crossings, and using such results to mark all zero-crossing matches in the region as unknown. However, for a region which contains a depth discontinuity, only part of the region will have the above characteristics. Zero-crossings in the rest of the region will have a unique match. Thus, when the statistical check on the number of unmatched points is performed, it is possible for the entire region to be considered in range, and thus all matches, including the incorrect ones of the occluded region, will be accepted.

4.4.3 Depth Discontinuities

It is interesting to comment on the effect of depth discontinuities for the different sized filters. For random dot patterns, the zero-crossings obtained from the larger filters tend to outline blobs or clusters of dots. Thus in general, the positions of the zero-crossings do not correspond to single elements of the underlying image. Suppose the dot pattern consists of one plane separated in depth from a second plane. In such a case, one might well find a zero-crossing that at one end of the zero-crossing contour belongs to dots on the first plane, and at the other end of the contour to dots belonging to the second plane. Such zero-crossings will be assigned disparities that reflect, to within the resolution of the channel, the structure of the image. The zero-crossings lying between the two ends of the contour will, however, receive disparities that smoothly vary from one extreme to the other. The largest channel would thus not see a plane separated in depth from a second plane, but rather a smooth hump.

For the smaller filter this does not occur, as the zero-crossing contours tend to outline individual dots or connected groups of dots. Thus the disparities assigned are such that the dots tend to belong to one plane or the other and the final disparity map is one of two separated planes.

To achieve perfect results from stereo, it is probably necessary to include in the $2\frac{1}{2}$-dimensional sketch a way of dealing competently with discontinuities. In a later section, we shall discuss this issue. In this connnection, it is interesting to point out that when one looks at a 5% random-dot stereogram portraying a square in front of its background, one sees vivid subjective contours at its boundary, although the output of the matcher does not account for this.

4.4.4 Constraint Checking

An integral part of most computational theories, proposed as models of aspects of the human

visual system, is the use of computational constraints based on assumptions about the physical world (Marr and Poggio, 1979, Marr and Hildreth, 1980, Ullman, 1979a). The constraints so derived are critical in the formation of the computational theory, and in the design of an algorithm for solving the problem. An interesting question to raise is whether the algorithm explicitly checks that the constraints imposed by the theory are satisfied. For example, Ullman's rigidity constraint in the analysis of structure from motion is explicitly checked by his algorithm. For the case of the Marr-Poggio stereo theory, two constraints were used, uniqueness and continuity of disparity values. It is curious that in the algorithm used to solve the stereo problem, the continuity constraint is explicitly checked while the uniqueness constraint is not. Uniqueness of disparity is required in one direction of matching, since only those zero-crossing segments of the left image which have exactly one match in the right image are accepted. However, it may be the case that more than one element of the right image could be matched to an element of the left image. When matching from the right image to the left, the same is true. Note that one could easily alter the algorithm to include the checking of uniqueness, thereby retaining only those disparity values corresponding to zero-crossing segments with a unique disparity value when matched from both images. However, the evidence of Braddick, discussed in the next section, would indicate that this is not the case. Hence, in the Marr-Poggio stereo theory, although both the requirement of uniqueness and continuity are subsumed, only one of these two constraints is explicitly checked by the algorithm. The reason the other constraint is not checked is probably because it is physically very unlikely to be violated.

4.4.5 Representations

There are a number of questions concerning the form of the $2\frac{1}{2}$-D sketch, which have yet to be firmly answered. Some of these problems, and the results of experimentation with the implementation as it relates to them, are indicated below.

The first critical question concerns whether the sketch uses the coordinates of the scene or of the working arrays. In the first case, the coordinates of the sketch would be directly related to the coordinates of the arrays of the entire scene. The advantage of this is that since disparity information about the scene is extracted from several eye positions, the representation of the disparities over the entire scene can readily be updated. However, this advantage also raises a difficulty. In order to store this information into a buffer with coordinate system connected to the image of the scene, explicit

information about the positions of the eyes is required. This is fine computationally, but for a model of the human visual system, it may be that such information is not available to the stereo process.

In the second case, no such problem arises. Here, the coordinates of the sketch are directly related to the coordinates of the retinal images. Such a system is called retinocentric, as it reflects the current positions of the eyes. As such, it does not require explicit knowledge of the eye positions relative to some fixed coordinate system within the scene, and thus it seems to be the most natural representation. This then raises the question of how information about disparities in the scene are maintained across eye movements.

The second question concerns the use of a fovea. Different sections of the images are analyzed at different resolutions, for a given position of the optical axes. An important consequence of this is that the amount of buffer space required to store the disparity will vary widely in the visual field, being much greater for the fovea than for the periphery. This also suggests the use of a retinocentric representation, because if one used a frame that had already allowed for eye-movements, it would require foveal resolution everywhere. Not only does such a buffer waste space, but it does not agree with our own experience as perceivers. If it were so, one should be able to build up a perceptual impression of the world that was everywhere as detailed as it is at the center of the gaze, and this is clearly not the case.

The final point about the $2\frac{1}{2}$-D sketch is that it is intended as an intermediate representation of the current scene. It is important for such a representation to pass on its information to higher level processes as quickly as possible. Thus, it probably cannot wait for a representation to be built up over several positions of the eyes. Rather, it must be refreshed for each eye position.

All of these factors combine to suggest a refinement to the implementation, as outlined above. In particular, a retinocentric representation, which represents disparities with decreasing resolution as eccentricity increases, should be used.

For the cases illustrated in this thesis, the $2\frac{1}{2}$-D sketch was created by storing fine resolution disparity values into a representation with a coordinate system identical to that of the scene. As we have argued above, a second alternative is to store values from all channels into a retinocentric representation, using disparity values from the smaller channels where available, and the coarser disparities from the larger channels elsewhere. In this way, a disparity representation for a single fixation of

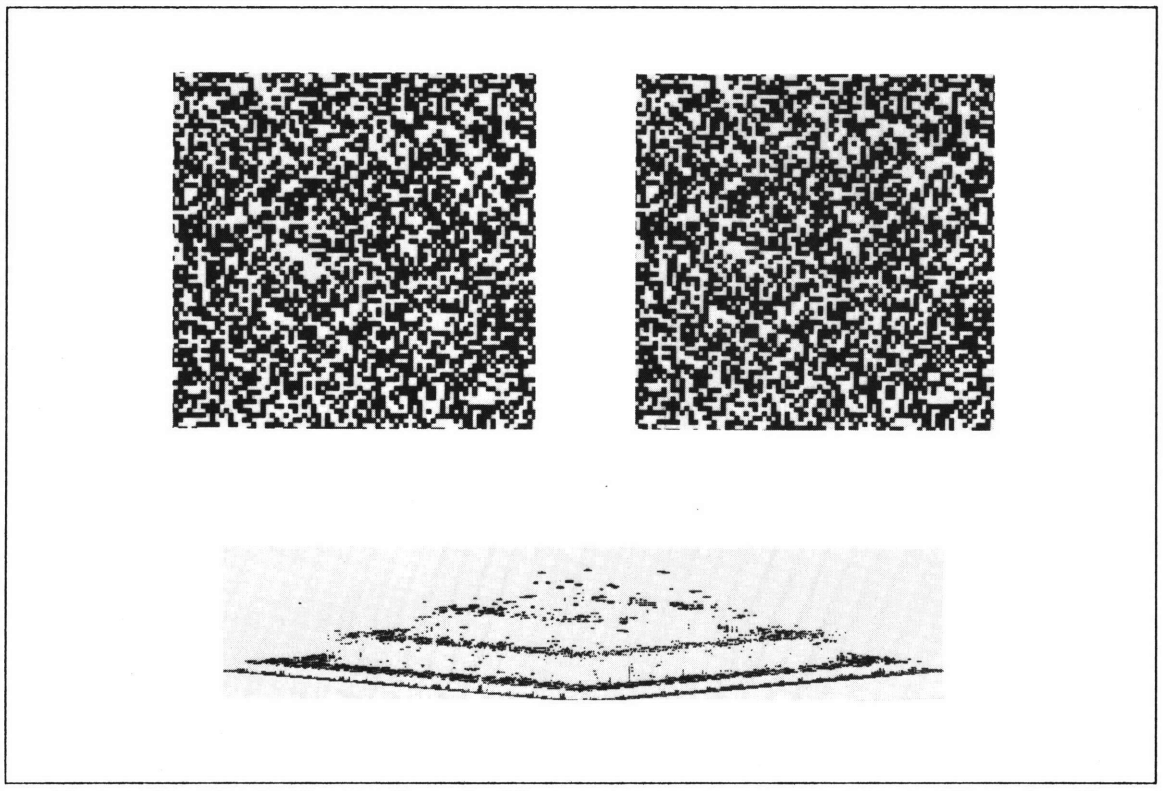


Figure 4.16. Single Fixation Example. The top images are a random dot wedding cake. The disparity map is that obtained by combining the different sized channels for a single fixation of the eyes. In this case, the eyes were fixated at the level of the bottom, outermost plane. It can be seen that the disparity values for the bottom plane are very sharp, since at this level the smallest channels are able to make the correspondences. For the region in the center of the image, the disparity values are much sparser and less accurate. This is since at this level, only the larger channels are within range of correspondence. The reader may check this perception by viewing the stereogram, while fixating only on the bottom plane.

the eyes may be constructed, with disparity resolution varying across the retina. Such a method of creating the $2\frac{1}{2}$-D sketch has been tested on the implementation, and is indicated in Figure 4.16.

4.4.6 Random Dot Patterns versus Natural Images

In the first part of this chapter, we have seen that the algorithm performs well on a wide range of random dot patterns, and in fact is consistent with human perception on that set of patterns. We have also seen that the algorithm can perform well on some natural images. However, there are some differences between using random dot patterns as input and using natural images as input, which can result in a difference in performance on natural images.

First, random dot patterns consist of synthesized intensity values, while the intensity values of natural images are subject to many factors of the imaging process. This can serve to add "noise" to the intensity values, and this in turn can affect the positions and orientations of the zero-crossings. Second, the vertical alignment of random dot patterns is a trivial matter, while for natural images, the vertical alignment is important. When random dot stereograms are synthesized, it is simple to ensure that there is exact vertical alignment between the elements. However, in natural images, this is not the case. Even if the two images of a natural scene are aligned vertically with respect to some object in the scene, the process of projection may cause other regions of the scene to be slightly misaligned vertically, unless the optical axes intersect.

The algorithm can be modified to account for this vertical deviation by allowing both horizontal and vertical alignment of the images to be controlled. That is, when the positions of the eyes are specified to the algorithm, they include a vertical as well as a horizontal displacement relative to one another. For example, suppose that the disparity values from one of the larger channels specify a particular horizontal alignment of the eyes. The algorithm will make this adjustment and match the zero-crossing descriptions of the smaller channels accordingly. If it is the case that the smaller channels do not obtain a match, it may be because of a slight vertical misalignment, and the matching is repeated for this horizontal adjustment, with a slight vertical alignment of the two images also taking place. In all the cases tested on the implementation, the total range of vertical deviation across the image was small, on the order of 2 or 3 image elements.

4.4.7 Failures of the Algorithm on Natural Images

In the first part of this chapter, I investigated the performance of the Marr-Poggio algorithm on a wide range of random dot patterns, and indicated that its performance was consistent with that of human perception. When turning to natural images, we have seen that the algorithm also seems to perform well. However, there are situations in which the algorithm can return disparity values inconsistent with other information in the image. The question is whether this reflects a basic error in the theory or its implementation, or whether there are other aspects of the visual process interacting with stereo which have not been accounted for in this implementation.

The results of testing the implementation on the broad range of images demonstrates that the matching module is acceptable as an independent module. In particular, the agreement between

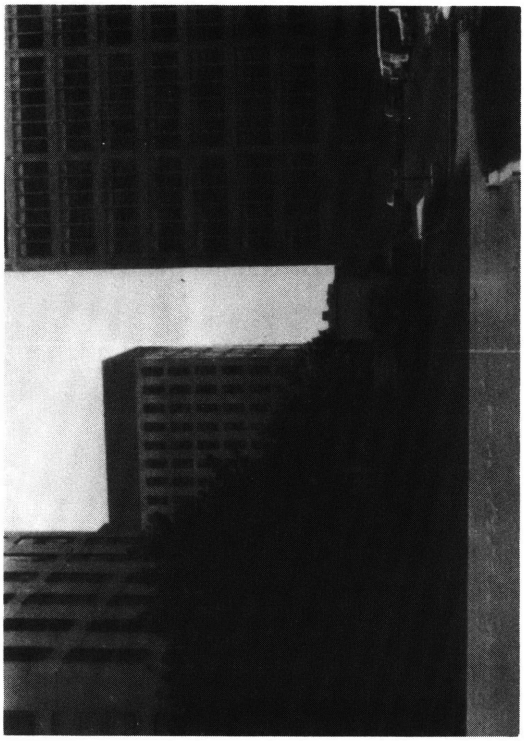
the performance of the algorithm and that of human observers on the many random dot patterns demonstrates that the matching module is acceptable, since in these cases, all other visual cues have been isolated from the matcher.

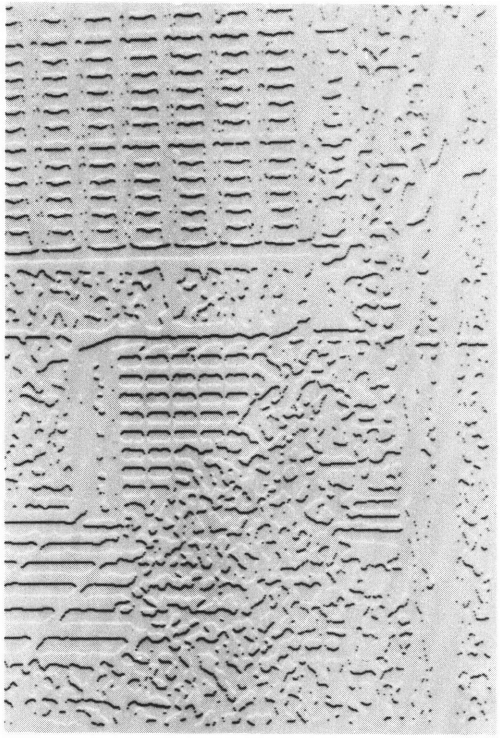
When turning to natural images, it is reasonable to expect that other visual modules may affect the input to the matcher and that they may alter the output of the matcher. For example, the evidence of Kidd, Frisby and Mayhew concerning the ability of texture boundaries to drive vergence eye movements indicates that other visual information besides disparity may alter the position of the eyes, and thus the input to the matcher. However, it does not necessarily imply that the theory of the matcher itself is incorrect.

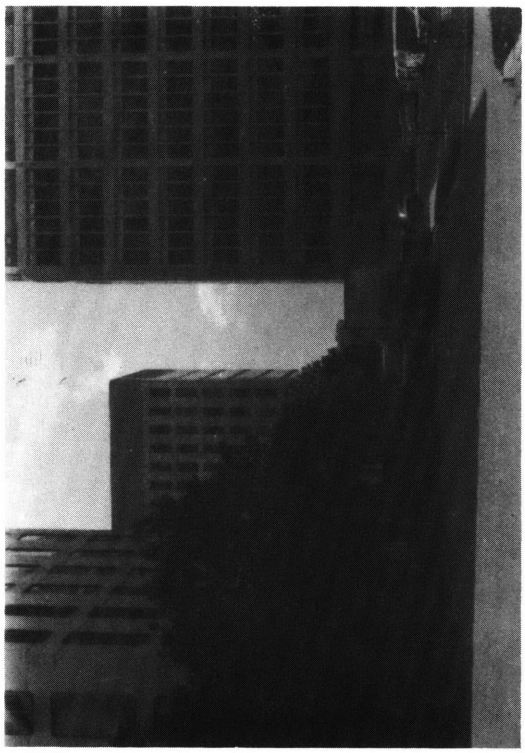
Interestingly, the performance of the implementation supports this point. The implementation, which is considered a distinct module, also performs very well on random dot patterns, where there is no possibility of interaction with other visual processes. For many natural images, this is still true. However, occasionally it is the case that a natural image provides some difficulty for the implementation. A particular example of this occurs in the image of Figure 4.17. Here, the regular pattern of the windows provides a strong false targets problem. In running the implementation, the following behavior was observed. If the initial vergence position was at the depth of the building, the zero-crossings corresponding to the windows were all assigned a correct disparity. If, however, the initial vergence position was at the depth of the trees in front of the building, the windows were assigned an incorrect disparity, due to the regular pattern of zero-crossings associated with them. Clearly, this seems wrong. Yet the question to ask is whether the implementation is wrong. Curiously, if one fuses the zero-crossing descriptions without eye movements, human observers have the same problem: if the eyes are fixated at the level of the building, all is well; if the eyes are fixated at the level of the trees, the windows are incorrectly matched. I would argue that this implies that the implementation, and hence the theory of the matching process is in fact correct. Given a particular set of zero-crossings, the module finds any acceptable match and writes it into a buffer. When the output of this buffer is sent to the $2\frac{1}{2}$-D sketch, it must be made consistent with other sources of information feeding the $2\frac{1}{2}$-D sketch. In this case, it is possible that some later processing module is capable of altering the disparity values, based on other information unavailable to the stereo process, and the correct depth is written into the $2\frac{1}{2}$-D sketch.

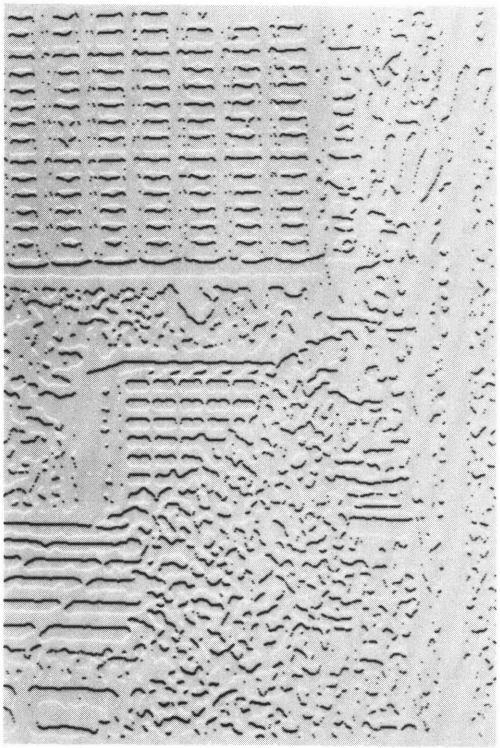
Thus, I would suggest that future refinements to the Marr-Poggio theory must account for the interactions of other aspects of visual information processing on the input and output of the matching module.

Figure 4.17. A Natural False Targets Problem. The zero-crossings corresponding to the windows of the furthest building form a regular pattern which accentuates the false target problem. The performance of the algorithm on these images is critically dependent on the positions of the eyes relative to the scene. If the fixation point is at the level of the building, the correct matching of the images takes place. If the fixation point is at the level of the trees in front of the building, then a "wallpaper" effect takes place, and the windows are matched incorrectly. Psychophysical experiments indicate that humans may have the same difficulty.









4.5 Development of the Implementation

We have indicated earlier that one reason for implementing a computational theory is that it offers us the opportunity to test the theory's adequacy.

A second reason for implementing a computational theory is that the implementation serves as a useful feedback device for the theory, indicating errors or omissions in the theory, as well as indicating areas whose difficulty had not been previously appreciated. Throughout the course of the development of the stereo implementation, a number of interesting observations were made. Some of these indicated equivalent methods of implementation which had interesting properties with regard to alternative theories of the process. Others served to correct assumptions made by the theory. Still other observations arose at surprising places, places where no difficulty was expected in the implementation process. In many of these cases, in finding a way around the problem, decisions of wide ranging effect were made. This is particularly true of the question of zero-crossings and the question of non-oriented filters. Thus, we shall see an example of a problem of implementation causing major changes in the theory of early visual processing. Without the act of implementation, such effects might not have been found. These observations are discussed in the following section.

4.5.1 From Which Image Do We Match?

Although the Marr-Poggio matcher is designed to match from one image into the other, there is no inherent reason why the matching process cannot be driven from both eyes independently. In fact, there may be some evidence that this is so, as is shown by the following experiment of O. Braddick (1978) on an extension to Panum's limiting case. First, a sparse random dot pattern was constructed. From this pattern, a partner was created by displacing the entire pattern by slight amounts to both the left and the right. Thus, for each dot in the right image, there corresponded two dots in the left image, one with a small displacement to the left and one with a small displacement to the right. The perception obtained by viewing such a random dot stereogram is one of two superimposed planes.

Suppose the matching process were driven from only one image, for example, from the right image to the left. In this case, the implementation would not be able to account for Braddick's results, since all the zero-crossings would have two possible candidates. However, suppose that the matching process were driven independently from both the right and left images, an unambiguous match from either side being accepted. In this case, although every zero-crossing in the right image would have an

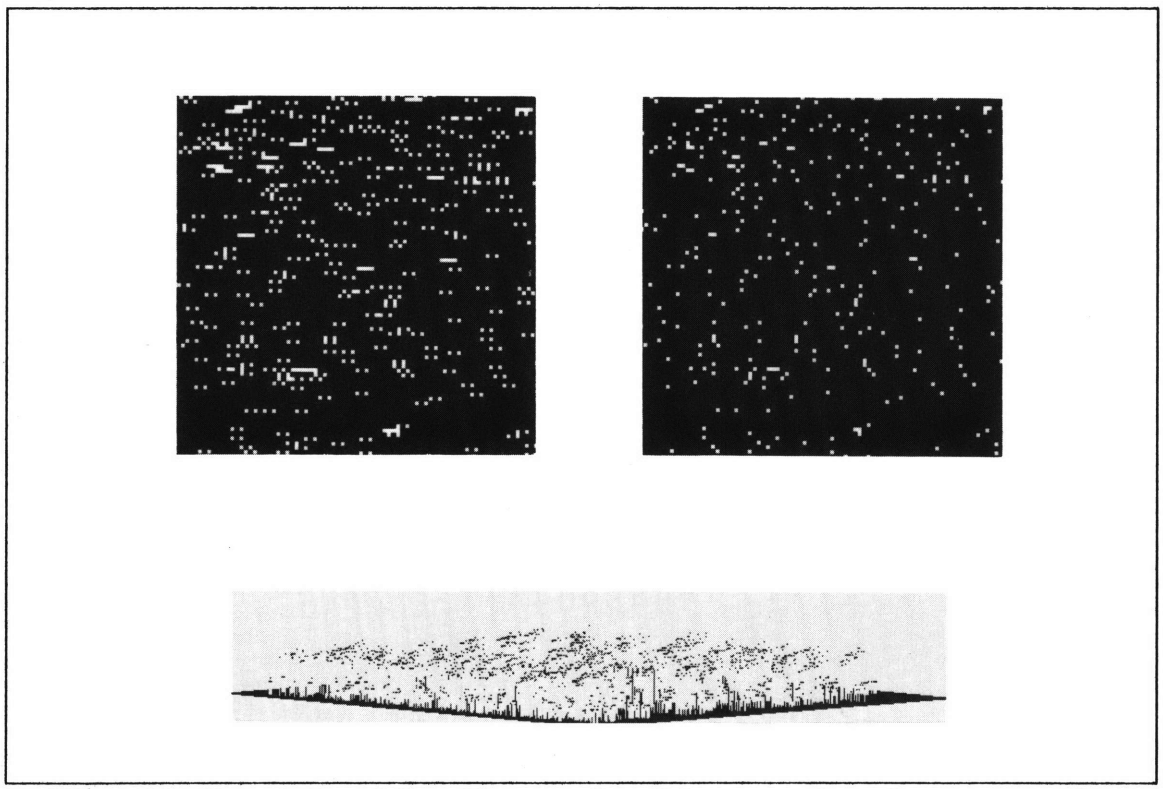


Figure 4.18. Panum's Special Case. The perception of the stereo pair is of a transparent plane above a second plane. The disparity map of the algorithm agrees with this.

ambiguous match, the program would obtain a unique match for each zero-crossing in the left image.

Braddick's case has been tested on the program, and the results are shown in Figure 4.18. It can be seen that the results of the implementation are that of two transparent planes, as in human perception.

An interesting idea is that there may be half stereo blind people who can see two planes when the images are presented such that the double image is in one eye, and who see none or one fuzzy plane when the double image is in the other eye.

4.5.2 Oriented Filters

Although this point has been extensively treated elsewhere (Marr and Hildreth 1979, Hildreth 1980), it is interesting to recount the historical development of the use of non-oriented filters.

In the original implementation, oriented filters were used rather than non-oriented ones, in part because the original Marr-Poggio theory was based on them. This was motivated by physiological

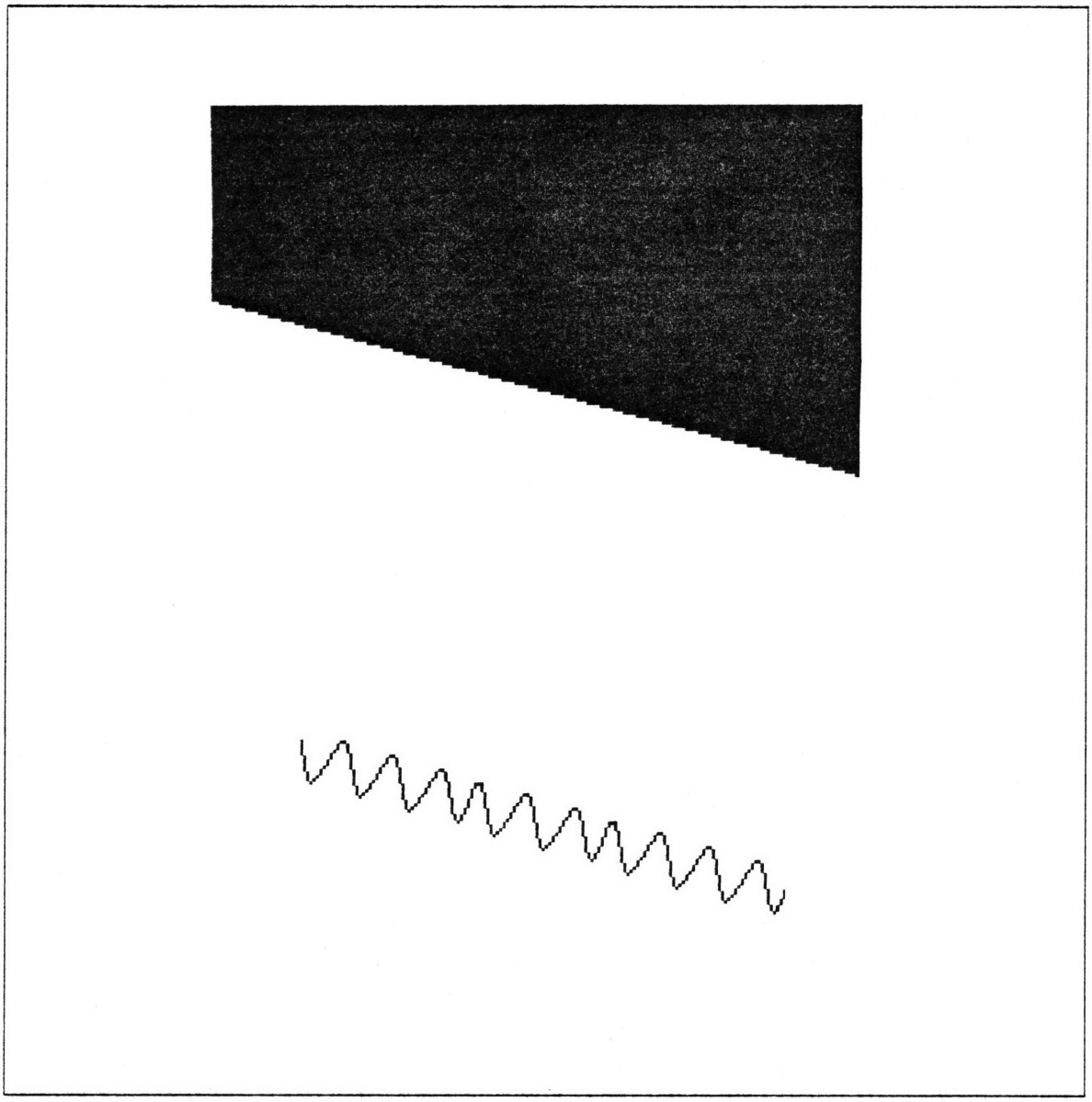


Figure 4.19. The Snake Effect. The ideal edge of the top figure was processed with an oriented filter whose orientation differed from that of the edge by 70 degrees. The zero-crossing in the bottom figure has the same overall orientation as the edge, but exhibits a wide fluctuation about this orientation.

considerations, but actual practise showed that there were severe difficulties with using such filters. Two effects were particularly noticeable. The first is that such bar-shaped filters tend to smear or stretch zero-crossings in the direction of the orientation of the filter relative to the actual edges in the image. For example, a bar filter which is oriented vertically tends to convert a circular object in a scene into an oval zero-crossing contour. This is not desirable, since any matching performed on such

zero-crossings will result in disparity values being assigned to locations in the image for which there is no evidence that such assignments are valid.

Some experiments with the aspect ratio of such bar shaped filters indicated, as might be expected, that the lower the aspect ratio, the smaller the amount of stretching of the zero-crossing contours.

A second effect concerned the case of an edge with an orientation different from that of the filter, illustrated in Figure 4.19. In this case, the resulting zero-crossings suffered from the following problem. Rather than giving a straight zero-crossing contour along the orientation of the edge, the filter produced a zero-crossing contour whose overall orientation was that of the edge, but which curved about the edge in a snake-like fashion. Thus, the resulting zero-crossing contour had a significant number of segments consisting of components in the direction of the filter orientation, rather than in the direction of the edge. Again, any matching process which assigns disparities based on such descriptors will be assigning disparity values which do not accurately reflect the structure of the underlying surface.

Both of these effects led to a questioning of the necessity for oriented filters, and in fact, such filters were replaced by circularly symmetric ones. A comprehensive analysis of non-oriented filters was developed by Marr & Hildreth (1979).

What is of interest here is the fact that attempts at implementing the early version of the stereo theory led to practical difficulties. In overcoming these problems, a major modification to the theory took place.

4.5.3 Statistics

We have already seen in Section 4.3 that the statistical analysis performed by Marr and Poggio is not consistent with the observed statistics of zero-crossings. This is due to the change in operators, from the oriented operators used by Marr and Poggio, to the non-oriented operators used in the implementation discussed here. By redoing the statistical analysis for non-oriented filters, I have been able to propose a modification to the matching algorithm which simplifies its operation.

4.5.4 Zero-Crossings

An earlier implementation of the theory did not use zero-crossings, but rather attempted to

create a symbolic description of the changes in an image using peaks of the convolved output. Several difficulties were encountered.

One difficulty concerned the side lobe effect. By this, I mean that even for an isolated edge, the convolved values would exhibit not only a peak corresponding to the edge, but a pair of side lobe peaks of the opposite sign. This not only made the matching task much harder, but also added to the symbolic description zero-crossings that did not reflect actual changes in the image intensities. Any matching based on such descriptors was therefore likely to assign disparity values which did not reflect the structure of the underlying part of the surface. Moreover, it was not possible to distinguish locally between *"real"* peak locations and *"false"* or side-lobe peaks.

A second difficulty concerned the difference between a one-dimensional peak and a two-dimensional peak. Since the matching takes place along horizontal slices of the convolved image, one could define a peak as any local extremum along that slice. However, the symbolic descriptions generated in this manner will differ greatly from those generated by locating local extrema in two-dimensions, where the point is required to be a local extremum along both axial directions. This is in contrast to the case of zero-crossings, where the zero-crossings generated by scanning along a horizontal direction are virtually identical to those generated by examining a two-dimensional neighbourhood. In fact, the zero-crossings not generated by the one-dimensional scan are not relevant to the stereo matching process, since they correspond to horizontally oriented zero-crossing segments, which do not have a precise disparity associated with them. Returning to the question of peaks, we see that if only the two-dimensional peaks are matched, the density of such features is much smaller than that of zero-crossings. On the other hand, matching of the one-dimensional peaks may lead to difficulties, since the locations of such peaks need not be sharply localized.

All of these difficulties led to the use of zero-crossings as a matching primitive rather than peaks.

This is a particularly interesting illustration of the role of an implementation in developing a computational theory. In this case, the implementation lead to questions about a particular aspect of the process whose resolution had wide reaching effects. Thus, the stereo implementation brought to light a problem which had not previously been considered, and the resolution of that problem has significantly altered the theory of several other processes, (for example, the theory of edge detection and the Primal Sketch, Marr and Hildreth, 1979).

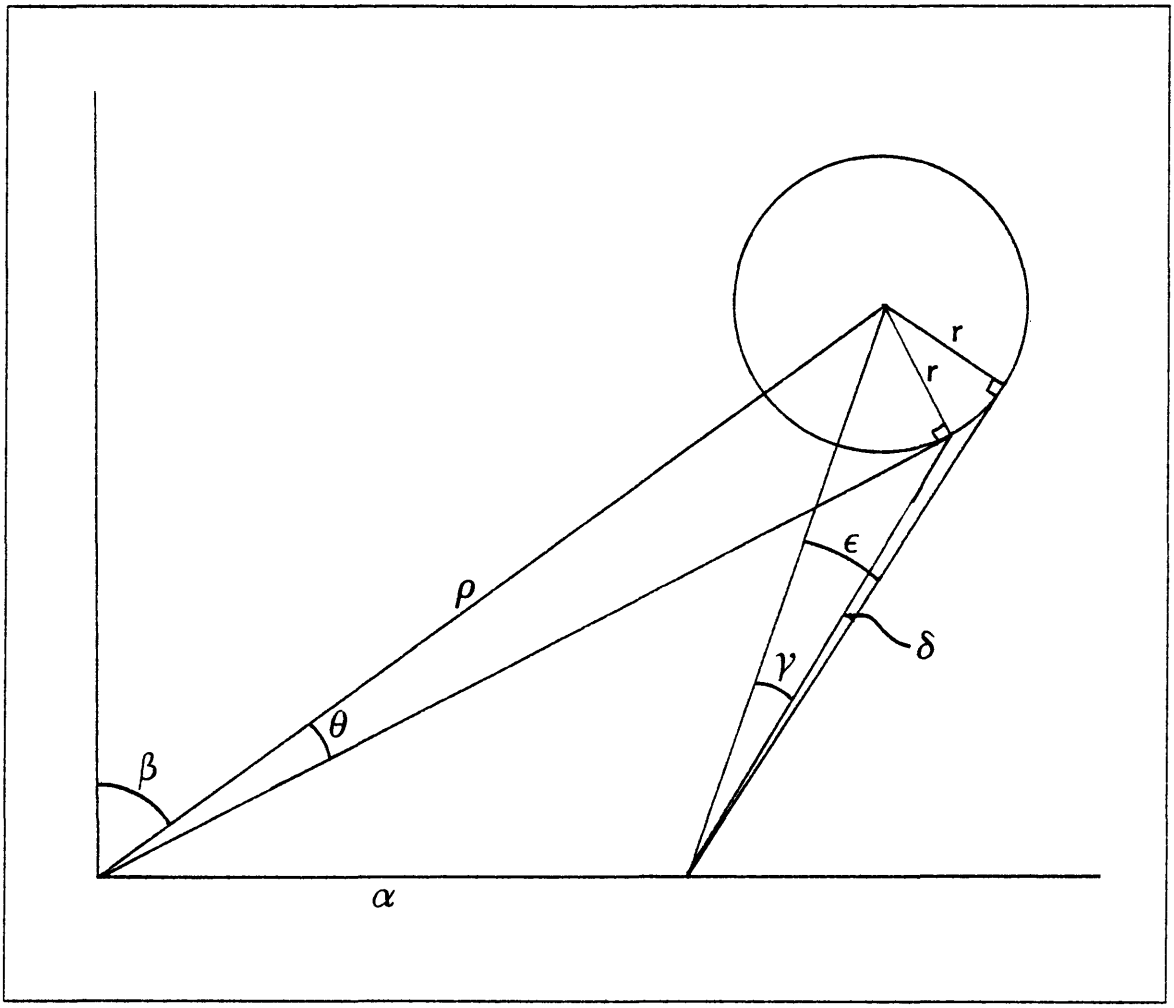


Figure 4.20. Geometry of Edge Effects. The angle δ measures the angular error in disparity associated with matching the occluding boundaries in both eyes.

4.6 Edge Effects

It should be noted at this point that the stereo theory has made an assumption concerning the relevance of zero-crossings. Recall the initial motivation for the use of zero-crossings. We wanted to match only those locations on a surface which gave rise to image properties that could be localized physically. Thus, surface scratches, texture markings, sharp reflectance changes all were characterized by strong local changes in intensity. Such changes in intensity will cause a zero-crossing in a second directional derivative, and it is exactly these descriptors which we have indicated will be matched by the stereo process. It is not apparent, however, that there is an isomorphism between zero-crossings in the convolved image and those object features whose position we want to match. In fact, there is one

major counter-example, namely occluding contours, or boundaries of objects. The design of the zero-crossing detector does not allow one to distinguish between zero-crossings caused by surface markings (photometric zero-crossings) and zero-crossings caused by changes in the gross shape of the surface (topographic zero-crossings). One must therefore ask whether one will be badly misled by allowing the stereo matcher to match topographic zero-crossings as well as photometric ones.

Consider a featureless cylinder, oriented vertically in space. The edges of the cylinder will in most situations cause a zero-crossing in the convolved image. However, the portion of the cylinder which will project to such a zero-crossing in the two eyes is different. It would seem that one does not want to match such descriptors since they do not correspond to the same physical location.

Let us consider this case carefully. The geometry of the situation is shown in Figure 4.20. I wish to determine the angle δ as a function of ρ, β, θ, and α, since δ corresponds to the difference in disparity between matching the correct point on the cylinder and matching the occluding contours as seen in each image. Trigonometric manipulation yields the following expressions, where $\epsilon = \gamma + \delta$,

$$\tan\epsilon = \frac{r}{\sqrt{\alpha^2 + \rho^2 - r^2 - 2\alpha\rho\sin\beta}}$$

$$\tan\gamma = \frac{r[\sqrt{\rho^2 - r^2} - \alpha\sin(\beta + \theta)]}{(\rho^2 + \alpha^2 - r^2) - 2\alpha\sqrt{\rho^2 - r^2}\sin(\beta + \theta) + r\alpha\cos(\beta + \theta)}$$

$$\tan\delta = \frac{\tan\epsilon - \tan\gamma}{1 + \tan\epsilon\tan\gamma}$$

Consider the special case of the cylinder being centered between the eyes, as shown in Figure 4.21, where the above expressions simplify. Note that in this case:

$$\sin\beta = \frac{\alpha}{2\rho}$$

$$\cos\beta = \frac{\sqrt{4\rho^2 - \alpha^2}}{2\rho}$$

$$\sin\theta = \frac{\sqrt{\rho^2 - r^2}}{\rho}$$

$$\cos\theta = \frac{r}{\rho}$$

so that the following expressions hold:

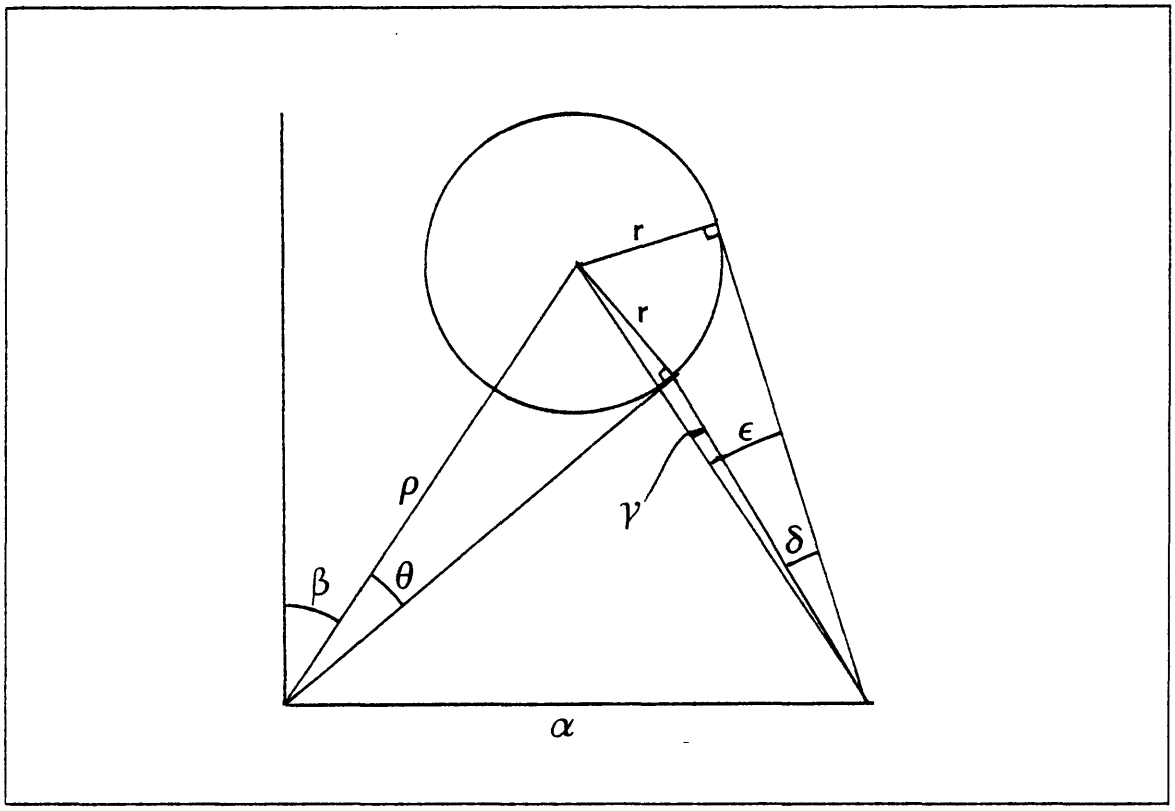


Figure 4.21. Special Case of Object Centered Between Eyes.

$$\tan\epsilon = \frac{r}{\sqrt{\rho^2 - r^2}}$$

$$\tan\gamma = \frac{r\left\{\sqrt{\rho^2 - r^2}(2\rho^2 - a^2) - ra\sqrt{4\rho^2 - a^2}\right\}}{2\rho^2(\rho^2 - r^2) - ra\sqrt{\rho^2 - r^2}\sqrt{4\rho^2 - a^2} + a^2r^2}$$

$$\tan\delta = \frac{ra^2}{2\rho^2\sqrt{\rho^2 - r^2} - ra\sqrt{4\rho^2 - a^2}}$$

If r and ρ are represented in units of the inter-ocular distance a, say $r = aa$ and $\rho = ba$, then algebraic manipulation yields:

$$a = \frac{2b^3}{\sqrt{\left(\cot\delta + \sqrt{4b^2 - 1}\right)^2 + 4b^4}}.$$

Thus, given some minimum acceptable angular error, for example $\delta < 0.5'$, one may graph the maximum allowed radius r of the cylinder as a function of the distance from the viewer, ρ. This

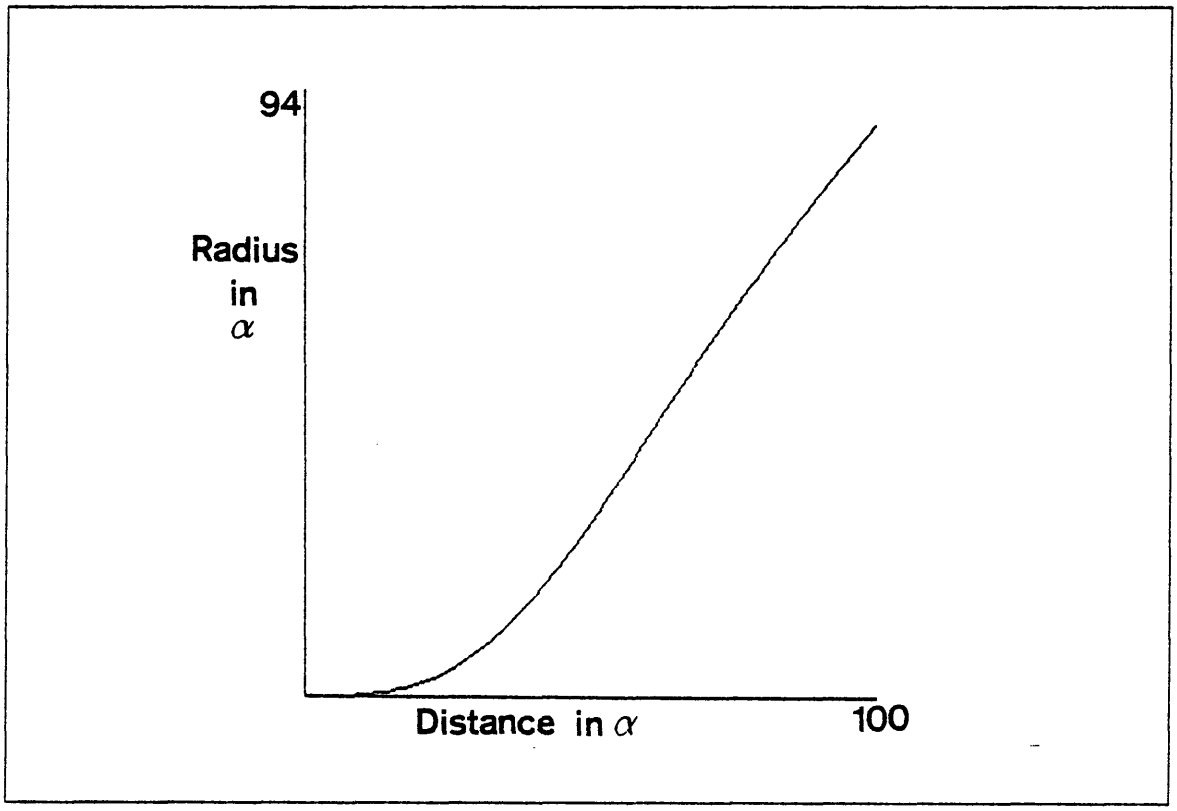


Figure 4.22. Graph of Effects of Matching Occluding Contours. The curve indicates the maximum radius of the cylinder, as a function of its distance from the viewer, which will cause an error in disparity of less than 0.5′ of arc.

particular case is graphed in Figure 4.22. It indicates that for almost all likely situations, the error in disparity associated with matching the occluding contours of a cylindrical object are negligible. At a distance from one eye of 1 inter-ocular unit, the maximum radius is 0.0003 inter-ocular units, a very tight constraint. For a distance of 10 inter-ocular units (roughly 2 feet), however, the maximum radius is 0.29 inter-ocular units (roughly 0.75 inches). This translates into cylinders of a diameter of roughly twice the width of a thumb, viewed at arm's length. In general, except for very small distances, the stereo algorithm will not introduce unacceptable errors in disparity when matching the zero-crossings corresponding to occluding boundaries.

There is an alternative method for estimating the error involved in matching occluding contours. Consider the geometry illustrated in Figure 4.23. The distance δ measures the error in the shape of the cylinder that would result from matching the occluding contours. Trigonometric manipulation yields the following expressions:

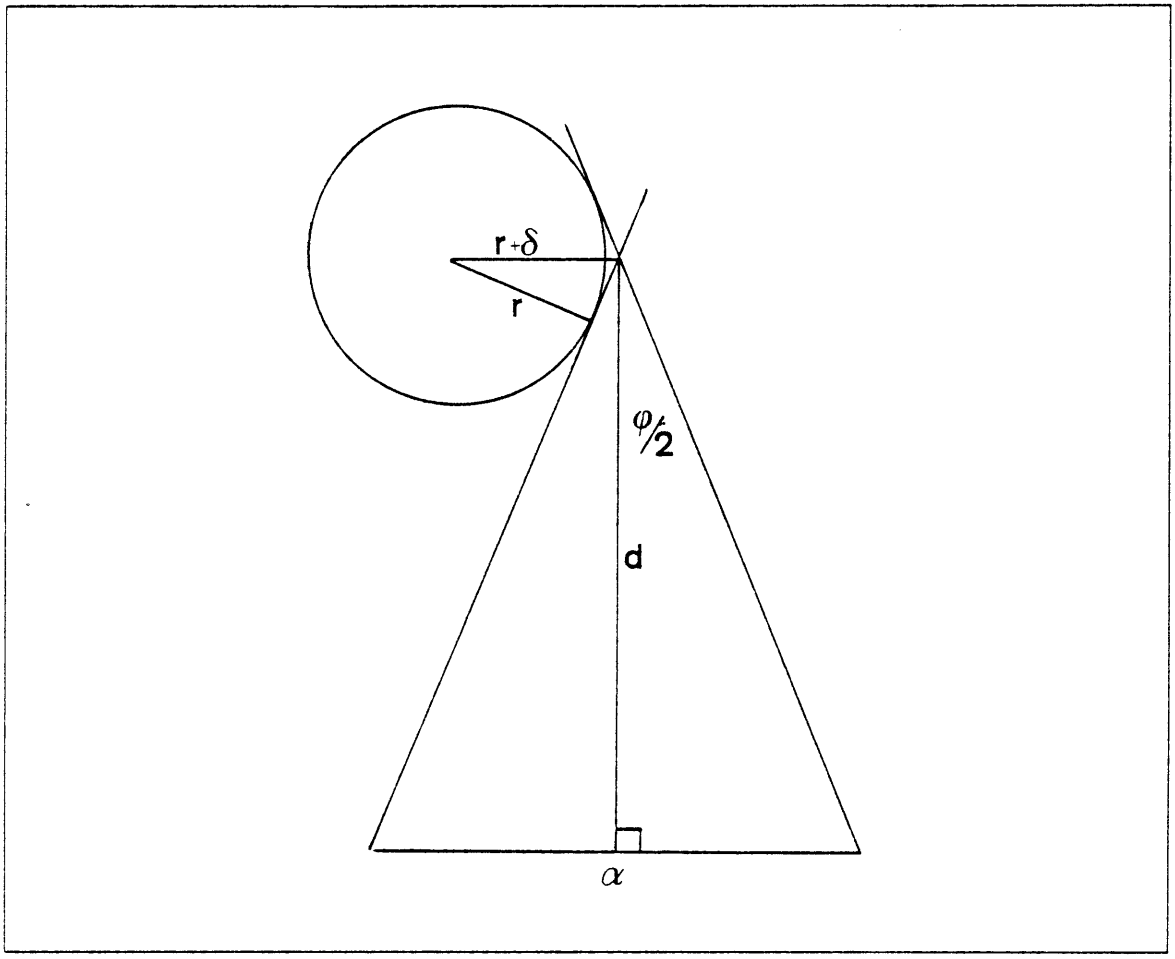


Figure 4.23. Geometry of Matching Occluding Contours. The distance δ, as a function of d and r, measures the error in the perceived shape of the cylinder that would result from matching the occluding contours.

$$\cos\frac{\phi}{2} = \frac{d}{\sqrt{\frac{a^2}{4} + d^2}} = \frac{r}{r+\delta}.$$

Thus,

$$\frac{\delta}{r} = \frac{\sqrt{\frac{a^2}{4} + d^2} - d}{d}.$$

and this is approximated by

$$\frac{\delta}{r} \approx \frac{-d + d(1 + \frac{a^2}{8d^2})}{d} \approx \frac{a^2}{8d^2}.$$

Thus, the error in the perceived shape of the cylinder relative to its size is very small for cylinders separated from the viewer by distances bigger than one inter-ocular unit.

This result implies that the error in depth associated with matching occluding contours is very minor, except for situations of very small separation from the viewer.

4.7 Transformations of Disparity

Since my ultimate goal is to describe the surface structure of objects in the scene, it is necessary to convert the disparity information into a form that more directly relates to the surface shape. There are two obvious transformations of disparity: depth or distance, and surface orientation. Again, the term *depth* will refer to the subjective distance to the object as perceived by the viewer, while *distance* refers to the objective physical distance from the viewer to the object. The surface orientation of a point is defined as the orientation of the normal vector at that point relative to some axes system.

4.7.1 Exact Distance

The geometrical situation involved in computing distance for a coordinate system centered between the eyes is illustrated in Figure 4.24. Two cases are shown: the simplest case of an object centered between the eyes, and the more complex case of an object off axis. The inter-ocular distance is denoted by a, and the disparity by $\phi = \phi_1 + \phi_2$.

In the simplest case, $\phi_1 = \phi_2$, and

$$\begin{aligned} d_1 &= \frac{a}{2} \cdot \cot \phi_1 \\ &= \frac{a}{2} \cdot \cot \frac{\phi}{2}. \end{aligned}$$

In the general case,

$$d_2 = \frac{a}{2} \cdot \sin \beta$$

and

$$x = \frac{a}{2} \cdot \cos \beta.$$

The law of sines implies that

$$\begin{aligned} x + y &= \frac{a \cos(\phi_2 + \beta)}{\cos \phi_2} \\ &= a(\cos \beta - \sin \beta \tan \phi_2). \end{aligned}$$

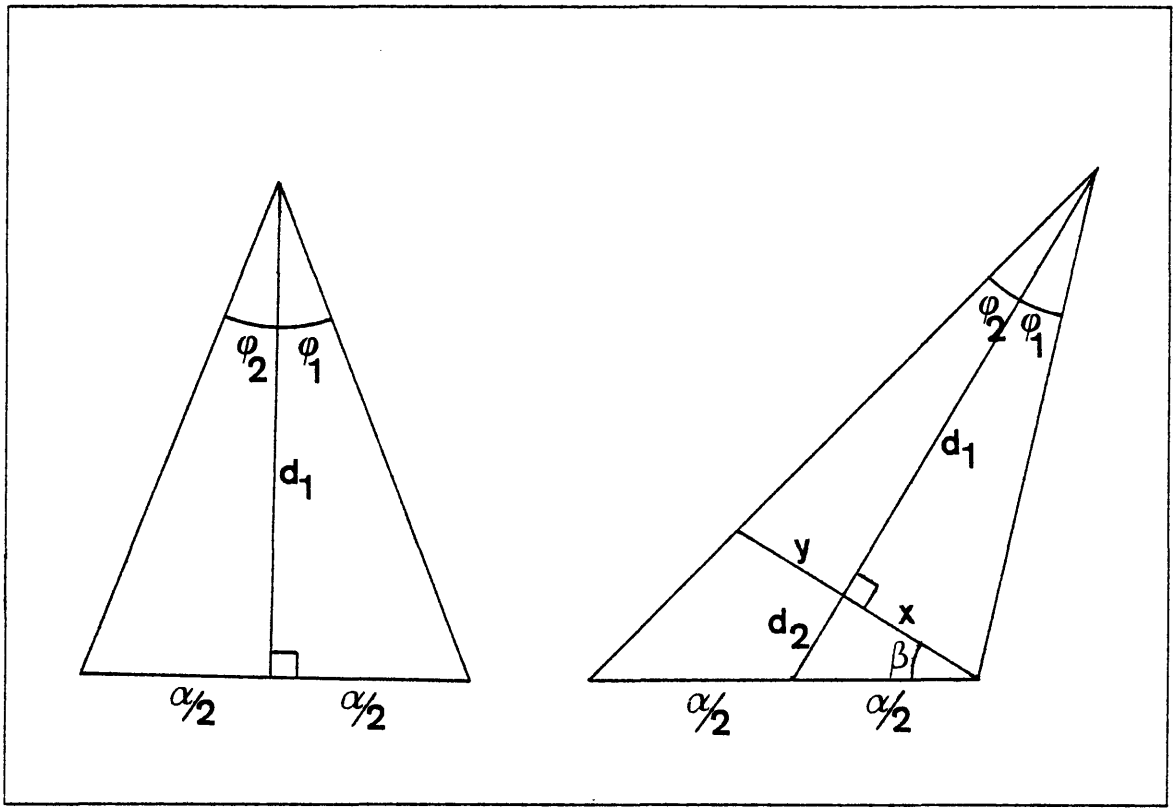


Figure 4.24. The Geometry of the Computation of Distance. The left diagram shows the simplest case of an object centered along the axis between the eyes. The inter-ocular distance is denoted by α, and the disparity is given by $\phi = \phi_1 + \phi_2$. The distance of interest is d_1, as a function of the inter-ocular separation α and the disparity ϕ. The right diagram shows the general case of an object lying off axis. Here, the distance of interest is $d = d_1 + d_2$ as a function of the disparity $\phi = \phi_1 + \phi_2$, the inter-ocular separation α, and the off axis angle β.

Thus

$$y = \frac{\alpha}{2}(\cos\beta - 2\sin\beta\tan\phi_2).$$

As a consequence,

$$\tan\phi_1 = \frac{\alpha\cos\beta}{2d_1}$$

$$\tan\phi_2 = \frac{\alpha\cos\beta}{2d_1 + 2\alpha\sin\beta}.$$

Thus,

$$\tan\phi = \frac{\tan\phi_1 + \tan\phi_2}{1 - \tan\phi_1\tan\phi_2}$$

$$= \frac{\alpha\cos\beta(4d_1 + 2\alpha\sin\beta)}{4d_1^2 + 4d_1\alpha\sin\beta - \alpha^2\cos^2\beta}.$$

Trigonometric and algebraic manipulation yields

$$d_1 = \frac{a}{2}\left(\cos\beta\cot\phi - \sin\beta + \sqrt{1+\cos^2\beta\cot^2\phi}\right).$$

Hence,

$$\begin{aligned} d &= d_1 + d_2 \\ &= \frac{a}{2}\left(\cos\beta\cot\phi + \csc\phi\sqrt{1-\sin^2\beta\cos^2\phi}\right). \end{aligned}$$

Note that if $\beta = 0$, then this expression reduces to

$$d = \frac{a}{2}(\cot\phi + \csc\phi),$$

and $\cot\phi + \csc\phi = \cot\frac{\phi}{2}$, so that the expression correctly reduces to the simple case.

To compute d exactly, it is necessary to know not only the disparity ϕ, but also the inter-ocular distance a and the angle β. In order to determine this angle, one needs to know the camera angles, that is the exact angular position of the eyes, relative to the coordinate system. For a general imaging system, this is not a problem. However, for the case of the human system, one must consider whether the system has access to these angles. For example, does the human system read the tensions on the eye muscles in such a way as to extract the optic axis angles relative to a fixed coordinate frame? Furthermore, one must also know the interocular separation, and one must again ask, for the human system, whether there is any evidence that this value is actually accessible to the system.

4.7.2 Relative Distance

Rather than attempting to measure the distance exactly, one may instead simply attempt to measure the distance relative to some fixed point [1]. The distance is given by

$$d = \frac{a}{2}\left(\cos\beta\cot\phi + \csc\phi\sqrt{1-\sin^2\beta\cos^2\phi}\right).$$

If the objects being viewed are assumed not to lie too far off axis, then β is small. (Note that this assumption is reasonable because if an object lies too far off axis, only one eye will be able to view it.) In this case, the distance expression reduces to the approximation:

$$d \approx \frac{a}{2}\cot\frac{\phi}{2}.$$

[1]Such as the horopter.

Thus,

$$d + \Delta d \approx \frac{a}{2} \cot \frac{\phi + \Delta\phi}{2}$$

and

$$\frac{\Delta d}{d} \approx \cot \frac{\phi + \Delta\phi}{2} \tan \frac{\phi}{2} - 1.$$

By taking a series expansion for the cot and tan functions, the above expression reduces to

$$\frac{\Delta d}{d} \approx \frac{-\Delta\phi}{\phi + \Delta\phi}\left(1 + \frac{2\phi^2\Delta\phi + \phi(\Delta\phi)^2}{12} + \ldots\right).$$

We know that the change in disparity $\Delta\phi$ must be less than 2^o (Marr and Poggio, 1979). Further, if any object is assumed to lie at least 10 centimeters from the viewer, then the disparity ϕ must be less than 34^o. These two factors combined yield an estimate for the maximum error in relative depth of less than 6%. Hence, the approximation

$$\frac{\Delta d}{d} \approx \frac{-\Delta\phi}{\phi + \Delta\phi}.$$

Thus, we see that even to determine relative depth, it is still necessary to have some estimate of the actual disparity to the point of fixation, as well as the relative disparities at nearby points in the image. The inter-ocular distance no longer plays a role, however.

Of course, for the general case of practical imaging systems, concern over the accessability of the camera angles and the separation of the cameras is not of major concern, since such parameters can easily be measured in such cases.

4.7.3 Surface Orientation

Another possible method of representing surface shape is to use surface orientation for small patches of the surface. One can, of course, compute surface orientation from depth values, but here the problem of computing surface orientation directly from disparity is considered. The geometry of the computation of surface orientation is illustrated in Figure 4.25.

In this system, the vector corresponding to a particular point is given by

$$\mathbf{r} = d\{\cos\psi\cos\theta, \cos\psi\sin\theta, \sin\psi\}$$

or

$$\mathbf{r} = d\mathbf{v}.$$

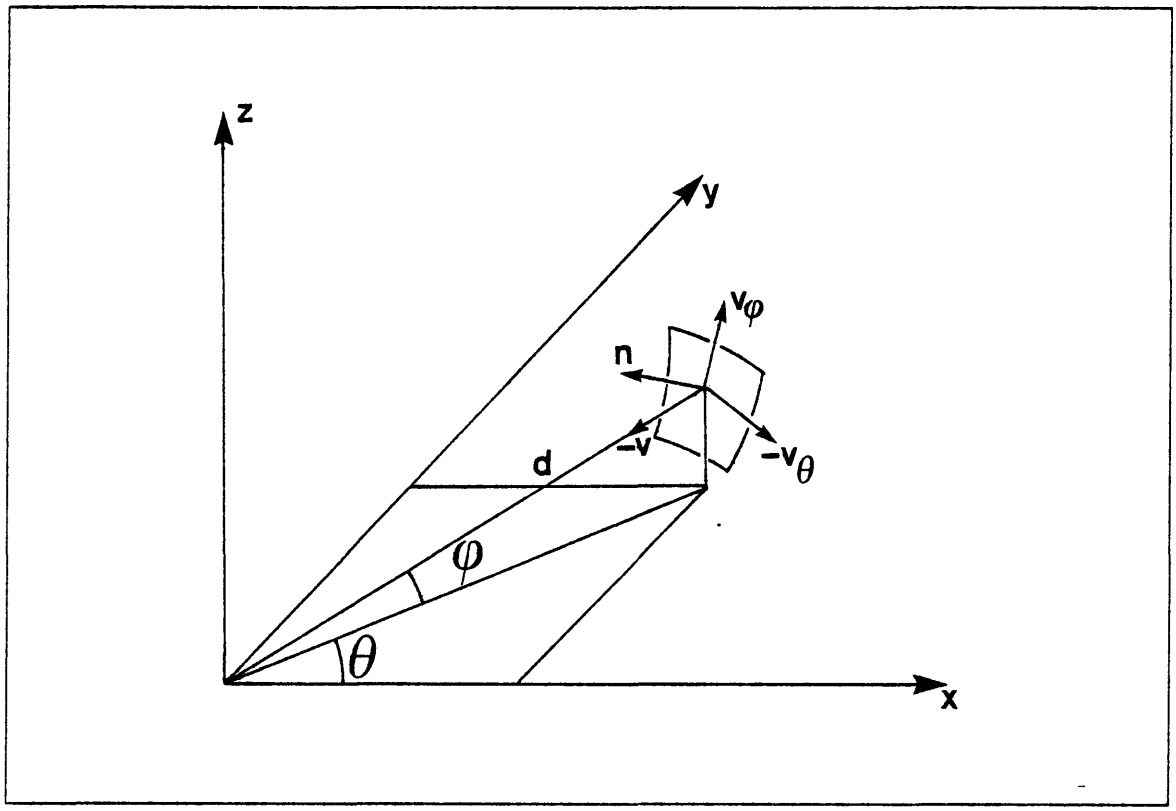


Figure 4.25. The Geometry of the Computation of Surface Orientation. The vector r describes the distance to a surface point in the space, as a function of θ, ψ and the distance d. The surface normal at the point is denoted by n and can be related to a new coordinate system based on the view vector, and spanned by the coordinated axes $-\mathbf{v}_\theta, \mathbf{v}_\psi$ and $-\mathbf{v}$.

The partial derivatives of this vector are

$$\mathbf{r}_\theta = d_\theta \mathbf{v} + d \cos\psi \mathbf{v}_\theta$$
$$\mathbf{r}_\psi = d_\psi \mathbf{v} + d\mathbf{v}_\psi$$

where

$$\mathbf{v} = \{\cos\psi\cos\theta, \cos\psi\sin\theta, \sin\psi\}$$
$$\mathbf{v}_\theta = \{-\sin\theta, \cos\theta, 0\}$$
$$\mathbf{v}_\psi = \{-\sin\psi\cos\theta, -\sin\psi\sin\theta, \cos\psi\}$$

The normal vector at this point is given by the cross product of the two partial derivatives.

$$dd_\psi \cos\psi \mathbf{v}_\psi + dd_\theta \mathbf{v}_\theta - d^2 \cos\psi \mathbf{v}$$

If this vector is related to a coordinate system spanned by the view vector and its partial derivatives, with unit coordinate vectors $-\mathbf{v}_\theta, \mathbf{v}_\psi, -\mathbf{v}$, then the normal is given in this system by

$$\mathbf{N} = \left\{-dd_\theta, dd_\psi \cos\psi, d^2 \cos\psi\right\}$$

The unit normal is given by

$$\mathbf{n} = \frac{\{-d_\theta, d_\psi \cos\psi, d\cos\psi\}}{\sqrt{d_\theta^2 + \cos^2\psi(d_\psi^2 + d^2)}}.$$

By rotating the original axis system about the x axis, one can set $\psi = 0$. Thus, the expression for the unit normal becomes

$$\mathbf{n} = \frac{\left\{-\frac{d_\theta}{d}, \frac{d_\psi}{d}, 1\right\}}{\sqrt{\frac{d_\theta}{d}^2 + \frac{d_\psi}{d}^2 + 1}}.$$

From our previous analysis of the case of relative depth, we know that

$$\frac{d_\theta}{d} \approx -\frac{\phi_\theta}{\phi + \phi_\theta}$$
$$\frac{d_\psi}{d} \approx -\frac{\phi_\psi}{\phi + \phi_\psi}.$$

Finally, if changes in distance are assumed to be small relative to the actual distance, $d_\theta \ll d$, $d_\psi \ll d$, then the unit normal may be approximated by

$$\mathbf{n} \approx \left\{\frac{\phi_\theta}{\phi + \phi_\theta}, -\frac{\phi_\psi}{\phi + \phi_\psi}, 1\right\},$$

where ϕ_θ and ϕ_ψ are the partial derivatives of disparity in the directions of the two coordinate axes.

It is apparent from these calculations that if one wishes to compute local surface orientation directly from disparity, one must be able to compute the partial derivatives of disparity. But since the Marr-Poggio stereo algorithm explicitly determines disparity only along zero-crossing contours, one is confronted with the task of determining derivatives from a sparse array. Provided the disparity array is not too sparse, this can be done. Perhaps the easiest method is to determine the local gradient of the disparity about a point by a least-squares planar fit. This would allow one in most situations to determine ϕ_θ and ϕ_ψ, and hence the local surface orientation.

CHAPTER 5

THE CONSTRAINTS ON INTERPOLATION

In the introduction to this thesis, we saw that for any process which transforms images into representations of surfaces, there are two stages. The first stage consists of computing explicit surface values (depth or surface orientation) at a particular set of points in the image. The second stage consists of interpolating between these known points to obtain a complete specification of the visible surfaces.

In the first half of this thesis, I described the method of stereo vision for achieving the first stage. This theory completes the first stage by computing explicit disparity values along the zero-crossing contours of the convolved image. Thus, to achieve the goal of complete specification of surfaces, we now turn to the second stage, the interpolation of surfaces between those known points.

Although the first half of this thesis dealt specifically with the method of stereo for computing surface values, the second half of this thesis will deal with the general problem of surface interpolation. There are two reasons for this general approach. The first is that there are several visual processes which will require a method of surface interpolation; these include structure from motion, shape from shading, and shape from surface contours. By dealing with the general question, independent of the source of information about the surfaces, a method for interpolating surfaces may be developed, which is valid for a wide range of visual processes. The second reason is that surface interpolation may be considered an independent module of the visual system. The only assumption made about the visual processes which feed the interpolation module is that they compute explicit surface information

along the zero-crossings of the convolved images. This is certainly true for stereo, structure from motion, and structure from surface contours. Hence, the method of surface interpolation developed in this half of the thesis should be applicable to each of these visual processes.

There are two parts to the problem of creating complete depth specifications. The most general problem is to consider a strictly mathematical question, independent of its relevance to the human visual system. Suppose one is given a visual process which determines surface information at points corresponding to relevant changes in the images. Can a probability density function be assigned to the set of surfaces consistent with this information which measures the inconsistency of each surface with the information? If so, can an algorithm for finding the surface be constructed which optimizes this probability density, by finding the least inconsistent surface? Secondary to this mathematical problem is the question of whether such a process is used by the human visual system.

Of course, the first problem is of considerable interest, irrespective of its relevance to the human system. There are many applications, such as high-altitude photomapping, hand-eye coordination systems, industrial robotics, inspection of manufactured parts, where it is useful to create a complete specification of the surfaces.

Thus, it is of interest, both from the view point of the human system, and from the view point of potential applications, to consider a computational theory of the process of creating complete specifications of the shapes of visible surfaces. There are three aspects to a computational theory which must be considered: the input representation, the output representation, and the constraints on the computation. The input to this process will be surface values, either depth or surface orientation, computed at those image locations corresponding to significant changes in image intensity.

It is desired that the output representation be a complete specification of surface information. However, the actual form of the representation could be any one of several forms, for example, distance, relative distance, or surface orientation. The choice of what kind of representation to use will in part relate to the applications to follow and the data available. We have already seen possible forms for the representation, and their relative merits.

Although surface values at all points of the image are important, there is another aspect of surface information which should be made explicit. This is the set of discontinuities in surfaces; the occluding contours, both subjective and objective. Marr (1978) argues that the $2\frac{1}{2}$-D sketch should be

a viewer-centered representation which includes both explicit surface information, such as depth and surface orientation, and explicit contours of surface discontinuities. In this thesis, the concentration is on the problem of creating explicit surface information at all points of the surface. The question of surface discontinuties will be outlined, and possible algorithms suggested, but an implementation of this stage has not been completed.

5.1 The Computational Constraint

We now turn to the heart of the matter, the computational constraints involved in the process of creating complete surface specifications.

Suppose one were to attempt to construct a complete surface description based only on the surface information known along the zero-crossings. An infinite number of surfaces would consistently fit the boundary conditions provided by these values. Even if restricted to smooth surfaces, an infinite number of possibilities remain. Yet there must be some way of deciding which surface, or at least which of a small family of surfaces, could give rise to the zero-crossing descriptions. This means that there must be some additional information available from the visual process which, when taken into account, will identify a class of nearly indistinguishable surfaces which represent the visible surfaces of the scene.

In order to determine what information is available from the visual process, one must first carefully consider the process by which the zero-crossing contours are generated. We have already relied on the fact that sudden changes in the reflectance of a surface, caused, for example, by surface scratches or texture markings will give rise to zero-crossings in the convolved image. Sudden or sharp changes in orientation shape of the surface will under most circumstances also give rise to zero-crossings. This fact will be used to constrain the possible shapes of surfaces which could give rise to particular surface values along zero-crossing contours.

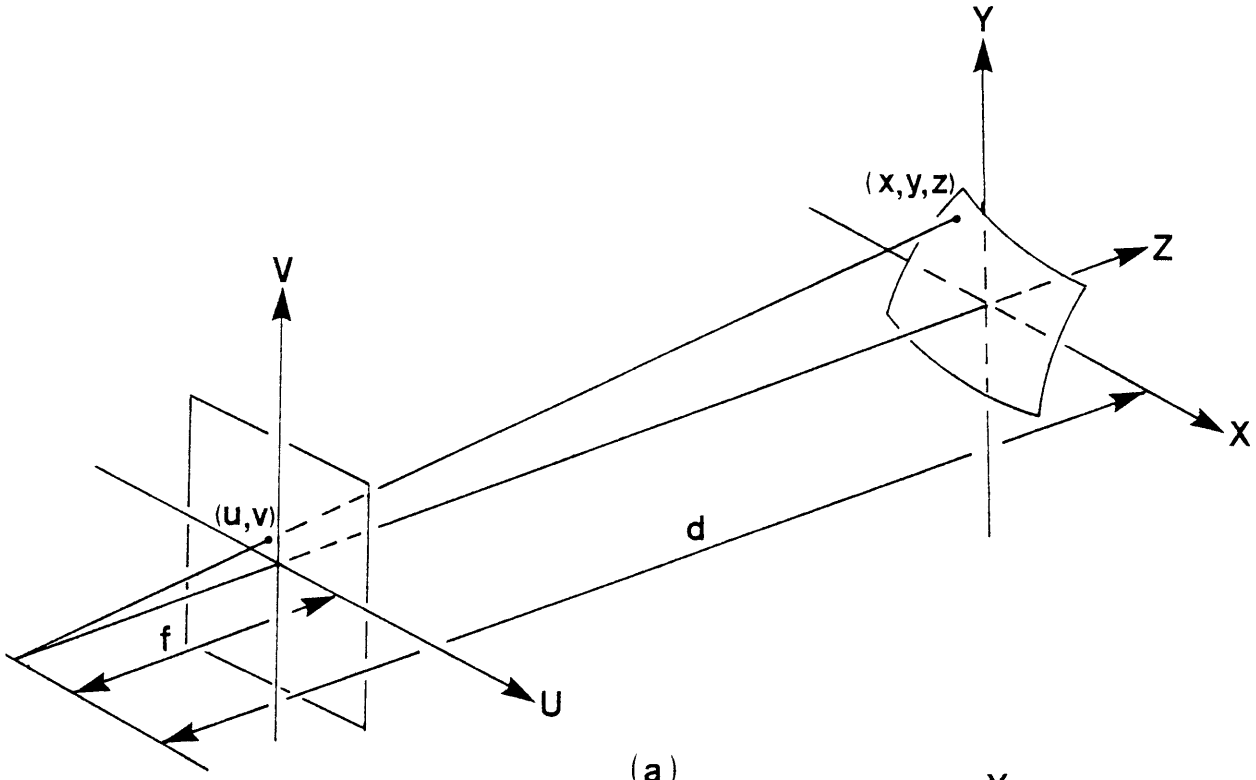
5.2 Image Formation

In order to examine the process of zero-crossing formation, a review of the processes involved in the formation of an image will be presented. One of these processes is concerned with the geometry of the projection from the scene to the image. The other is concerned with the process by which image intensity values at a point are formed. The analysis of these processes has been undertaken by several investigators. In this section, the work of Horn (1975, 1977) and Woodham (1978) is relied upon.

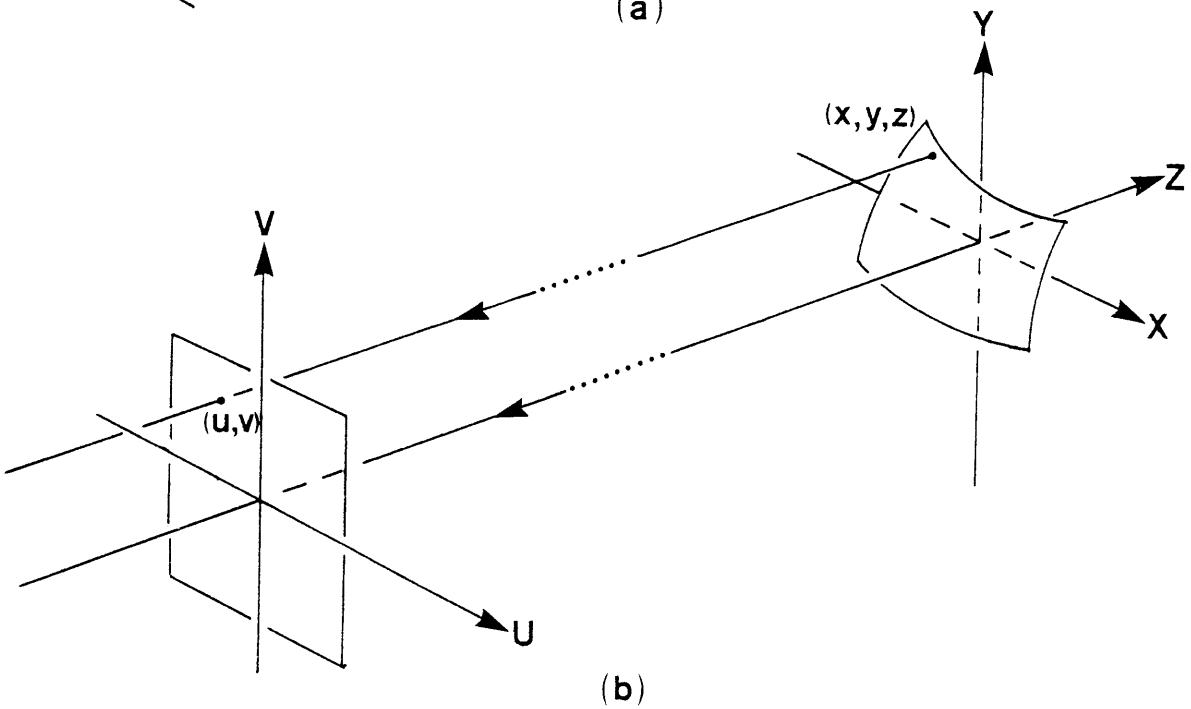
5.2.1 From the Object to the Image

Consider the projection of a surface point onto the image plane as illustrated in Figure 5.1. Note that it is convenient to think of the image plane as being in front of the lens rather than behind it. This avoids inversion of the image. Position the lens at the origin and the image plane perpendicular to the z axis. The focal length of the lens, that is, the distance between the view point and the image plane, will be represented by f.

Figure 5.1. The Image Projection. The top figure shows a perspective projection. The focal length is given by f and the distance from the object coordinate origin to the center of the lens is given by d. The image plane is placed in front of the lens to avoid image inversion. The bottom figure shows an orthographic projection. This occurs in the case of objects that are small relative to the viewing distance, where the focal length f is infinite, causing all rays from the object to the image plane to be parallel.



(a)



(b)

The proportionality of similar triangles yields

$$\frac{u}{f} = \frac{x}{z+d} \qquad \frac{v}{f} = \frac{y}{z+d}$$

so that

$$u = \frac{f}{z+d} \cdot x \qquad v = \frac{f}{z+d} \cdot y .$$

These equations define the standard *perspective projection*, determining an image point (u, v) corresponding to an object point (x, y, z). If the size of the objects in the scene is small compared to the viewing distance, then for all surface points (x, y, z), z is nearly constant and the equations become, after scaling,

$$u = x \qquad v = y .$$

These equations define the standard *orthographic projection*. Note that in the case of orthographic projection, all rays from the surface to the image plane are parallel, so the use of separate image coordinates is somewhat redundant. Thus, the image coordinates and object coordinates can be referred to interchangeably.

This determines where a point on a surface will appear in the image.

5.2.2 Grey-Level Formation

It is now necessary to determine what intensity value will be associated with a particular image location. There are four factors involved in the formation of the grey-level values associated with each pixel location in the image. Horn (1970, 1975) and Woodham (1978) identify these as:

(1) the imaging geometry,

(2) the incident illumination,

(3) the surface photometry,

(4) the surface topography.

Surface photometry refers to how light is reflected by the object surface. It is determined by optical constants of the object material and by the surface microstructure. Surface topography is the surface detail which is within the resolution limits of the imaging hardware. It refers to the gross object shape relative to the viewer. Thus, for example, surface texture or colour markings refer to changes in the surface photometry, while sharp changes in the orientation of the surface are effects of the surface topography. Both factors can cause zero-crossings in the convolved image.

IMAGE FORMATION

It is now necessary to determine how much light is radiated by different parts of the object towards the imaging device. In general, as noted, the amount of light radiated in a particular direction by a surface element depends on a number of factors. Each is dealt with in turn.

The simplest case of illumination geometry is that of a single point source of light. In this case, the geometry of reflection is governed by the three angles shown in Figure 5.2. The angle between the local surface normal and the incident ray is called the incident angle and is denoted by i. The angle between the local surface normal and the emitted ray is called the view angle and is denoted by e. The angle between the incident and emitted rays is called the phase angle and is denoted by g. The fraction of incident illumination at a given surface point that is reflected in the direction of the viewer per unit solid angle per unit area is given by the reflectance function $\phi(i, e, g)$. It is generally assumed (Silver, 1980) that cases with a more complicated light source distribution can be modelled by a superposition of single point sources, excluding the effects of shadows and mutual illumination.

The following quantities are of use in the problem of image formation. Let the object irradiance at a surface point (x, y, z) be denoted by $a(x, y, z)$. This is usually constant or obeys some inverse square law.

The ratio of image irradiance to scene radiance is denoted by $t(x, y, z)$. If the surface is composed of a single material, this is a constant. However, for cases in which the surface photometry changes, such as at texture changes or colour changes, t is a function of its position on the surface.

Thus one can set the scaled object irradiance, frequently called the *albedo*, to be $A(x, y, z) = ta(x, y, z)$. Let $\mathrm{r} = (x, y, z)$ be the coordinates of a visible point on an object and $\mathrm{r}' = (x', y', f)$ be the coordinates of the corresponding point in the image. This is given by the geometry of the projection, which need not be orthographic.

Finally, let $b(x', y')$ be the image irradiance measured at the image point (x', y'). Since the scene radiance is proportional to the image irradiance (Horn and Sjoberg, 1978),

$$A(\mathrm{r})\phi(i, e, g) = b(\mathrm{r}')$$

This is the general equation of image formation, derived by Horn (1970, 1975, 1977, also Horn and Sjoberg, 1978).

A number of simplifying assumptions are frequently made, and the effects of such assumptions

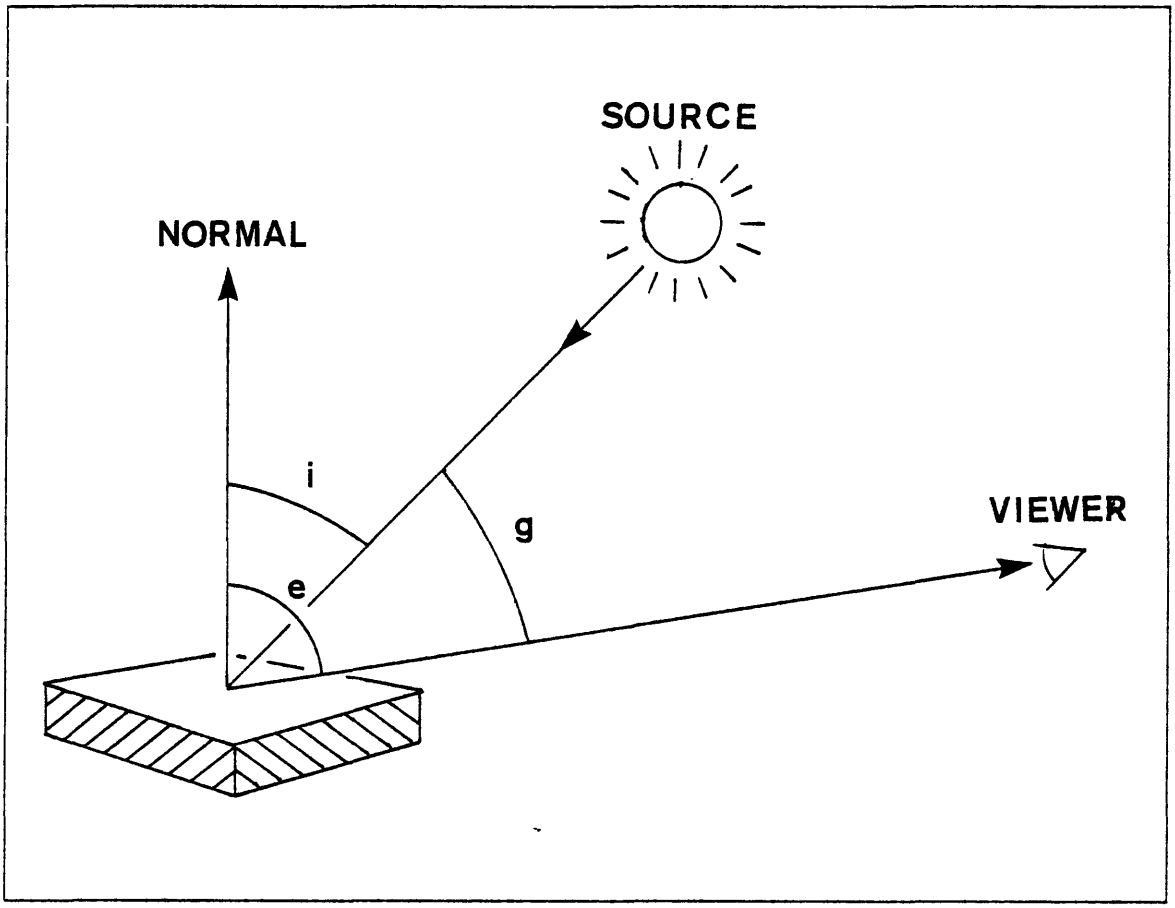


Figure 5.2. The Imaging Geometry. The incident angle i is the angle between the incident ray and surface normal. The view angle e is the angle between the emergent ray and the surface normal. The phase angle g is the angle between the incident and emergent rays.

are briefly listed. If it is assumed that the object is constructed of uniform material, and that each surface point receives the same incident illumination or irradiance, then $A(r)$ is constant or obeys some inverse-square law with respect to distance from the source. Of course, if the surface material changes, then this will not be true.

If the projection is orthographic, so that an object point (x, y, z) maps into an image point (x, y), then one may write

$$b(r') = I(x, y)$$

where $I(x, y)$ is the image intensity recorded at a point (x, y).

If the light source is distant relative to the viewer and object, then the phase angle g is roughly constant. The remaining angles i and e then depend on the local surface normal. If the equation of a

smooth surface is given as $z = f(x, y)$, then the surface normal toward the viewer at a point (x, y) is

$$\{f_x(x, y), f_y(x, y), -1\}.$$

It is conventional to define

$$p = f_x(x, y) \qquad q = f_y(x, y)$$

so that the surface normal becomes $\{p, q, -1\}$. Since g is constant in this case, $\Phi(i, e, g)$ can be rewritten as $R(p, q)$, where R is the reflectance associated with a particular surface normal.

Note that in many cases one can decompose the surface photometry from the surface topography in such a manner that the image intensities depend only on p and q. In fact, Woodham (1978) observes:

> "No matter how complex the distribution of incident illumination, for most surfaces, the fraction of the incident light reflected in a particular direction depends only on the surface orientation."

Thus, if attention is restricted to situations in which the light source can be considered distant relative to the separation of object and viewer, and to situations in which the image projection is orthographic, then the image equation becomes

$$I(x, y) = A(x, y, z)R(p(x, y), q(x, y))$$

For most surfaces, and most illumination geometries, this decomposition of the image equation into a product of two factors is valid (Horn, 1977; Horn and Bachman, 1977). The albedo function $A(x, y, z)$ describes the effects of the surface photometry on the image intensities. The reflectance function $R(p(x, y), q(x, y))$ describes the effect of surface topography on the image intensities. This will be taken as the basic image formation equation, that is, the means by which an individual grey level is assigned to a particular image location.

5.3 No Information Is Information

In general, any one of a multitude of widely varying surfaces could fit the boundary conditions imposed by the stereo algorithm. The intention in this section is to show that to be completely consistent with the stereo process, such surfaces must meet both explicit conditions and implicit conditions.

The explicit conditions are given by the depth or surface orientation values along the zero-crossing contours. The implicit conditions are that the surface not give rise to any other zero-crossing contours which do not appear in the convolved image. Thus the assertion:

The absence of zero-crossings constrains the possible surface shapes.

Just as the presence of a zero-crossing tells us that something is happening at a given location, the absence of a zero-crossing tells us the opposite. It is this implicit information which I now shall attempt to enunciate.

In order to make explicit any constraints on the shape of the surface, for locations in the image not associated with a zero-crossing, one must carefully examine the image formation equation. Two goals will be kept in mind. The first is to determine what conditions will cause a local change in intensity, and the second will be to combine this constraint with the input from the visual processes, such as stereo or structure from motion.

The image formation equation is given by:

$$I(x, y) = A(x, y) \cdot R(p(x, y), q(x, y))$$

and its derivatives are given by:

$$\begin{aligned} I_x(x, y) &= A_x R + A R_x \\ I_y(x, y) &= A_y R + A R_y \\ \nabla^2 I(x, y) &= \nabla^2 A \cdot R + A \cdot \nabla^2 R + 2\big(A_x R_x + A_y R_y\big) \end{aligned}$$

where by the chain rule:

$$\begin{aligned} R_x &= R_p p_x + R_q q_x \\ R_y &= R_p p_y + R_q q_y \\ R_{xx} &= R_{pp} p_x^2 + 2R_{pq} p_x q_x + R_{qq} q_x^2 + R_p p_{xx} + R_q q_{xx} \\ R_{yy} &= R_{pp} p_y^2 + 2R_{pq} p_y q_y + R_{qq} q_y^2 + R_p p_{yy} + R_q q_{yy} \end{aligned}$$

If the surface is planar, (a linear function of x and y), then the partial derivatives p_x, p_y, q_x, q_y all vanish and hence so do R_x, R_y, R_{xx}, R_{yy}. Thus, $\nabla^2 I = \big(\nabla^2 A\big)R$. Hence, $\nabla^2 I = 0$ if and only if $\nabla^2 A = 0$ or $R = 0$. The condition of $R = 0$ is equivalent to no light being emitted by the surface. This as an uninteresting case and is not found in practice. The other case, $\nabla^2 A = 0$, implies that photometric zero-crossings, due to sudden pigment changes or sudden texture changes,

are encompassed by this method. Even for nonplanar surfaces, a sudden photometric change will usually force a zero-crossing in the intensities. It was exactly these types of surface markings that led to the detection of zero-crossings as a component of the Marr-Hildreth edge detection theory.

However, the form of the equations indicates that there may be topographic effects which could also cause sharp changes in intensity. We are interested in the surface conditions which would cause these types of intensity changes.

Consider the following example. Suppose one is given a closed zero-crossing contour, within which there are no other zero-crossings. An example would be a circular contour, along which the disparity is constant. One surface which is consistent with this set of boundary conditions is a flat disk. However, one could also fit other smooth surfaces to this set of boundary conditions. For example, the highly convoluted surface formed by $\sin\left(\sqrt{x^2+y^2}\right)$ would be consistent with the known disparity values. Yet in principle, such a rapidly varying surface should give rise to other zero-crossings, at least in the unfiltered image. This follows from the observation that if the surface orientation undergoes a periodic variation, then it is likely that the intensity values will also undergo such a variation. Such variation would give rise to zero-crossings in the convolved image. Since the only zero-crossings are at the borders of the object, this implies that the surface $\sin\left(\sqrt{x^2+y^2}\right)$ is not a valid representative surface for this set of boundary conditions.

Hence, our hypothesis, which we shall check in the following sections, is that the set of zero-crossing contours contains implicit information about the surface as well as explicit information. If a set of conditions on the surface shape that cause inflections in the intensity values can be determined, then one may be able to determine a likely surface structure, given a set of boundary conditions along the zero-crossing contours.

The basic problem is, under what conditions does bending of the surface force an inflection in the intensity array? This question will be answered by considering specific cases in the following sections. The arguments will deal with the filtered intensities of the form $\nabla^2 I$, whereas the Marr-Hildreth theory uses operators of the form $\nabla^2(G * I)$. The additional smoothing introduced by the Gaussian G will be discussed later. In what follows, we assume that A, R and z are functions which have continuous second order partial derivatives.

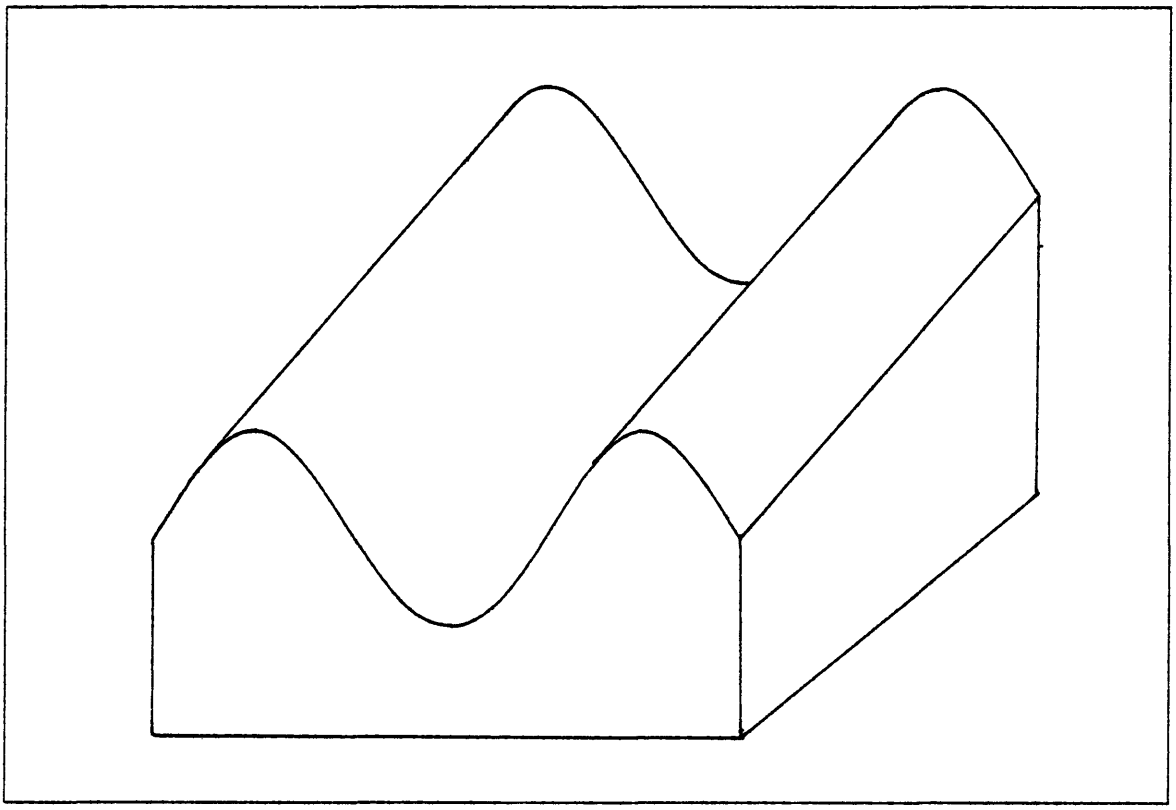


Figure 5.3. An Example of a Developable Surface. The component of the surface orientation in the y direction is constant for this region of the surface, so that the only variations in surface orientation take place in the x direction.

5.3.1 The One-Dimensional Case

I shall consider first the one-dimensional case of a developable surface. Note that the Laplacian ∇^2 is orientation independent, so that without loss of generality, one may rotate the coordinate system of the image to suit our needs. One may assume that the surface has the form $z(x, y)$ such that $z_y(x, y) = q(x, y) = c$, in the local region under consideration. Hence, $q_x = q_y = 0$, and $q_{xx} = q_{xy} = q_{yy} = 0$. Furthermore, for second differentiable surfaces, $p_y = q_x$ so that $p_y = 0$, and $p_{xy} = p_{yy} = 0$. A sample surface is shown in Figure 5.3, (similar surfaces have been studied by Stevens, 1979).

Suppose that a one-dimensional slice in the x direction of the surface contains at least two inflection points. Figure 5.4 indicates a sample surface and its derivatives. Since the intensities with which we shall deal are caused by both surface photometry and surface topography, it is possible to have complex interactions between the two effects. In the case in which both factors have roughly

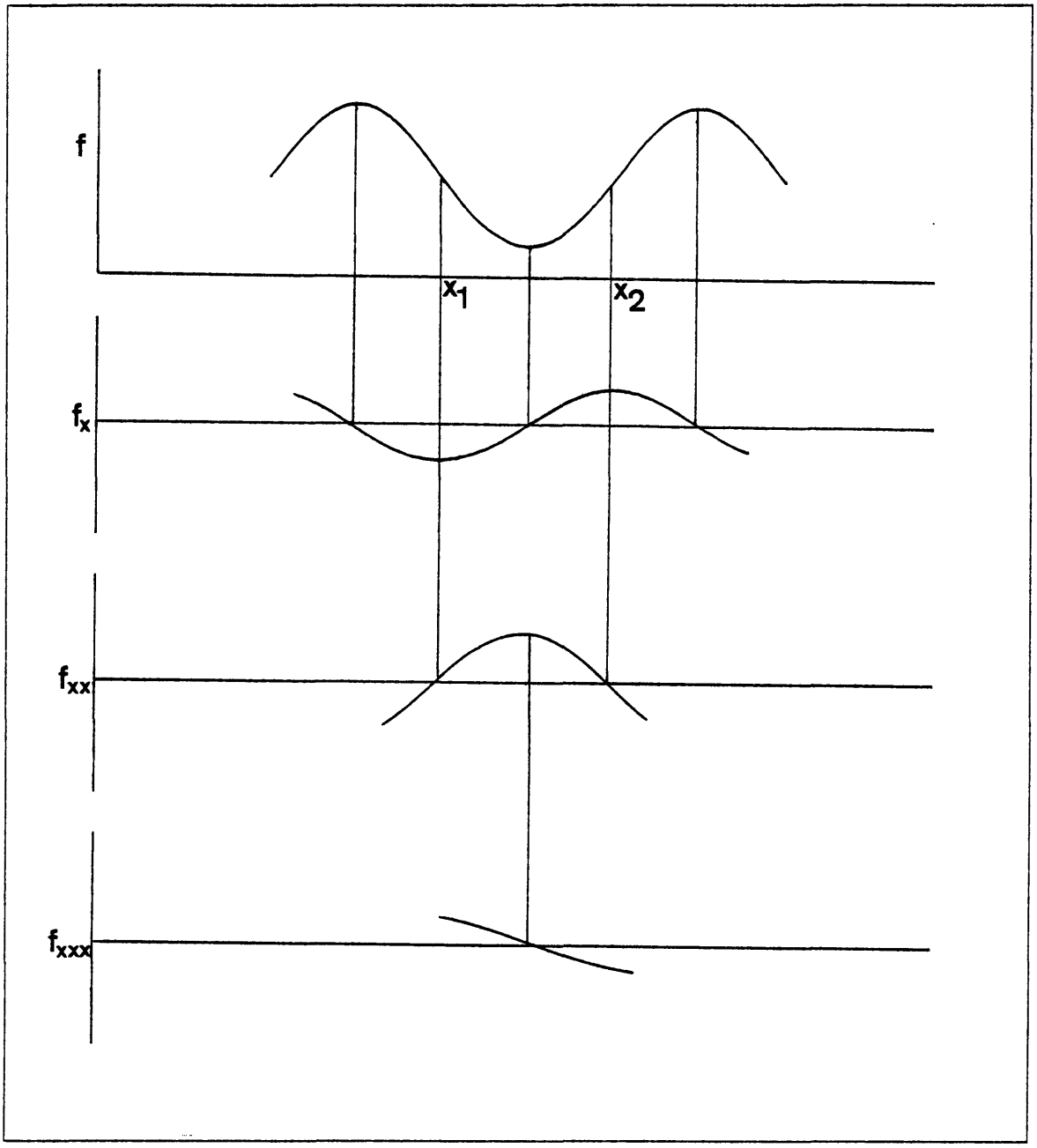


Figure 5.4. One-dimensional example. The top figure illustrates a slice of the surface, containing two inflection points. The second figure illustrates the first derivative of the surface function. The two inflection points of the surface correspond to extrema in the first derivative. The third figure illustrates the second derivative of the surface function. The two inflections in the surface correspond to zero-crossings in the second derivative. The bottom figure illustrates the third derivative of the surface function. Between the points corresponding to the two inflection points of the surface, the third derivative contains a zero-crossing.

equivalent magnitudes, one could construct situations in which the surface topography changes radi-

cally, yet the surface photometry also changes sufficiently so that there are no noticable changes in the image intensities, for one eye. However, note that in such situations it is probably the case that balancing of the two effects requires a specific viewer position, to properly account for the photometric effects. Since we have two eyes, the view point for the second eye will usually be sufficiently different to alter the photometric effects, and the cancellation of the topographic effects by the photometric effects will no longer be valid. Moreover, this situation is very unlikely.

Hence, the concentration will be on those situations in which at any point the changes in one of the two components, albedo or reflectivity, dominates changes in the other factor.

We have assumed that the surface is developable such that q is constant along this slice of the surface. In this case, the derivatives of the intensity equation are given by:

$$\nabla^2 I = \left(\nabla^2 A\right) \cdot R + 2A_x R_p p_x + A\left(R_{pp} p_x^2 + R_p p_{xx}\right).$$

Recall that the operation of finding fixed surface locations from an image consisted of locating the zero-crossings of the Laplacian $\nabla^2 I$. The goal is to determine what surface variations cannot occur in the absence of a zero-crossing. To do this, we will first determine what topographic factors can cause a zero-crossing in the Laplacian. The inverse will then specify what surface variations are disallowed.

Perhaps the simplest conditions for a zero-crossing to occur are given in the following result.

Theorem 1: Consider a portion of a second differentiable developable surface oriented along the y axis such that $z_y(x, y) = q = c$, for some constant c. If the following conditions are true:

(1) **The surface portion contains exactly two inflection points in the x direction, at x_1 and x_2,**

(2) **At the points x_1 and x_2, normalized changes in albedo are dominated by normalized changes in reflectance,**

$$\left|\frac{\nabla^2 A}{A}\right| < \left|\frac{R_p p_{xx}}{R}\right|,$$

(3) **The reflectance R does not pass through an extremum in this region of the surface,**

(4) **The reflectance R is not constant over this region of the surface,**

(5) **The albedo A is non-zero,**

then there exists a point $x_1 < x' < x_2$ such that $\nabla^2 I(x') = 0$.

Proof: The signum function is defined by:

$$sgn(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \end{cases}$$

From the derivatives of the image equation,

$$\nabla^2 I = (\nabla^2 A)R + 2A_x R_p p_x + A(R_{pp} p_x^2 + R_p p_{xx}).$$

At the inflection points x_1 and x_2, $p_x(x_i) = 0$. Hence, evaluation of the equation yields

$$\nabla^2 I(x_i) = \nabla^2 A(x_i) R(p(x_i)) + A(x_i) R_p(p(x_i)) p_{xx}(x_i) \quad \text{for } i = 1, 2$$

Condition (2) implies that the albedo changes are negligible in this region, so that the first term may be ignored,

$$sgn(\nabla^2 I(x_i)) = sgn(A(x_i) R_p(p(x_i)) p_{xx}(x_i)).$$

Condition (5) implies that $sgn(A) = 1$. Observing that

$$sgn(xy) = sgn(x) \cdot sgn(y),$$

the sign of the convolved intensity function at the surface inflection points is given by

$$sgn(\nabla^2 I(x_i)) = sgn(R_p(p(x_i)) p_{xx}(x_i)).$$

Condition (3) implies that R_p does not change sign in this region of the surface and hence

$$sgn(R_p(p(x_1))) = sgn(R_p(p(x_2))).$$

Note that $p_x = 0$ at x_1, x_2. The fact that there are exactly two inflections on the surface implies that $p_x \not\equiv 0$ over the neighbourhood (x_1, x_2). Thus,

$$sgn(p_{xx}(x_1)) \neq sgn(p_{xx}(x_2)),$$

and thus

$$sgn(\nabla^2 I(x_1)) \neq sgn(\nabla^2 I(x_2)).$$

The second differentiability of the surface implies that there exists a point $x' \in (x_1, x_2)$ such that $\nabla^2 I(x') = 0$. ∎

The importance of this theorem is that its contrapositive states something important about the possible surfaces which can fit the known depth information. Specifically, if there is a set of known depth points to which a surface is to be fit, we know that, given the conditions of the theorems, between any two zero-crossing points, there cannot be two or more inflections in the surface. If there were, there would have to be an additional zero-crossing there as well.

Corollary: Suppose one is given a set of known depth points at a set of zero-crossings, along a developable surface. If the albedo A is non-zero and the reflectance R is not constant for this region of the surface, then the surface cannot contain two or more inflection points between any pair of adjacent zero-crossings. ∎

Thus, this suggests that between known values, any surface fitting the boundary conditions should allow at most one inflection. However, this can be restricted even further.

Theorem 2: Consider a portion of a second differentiable developable surface oriented along the y axis such that $z_i(x, y) = c$, for some constant c. If the following conditions are true:

(1) The surface contains exactly one inflection point in the x direction at x_1, and the reflectivity R achieves an extremum at the point x_2, $x_1 \neq x_2$,

(2) At the point x_1 the normalized changes in albedo are dominated by normalized changes in reflectance of the form,

$$\left|\frac{\nabla^2 A}{A}\right| < \left|\frac{R_p p_{xx}}{R}\right|,$$

and at the point x_2 the normalized changes in albedo are dominated by normalized changes in reflectance of the form,

$$\left|\frac{\nabla^2 A}{A}\right| < \left|\frac{R_{pp} p_x^2}{R}\right|,$$

(3) The reflectance R is not constant,

(4) The albedo A is non-zero,

then there exists a point $x_1 < x' < x_2$ such that $\nabla^2 I(x') = 0$.

Proof: As in the proof of the previous theorem, at the point x_1,

$$sgn(\nabla^2 I(x_1)) = sgn(R_p(p(x_1))p_{xx}(x_1)).$$

At the point x_2, $R_p = 0$ so that

$$\nabla^2 I(x_2) = \nabla^2 A(x_2) R(p(x_2)) + A(x_2) R_{pp}(p(x_2)) p_x^2(x_2).$$

Condition (2) then implies that at this point, the normalized albedo changes are dominated by the normalized reflectance changes,

$$sgn(\nabla^2 I(x_2)) = sgn(A(x_2)) sgn(R_{pp}(p(x_2))) sgn(p_x^2(x_2)).$$

Condition (4) implies that $sgn(A) = 1$, so that

$$sgn(\nabla^2 I(x_2)) = sgn(R_{pp}(p(x_2))) sgn(p_x^2(x_2)).$$

There are two subcases. In the first subcase, $p_{xx}(x_1) > 0$. Since there is only one inflection point in the surface, this implies that $p(x_1) < p(x_2)$. Then $R_{pp}(p(x_2)) < 0$ implies that

$$R_p(p(x_1)) > R_p(p(x_2)) = 0.$$

Conversely, $R_{pp}(p(x_2)) > 0$ implies that

$$R_p(p(x_1)) < R_p(p(x_2)) = 0.$$

In either case,

$$sgn(R_{pp}(p(x_2))) \neq sgn(R_p(p(x_1)) p_{xx}(x_1)).$$

In the second subcase, suppose that $p_{xx}(x_1) < 0$. This implies that $p(x_1) > p(x_2)$. Then $R_{pp}(p(x_2)) < 0$ implies that

$$R_p(p(x_1)) < R_p(p(x_2)) = 0.$$

Conversely, $R_{pp}(p(x_2)) > 0$ implies that

$$R_p(p(x_1)) > R_p(p(x_2)) = 0.$$

In either case,

$$sgn(R_{pp}(p(x_2))) \neq sgn(R_p(p(x_1)) p_{xx}(x_1)).$$

Thus, we see that

$$sgn(\nabla^2 I(x_1)) \neq sgn(\nabla^2 I(x_2))$$

and as before the second differentiability of the surface implies that there exists a point $x' \in (x_1, x_2)$ such that $\nabla^2 I(x') = 0$. ∎

Corollary: Suppose one is given a set of known depth points at a set of zero-crossings, along a developable surface. If the albedo A is non-zero and the reflectance R is not constant for this region of the surface, then if the reflectance function R passes through an extremum, the surface cannot contain any inflection points between any pair of zero-crossings. ∎

This theorem is unfortunately not as strong as the previous one. Here, one can exclude a surface with one inflection point as being inconsistent with the zero-crossings only if there is also an extremum in the reflectance function in this range of the surface as well. However, one will not be able to gain explicit knowledge of the reflectance function, so one seems to be unable to make use of this theorem. Indeed, surfaces with one inflection point cannot explicitly be banned. However, they are unlikely to be the correct underlying surface.

To see this, let $p_0 = \min_x p(x)$ and $p_1 = \max_x p(x)$. Then the range of values which are taken by p over this region of the surface is given by $[p_0, p_1]$. The above theorem implies that the probability of any surface with one inflection point forcing an inconsistent zero-crossing in this region is proportional to the probability of the surface reflectance function $R(p(x))$ reaching an extremum, given by some probability measure $\rho_R([p_0, p_1])$. If this probability can be reduced, then the resulting surface will have less probability of being inconsistent, and therefore will be a more likely candidate for the underlying surface. That is, since there are no additional zero-crossings evident between the two endpoints of the region, the more likely a surface is to contain a zero-crossing, the more likely it is to be inconsistent with the image intensities. Since the probability of the surface forcing an inconsistent zero-crossing is proportional to $\rho_R([p_0, p_1])$, this function can serve as a measure of the potential inconsistency of the surface.

Note that one may not be able to determine the exact form of the probability function, since it would require explicit knowledge about the positions of the light sources, as well as the form of the reflectance function R. However, this may not be important, as we shall see in the next chapter.

Further, it is possible that the original surface did contain a single inflection in this region, but that the reflectance function was such that the surface inflection was not manifest in the intensity functions. The point of the above argument is that such surfaces are unlikely, since they require a particular arrangement of the light source and viewer in order to hide such an inflection from the viewer. It is interesting psychophysically to question whether, under such conditions, we perceive

the surface as being uninflected. Certainly, this would seem likely, since there is simply no available information about the surface inflection point in the image intensities.

There is one final case to consider, namely that in which the surface does not contain any inflection points in the region bounded by two zero-crossings. We again seek a method for determining the probability of inconsistency of a particular surface.

Theorem 3: Consider a second differentiable developable surface oriented along the y axis such that $z_y(x, y) = c$, for some constant c. If the following conditions are true:

(1) At the point x_1, the surface becomes self-shadowing, that is $R(x_1) = 0$,

(2) The reflectivity R achieves an extremum at the point x_2, $x_1 \neq x_2$,

(3) At the point x_2 normalized changes in albedo are dominated by normalized changes in reflectance,

$$\left|\frac{\nabla^2 A}{A}\right| < \left|\frac{R_{pp}p_x^2}{R}\right|,$$

(4) The reflectance R is not constant over this region,

(5) The albedo A is non-zero,

then there exists a point $x_1 < x' < x_2$ such that $\nabla^2 I(x') = 0$.

Proof: The fact that the surface becomes self-shadowing implies that there is a region of the surface, beginning at x_1, such that R is constantly zero. The fact that there is an extremum in R for some other point implies that the intensity function must be concave down in the region of the extremum and concave up in the region of self-shadowing. There must be an inflection point in between and hence there must be a point x' such that $\nabla^2 I(x') = 0$. ∎

Theorem 4: Consider a second differentiable developable surface oriented along the y axis such that $z_y(x, y) = c$, for some constant c. If the following conditions are true:

(1) At the point x_1, the reflectivity R achieves an inflection point,

(2) There exist points $x_0 < x_1 < x_2$ such that reflectance changes dominate albedo changes,

$$\left|\frac{\nabla^2 A}{A} + \frac{R_p}{R}\left(2\frac{A_x}{A}p_x + p_{xx}\right)\right| < \left|\frac{R_{pp}}{R}p_x^2\right|,$$

(3) The first derivative of the surface, p, is monotonic in this region, (i.e. it does not achieve an extremum),

(4) The reflectance R is not constant over this region,

(5) The albedo A is non-zero,

then there exists a point $x_1 < x' < x_2$ such that $\nabla^2 I(x') = 0$.

Proof: The proof is very similar to the previous ones, except that in this case, by condition (2),

$$sgn(\nabla^2 I) = sgn(A)sgn(R_{pp})sgn(p_x^2).$$

at the points x_0, x_2. Then condition (4) implies that

$$sgn(\nabla^2 I) = sgn(R_{pp}).$$

Condition (3) implies that

$$p(x_0) < p(x_1) < p(x_2)$$

or

$$p(x_2) < p(x_1) < p(x_0)$$

In either case, condition (1) then implies that

$$sgn(R_{pp}(p(x_0))) \neq sgn(R_{pp}(p(x_2))).$$

Hence,

$$sgn(\nabla^2 I(x_0)) \neq sgn(\nabla^2 I(x_2)),$$

and as before, the second differentiability of the surface implies that there exists a point $x' \in (x_0, x_2)$ such that $\nabla^2 I(x') = 0$. ∎

The previous two theorems also allow one to estimate the probability of inconsistency of a particular surface. That is, if $p_0 = \min_x p(x), p_1 = \max_x p(x)$, then the probability that the surface forces an inconsistent zero-crossing in the image intensities is proportional to the probability that the reflectance function contains an inflection for this region of the surface, and is also proportional to the probability that the surface is self-shadowing for the particular imaging geometry.

All of these theorems relate the existence of a change in intensity to a set of conditions on the surface and its reflectivity function. Some of these conditions are of major interest, since they will enable us to measure the probability of a particular surface being consistent with the image information. Note that the other conditions of the theorems all seem reasonable. The requirement that A be non-negative is a trivial one, since it is only intended to exclude the case of a zero albedo factor, for which no light is reflected from the surface. Similarly, the condition that R be non-constant is to exclude the case in which a uniform amount of light is observed for the surface, independent of the orientation of the surface, which is probably physically impossible. The final condition usually concerns the relative strengths of albedo (or photometric) changes as opposed to reflective (or topographic) changes. The basic assuption in these theorems is that the changes in intensity due to albedo changes do not dominate the changes in intensity due to topographic changes.

Parenthetically, note that if in fact the albedo changes are much bigger than the topographic changes, this results in the situation where

$$\nabla^2 I \approx \left(\nabla^2 A\right) R$$

which means that photometric changes will result in zero-crossings in the Laplacian of the intensity. These are precisely the types of features used as motivation by Marr and Hildreth in the design of their method for creating primal sketch descriptions.

It is worth noting that the condition concerning the relative effects of albedo and topography in all the theorems can be relaxed from requiring the inequality at the points of inflection, to simply requiring that the inequality hold for some point within a neighbourhood about the points of inflection.

In general, note that in situations where the surface markings are strong, the density of zero-crossings is high, and the surface is well constrained by the known depth points. In situations where the surface markings are sparse, shading analysis constraints such as have been outlined in the above theorems will constrain the surface.

Thus we see that provided the changes in surface topography dominate changes in albedo, we have a method for determining the probability of a particular surface being the underlying surface of the image. Given the form of the surface, the probability of the surface being inconsistent with the image intensity information contained in the zero-crossings can be measured, since this is related to the probability of the surface forcing an inflection in the intensities which has not been detected

by the zero-crossings detector. This gives a method of assigning a probability function to the class of possible surfaces, measuring the acceptability of such surfaces.

To illustrate these arguments, Figures 5.5, 5.6, 5.7 and 5.8 show examples of a one-dimensional slice of a developable surface, with a Lambertian reflectance function, and the intensity values obtained for different positions of a point light source. Figure 5.5 indicates the sample surfaces and the rough positions of the light sources for the different examples. Figures 5.6, 5.7 and 5.8 indicate sample surfaces and the corresponding intensity profiles for different positions of the light source, as indicated in Figure 5.5.

NO INFORMATION IS INFORMATION

Figure 5.5. Examples of One-Dimensional Surfaces. The top figure shows one surface and the arrows indicate the rough orientations of the light source. The numbers refer to the intensity profiles in Figure 5.6. The bottom figure shows a second surface, with a set of rough orientations of the light source. The numbers refer to the profiles of Figures 5.7 and 5.8.

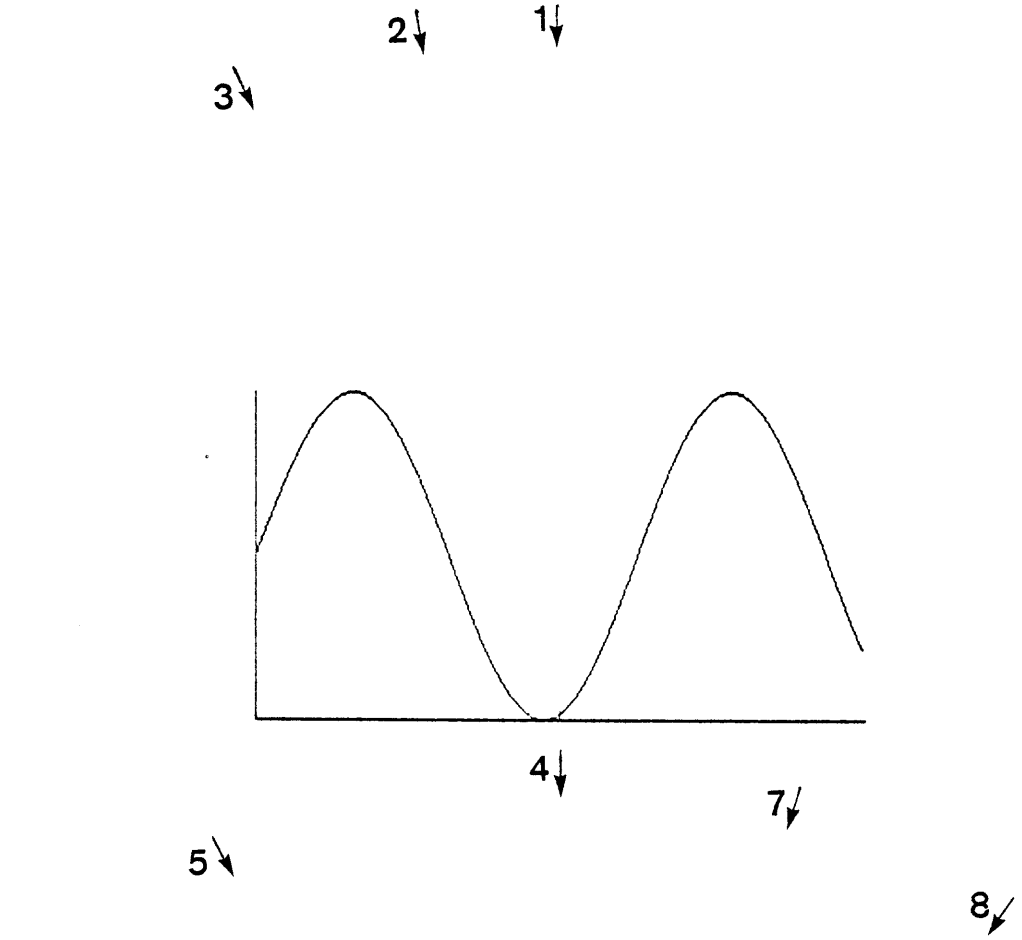
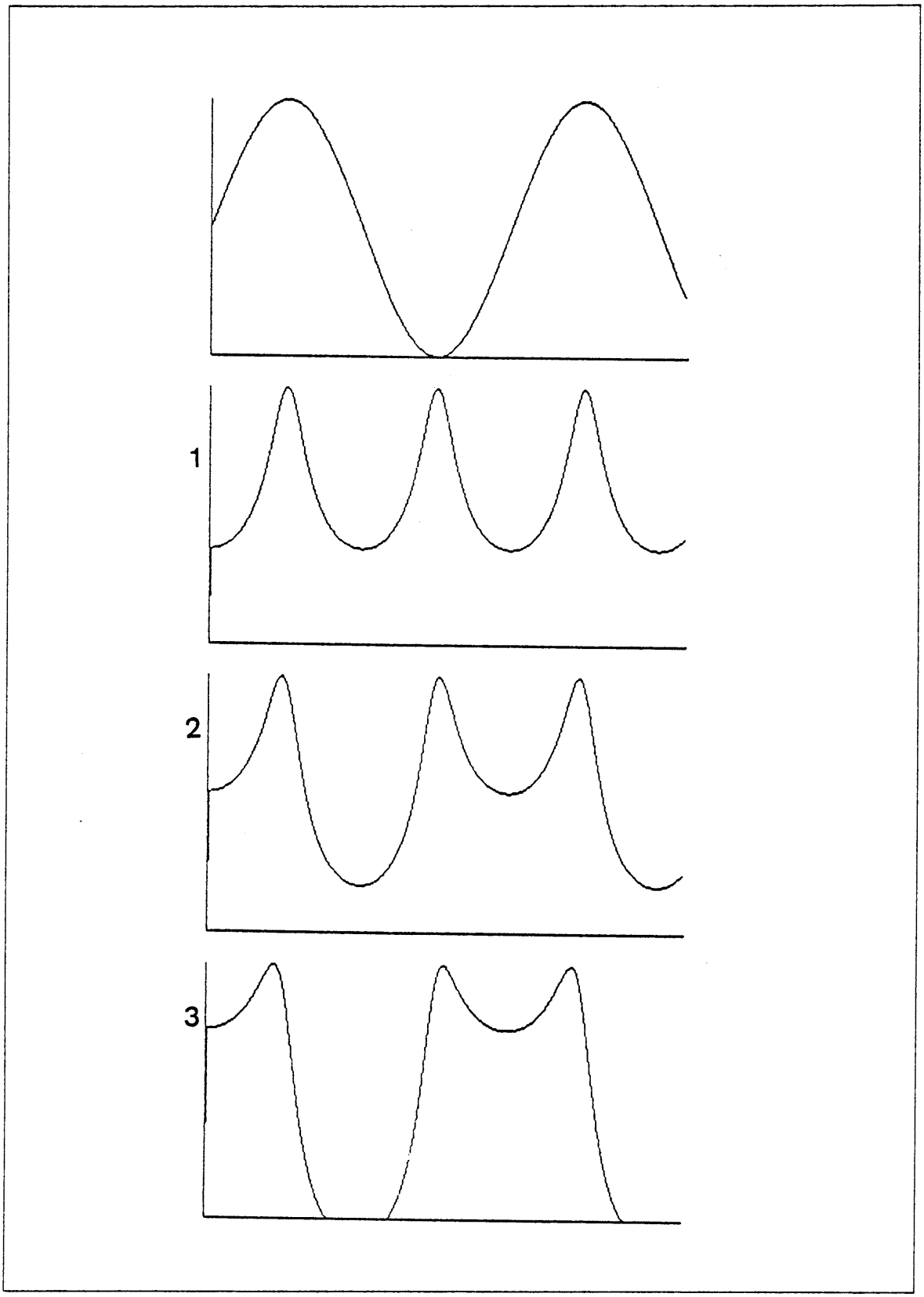

2
1
3
4
7
5
8
6

Figure 5.6. Examples of a Surface with Two Inflections. The top figure shows a slice of a surface. The bottom three figures indicate intensity profiles for different positions of the light source. Note that in all cases, there are six intensity inflections. In case 3, the intensities also undergo a self-shadowing, where the intensity value is zero.


1
2
3

NO INFORMATION IS INFORMATION

Figure 5.7. Examples of a Surface with One Inflection. The top figure shows a slice of a surface. The bottom three figures indicate intensity profiles for different positions of the light source.

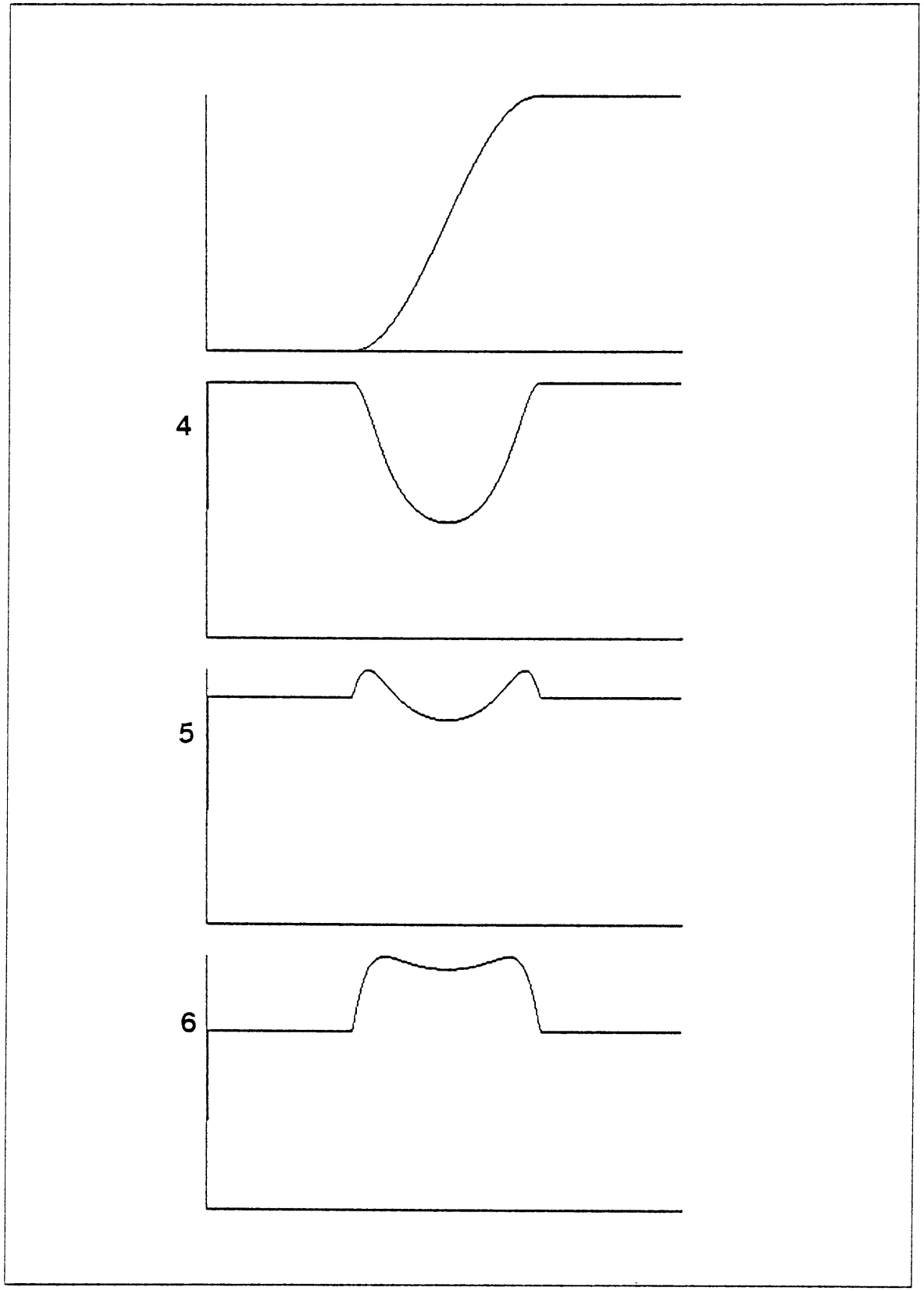
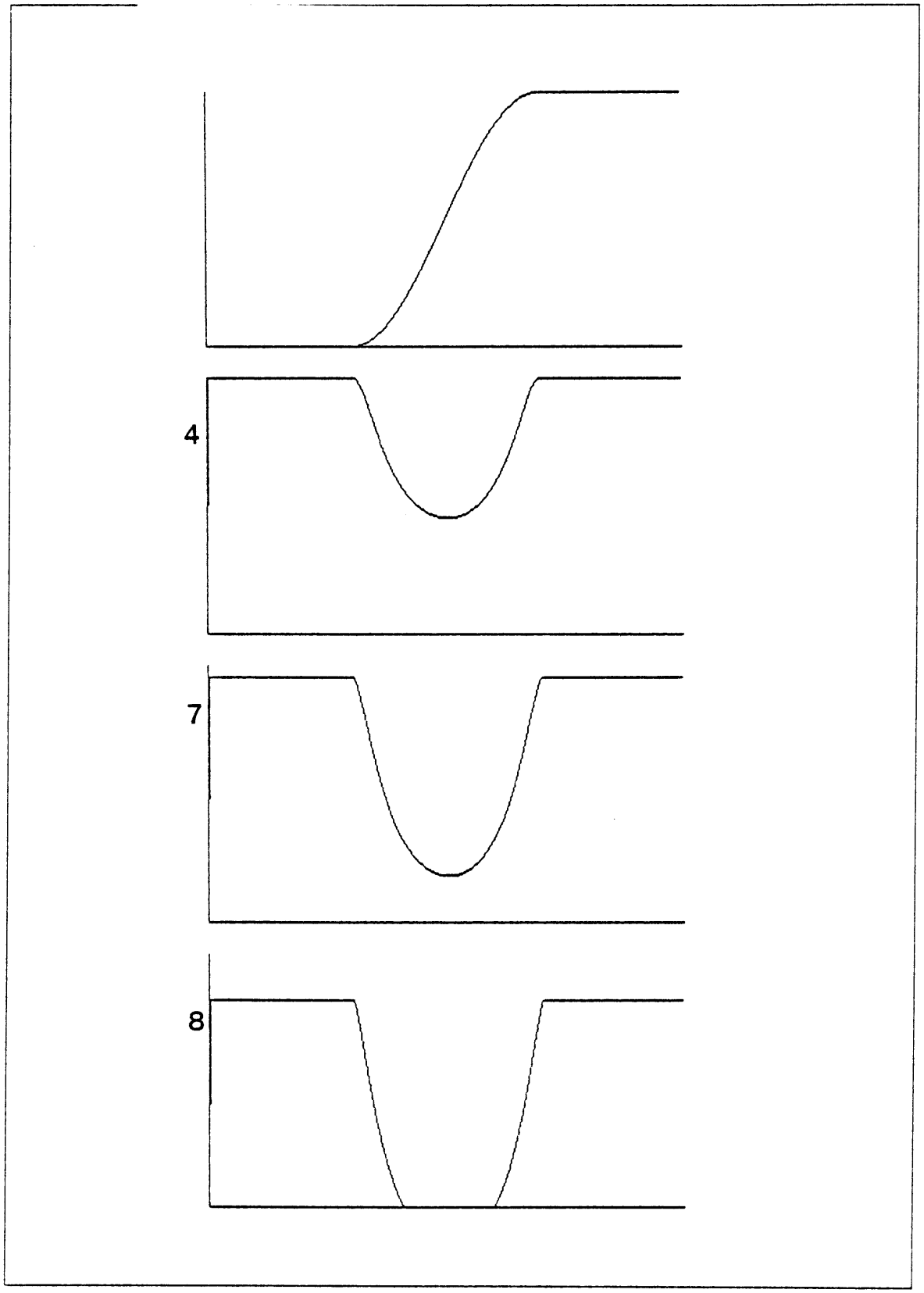

4
5
6

Figure 5.8. Examples of a Surface with One Inflection. The top figure shows a slice of a surface. The bottom three figures indicate intensity profiles for different positions of the light source. Note that in case 8, the intensity profile undergoes a self-shadowing.


4
7
8

5.3.2 The Two-Dimensional Case

One would like to consider extensions of these theorems to cases other than developable objects. In particular, there are other conditions in which one obtains topographically induced zero-crossings in $\nabla^2 I$, where the surface need not be developable. In the previous section, a set of theorems were proven analytically. In considering general extensions of these theorems in this section, a set of geometric arguments will be sketched.

For all of the arguments in this section, assume that the albedo effects can be ignored relative to the topographic, that is, the albedo factor may be considered constant. The general form of the argument is as follows. Consider a surface, and take a planar slice of the surface along some direction (such as the x direction). At each point along the resulting curve, one may associate a surface orientation, or gradient, given by a pair of partial derivatives, $p(x, y) = f_x(x, y)$, and $q(x, y) = f_y(x, y)$. Thus, one may construct a two-dimensional space spanned by a coordinate system with axes given by p and q, the gradient space introduced by Huffman (1971) and used by Mackworth (1973), and by Horn (1977) to relate the geometry of image projections to the radiometry of image formation. The curve obtained by the planar intersection of the surface transforms into a new curve in gradient space.

Because the albedo A is roughly constant, the image intensities are determined by the reflectance factor $R(p, q)$. Thus, the intensities may be related to Horn's reflectance map, in which one considers the surface defined by $R(p, q)$ in gradient space. Here, the curve on the original surface will map onto a curve on the reflectance surface $R(p, q)$.

The interest is in what conditions on the original surface and the reflectance surface will cause a sharp change in the image intensities. Note that the intensities along any parametric curve are given by the height of the corresponding point on the parametric curve on the reflectance surface. How does one relate an inflection in intensity to the form of this three-dimensional curve? Suppose one traces along the length of this curve. At any point consider the component of the gradient of R in the direction of motion — that is, as one continues to move along the curve, does one move upwards, downwards, or on the level? A change in the direction of the local gradient, from up to down or down to up, will correspond to a local extremum in intensity along the slice of the surface. Furthermore, a pair of changes in direction will imply that between them, there must be an inflection in intensity. This is the condition in which we are interested.

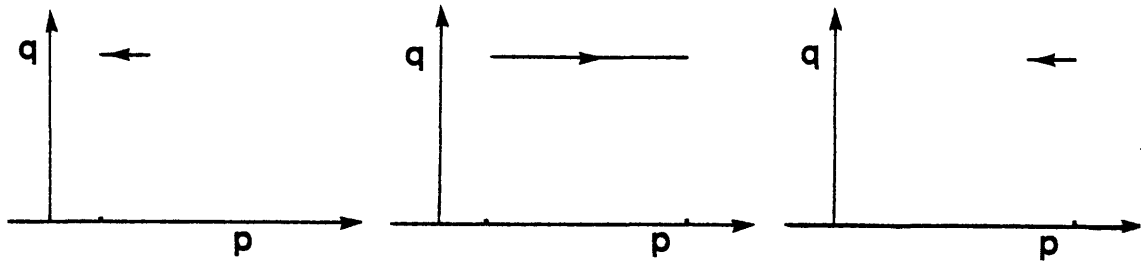


Figure 5.9. Gradient trace of surface with two inflections. If q is constant, then tracing along a surface slice with two inflection points results in three overlapping curve segments.

We can now recast our previous theorems in this new manner. The case of a developable surface is that in which q is constant along the slice of the surface. If the original surface has two inflections, along the surface slice, the function p has a local minimun: and a local maximum. The translation of this curve to gradient space results in a curve which consists of three overlapping segments as shown in Figure 5.9.

Thus, the first theorem states that provided the reflectance surface is not constant in this region of gradient space, there must be an inflection in image intensities. This can easily be seen. In particular, at the gradient points corresponding to the extrema in p, the local gradient of the curve in reflectance space must undergo a change in direction (since R is not constant). Thus, there are two such changes in direction and hence there must be an inflection in intensity at some point along the curve.

The second theorem concerned the case of only one inflection in the surface. In this case, the function p, as traced along the curve, will have only a single extremum. Hence, there will be only one change in the direction of the local gradient of the curve in reflectance space. However, if there is a point along the curve at which the reflectance function R undergoes an extremum, then this will result in a second change in the direction of the local gradient, and once again there must be an inflection in the image intensities.

Similarly, if the reflectance function passes from a self-shadowing through an extremum as one

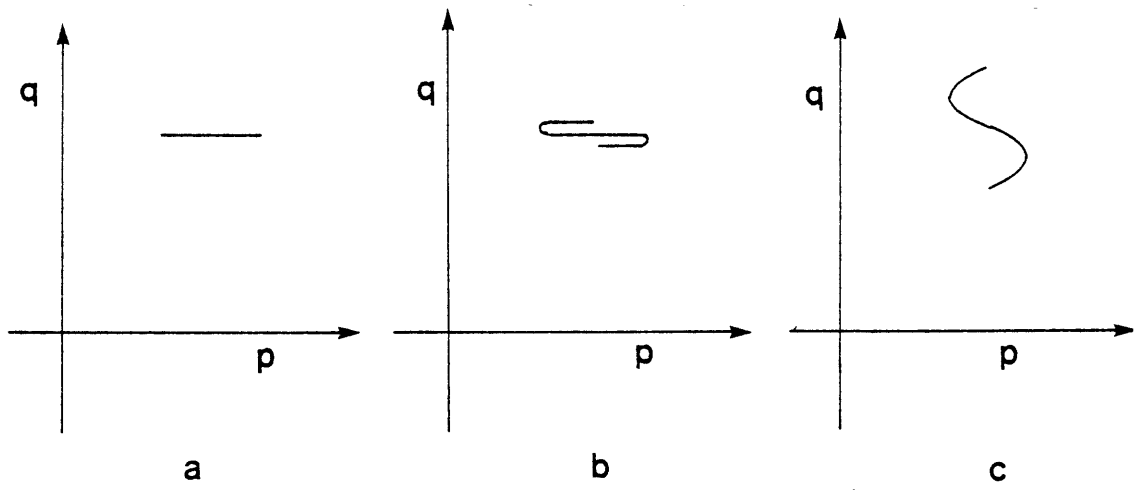


Figure 5.10. Gradient space curves. Figure (a) illustrates the case of q constant as p passes through two inflections along the slice. Figures (b) and (c) are cases of surface slices in which q is not constant along the curve. In case (b), almost all reflectance surface R will be such that two changes in the local gradient occur, as one traces along the curve. In case (c), this is less likely.

traces along the $p - q$ curve, then there again are two changes in the direction of the local gradient of the $p - q$ curve, and again there must be an inflection in intensity.

Suppose the condition that q be constant is relaxed. Do the same theorems still hold? Consider first the case in which the original surface has two inflections along the x direction, and hence, two extrema in p. In the case of q constant, we saw that the corresponding curve in gradient space had two sections of overlap. If q is not constant, then this is not the case, although the gradient space curve still contains two points where the curvature changes sharply. This is illustrated in Figure 5.10.

The case illustrated in Figure 5.10a is that of a developable surface. We have already discussed this. Consider the case illustrated in Figure 5.10b, where the reflectance space curve undergoes a sharp change in curvature at two points. It is possible for the reflectance surface to be such that the local gradient of the curve does not change direction at these points. This would be true both for the case of a constant reflectance, and for the case of a planar reflectance surface with no component of gradient in the p direction. However, for most reflectance surfaces, the direction of the local gradient will change, and once again there must be an inflection in the intensities. Thus, although the conditions under which we can explicitly state that there must be an intensity inflection are not as strong as in the

case of constant q, it is still the case that for most reflectance surfaces R, such changes in p must cause an intensity inflection. This is particularly true for cases where the changes in q are small compared to the changes in p, as indicated in Figure 5.10b. As the changes in q increase in magnitude, the $p - q$ curve approaches the form shown in Figure 5.10c. As this happens, the probability of the reflectance surface R having a form such that the local gradient along the curve changes twice decreases.

In a similar manner, one may argue that if the curve in gradient space passes through an extremum in the reflectance surface, and contains a point of sharp curvature (corresponding to an extremum in p in the original surface), then it is probable that there is an inflection in intensity. In the case of the gradient curve passing from a region of self-shadowing through an extremum in the reflectance surface, the same sort of argument holds again, as the curve must change its direction of local gradient at both locations.

Hence, we have argued geometrically that the following is true.

Theorem 5: Consider a surface, and some planar slice of that surface. This defines a curve on that surface, and hence a curve in gradient space. If any of the following are true:

(1) The gradient curve contains two points of sharp curvature (and in the limit, infinite curvature as the curve doubles back upon itself),

(2) The gradient curve contains one point of sharp curvature, and also passes through a local extremum (defined in terms of the intersection of the curve and the reflectance surface) of the reflectance surface,

(3) The gradient curve passes from a region of self-shadowing through a local extremum (defined in terms of the intersection of the curve and the reflectance surface) of the reflectance surface,

Then, for almost all reflectance surfaces, there will be an inflection in the image intensities taken along the line of the curve on the original surface. ∎

Note that by "almost all reflectance surfaces" I mean the following. In order for one of the above conditions to be true and not give rise to an inflection in image intensities, one requires a very particular form for the reflectance surface. In the space of all possible reflectance surfaces, such peculiar surfaces will form an extremely small subspace, and in general will not arise.

Furthermore, there is an interesting consequence of assuming that the viewer position is not

fixed. Suppose one considers a situation such as that given by Figure 5.10b. In order to avoid an intensity inflection, the reflectance surface must have a local form such that the surface gradient has no component in the p direction. If the viewer position is changed slightly, this will result in a transformation of the functions p and q, and hence the gradient curve will change, while keeping its basic form. Thus, if one is still to avoid an intensity inflection, the reflectance surface must again have a very restrictive form, such that the new region of gradient space spanned by the new curve also has no component in the direction perpendicular to the curve. This is even more unlikely to occur than the first situation, since we are applying even more stringent conditions upon the form of the reflectance surface. Thus, for most reflectance functions, small alterations in viewer position will tend to eliminate the conditions under which sharp changes in the gradient curve do not force inflections in intensities. Thus, the theorems we have developed in the preceeding section hold in general for almost all reflectance functions.

One final note should be made concerning the arguments of Theorem 5. These arguments are based on the use of a directional derivative along the line of the one-dimensional slice of the surface. Yet, the stereo algorithm uses the Laplacian as its differential operator. Fortunately, the arguments used in proving this theorem will still hold in the case of using a Laplacian, under a simple assumption. Marr and Hildreth (1980) note that if the *condition of linear variation*, which states that the intensity variation near and parallel to the line of zero-crossings should locally be linear, is valid, then one is justified in using a Laplacian rather than a directional derivative. If we assume the condition of linear variation as part of Theorem 5, then the arguments used to prove that theorem in the case of developable surfaces will still apply in the case of more general surfaces with linear intensity variation parallel to the zero-crossings.

5.3.3 Summary

Let us recapitulate the important aspects of this section. The basic hypothesis was that in order for a surface to be consistent with a given set of zero-crossings, not only must it give rise to an inflection in the intensities at those points, but it must also **not** give rise to zero-crossings anywhere else. This led to the assertion that under most situations, such a restriction would require that the surface not change in a radical manner, that it not contain a large number of inflection points, (defined along one-dimensional slices of the surface). It is difficult to prove this in general, since the image for-

mation equation includes terms dependent on the imaging process and on the light source geometry, as well as factors dependent on the photometric properties of the surface itself. However, under some fairly weak assumptions concerning the relative strengths of photometric and topographic changes, it is the case that a pair or more of inflections in the surface will cause an inflection in intensity. Further, although a single surface inflection will not in general cause a change in intensity, if the reflectivity function passes through an extremum as well, then the intensity function must contain a sharp change. Finally, if the surface does not contain any inflections, it is still possible to have an intensity change, either if the reflectivity function passes through an inflection, or if the reflectivity function passes through an extremum and the surface is self-shadowing.

The importance of these theorems is that they lead to a method for measuring the probability of a particular surface being inconsistent with the zero-crossing information. This in turn suggests that it will be possible to derive a method for measuring the best possible surface to fit the known information. We shall turn to this problem in the next chapter.

One final note concerns the use of $(\nabla^2 G) * I$ rather than $\nabla^2 I$. The inclusion of a Gaussian in the operator will result in a certain amount of smoothing of the image intensities, before the zero-crossings are detected. This means that the original surface may contain a certain amount of high frequency fluctuation which will give rise to zero-crossings, and hence will not be present when we reconstruct the surface. This is not criticial, since we are interested in a reasonable approximation to the original surface, and such fluctuations as are removed by the Gaussian will not constitute a significant alteration of the shape of the reconstructed surface, at that scale.

CHAPTER 6

THE COMPUTATIONAL PROBLEM

We are now ready to consider the computational problem associated with the task of constructing complete surface specifications consistent with the information contained in the zero-crossings. The modules of early visual processing, such as stereo or structure from motion, provide explicit information about the shapes of the surfaces at specific points in the images, corresponding to the zero-crossings of the convolved images. The theorems of the previous chapter indicate implicit information about the shapes of the surfaces between the zero-crossings. In this chapter, these two effects will be combined, in order to obtain a complete surface specification. In later chapters, an algorithm to solve this problem will be considered.

6.1 Formulating the Surface Consistency Constraint

The incoming information, with which the interpolated surfaces must be consistent, is that of depth values along the zero-crossings of the convolved image. Additional information can be derived from the theorems of Chapter 5. The fact that there do not exist other zero-crossings between the known ones, together with these theorems imply a constraint on the surface. I shall call this or *the surface consistency constraint.*

For simplicity, consider first the case of a developable or cylindrical surface. What do the theorems state about the shapes of such surfaces? Consider an example. Suppose that one is given the two boundary points as shown in Figure 6.1, and no additional zero-crossings. There are infinitely

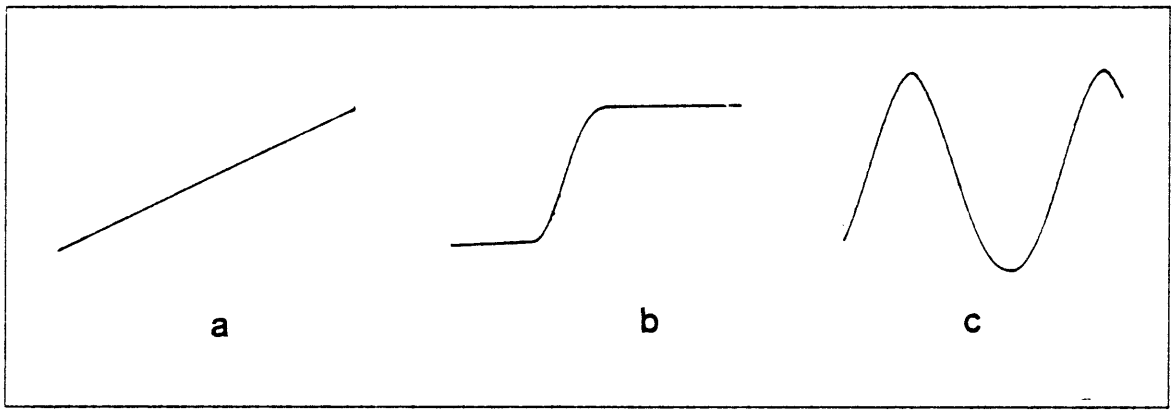


Figure 6.1. A boundary condition and some sample surfaces.

many second differentiable surfaces which will fit these conditions. Some examples are shown in the figure. How does one choose which surface of this collection is the "correct" one? The theorems of the previous chapter indicate that surface (c) is impossible under almost all conditions, because it would yield additional zero-crossings interior to the boundary zero-crossings, and no such zero-crossings exist. Surface (b) is less likely than surface (a), since there is a larger probability that (b) will be inconsistent with the known image intensities, in particular with those locations in the image corresponding to intensity inflections. Thus, for these three sample surfaces, the least inconsistent candidate is surface (a).

Yet, the problem is that the best one can do is simply state that (a) is more likely than (b), rather than stating that (a) is the correct surface. The reason for this is that we do not have enough information to determine exactly the correct surface. To see this, note that the theorems of Chapter 5 have the form: If some condition on the surface is true and some conditions on the reflectivity are true, then the convolved image intensities must have a zero-crossing. The contrapositive of this states that since there is no zero-crossing between the known ones, either the condition on the surface is false or one of the conditions on the reflectivity is false. One can safely assume that most of the conditions on the reflectivity are true, in particular, the conditions on the albedo not being zero and the reflectance not being constant. Thus, one must state that either the surface does not have a particular form, or that the reflectivity does not attain an extremum, inflection or self-shadowing. To decide which of these possibilities is the case would require explicit information about the form of the reflectivity function of the surface, including knowledge of the surface material, as well as explicit information

concerning the positions of the light sources. Such information will not generally be available in this situation. Since one of these two conditions must be incorrect, the probability that the surface must be constrained is directly related to the probability that the reflectivity function for that surface has an extremum, an inflection, or a self-shadowing.

This implies that in this situation, one will not be able to construct a direct relation between image intensities and surface shape. Note that this differs from the case of shape from shading (Horn, 1970) in which such a relationship is constructed. Although one cannot construct a direct relationship between the image intensities and the surface shape, one can construct a relationship between changes in the image intensities (the zero-crossings of the convolved images) and the surface shape. The theorems of the previous chapter indicate that this relationship is a probabilistic one. For a particular surface shape between a set of zero-crossing contours, one can associate a probability measure of that surface being inconsistent with the known inflections in the image intensities, based on the reflectance function of the surface. Using this measure of surface inconsistency, one can then find the surfaces that minimize this probability.

The interpolation problem can now be stated. Given a set of known depth points, consider all possible surface fitting through those points. One would like to be able to compare different surfaces, in order to determine which is the least inconsistent. The normal method for comparing surfaces is to assign to each surface a real number. Then, in order to compare the surfaces, one need only compare the corresponding real numbers. The assignment of real numbers to possible surfaces is accomplished by defining a functional, mapping the space of possible surfaces into the real numbers, $\Theta: X \mapsto \Re$. This functional should be such that the less inconsistent the surface, the smaller the real number assigned to it. This means that the functional should somehow capture the essence of the theorems of Chapter 5. In this case, the least inconsistent surface will be that surface which is minimal under the functional.

Interpolation refers to the case in which the surface exactly fits the set of known points. The problem can be relaxed somewhat into an approximation problem, requiring only that the surface approximately fit the known data and be smooth in some sense.

For both interpolation and approximation, there are two distinct classes of methods available — global and local methods. Of course, questions of global versus local belong at the level of the

algorithm, not the computational theory, which is independent of such considerations. I shall consider such methods in the next chapter, when we discuss possible algorithms for solving the computational problem.

6.1.1 The Form of the Functional

The major problem is to determine the functional Θ. The theorems of the previous chapter have related the consistency of a surface to the probability that the surface reflectance function contains an extremum, an inflection, or a self-shadowing. In principle, the functional Θ should measure this probability, denoted by ρ_R. If one could construct such a probability function, then finding the surface which minimizes this probability of inconsistency between the surface and the image intensities would be equivalent to finding the least inconsistent surface. This would constitute an optimal solution to the problem. Without explicit knowledge of the surface material and light source positions, it is not possible to explicitly construct this function ρ_R.

Since ρ_R is a probability measure, however, note that:

$$\rho_R(A \cup B) = \rho_R(A) + \rho_R(B)$$

$$\rho_R(A) \geq 0$$

for all elements A, B in the σ-algebra of ρ_R.

Although it may not be possible to directly minimize ρ_R, this observation implies that its value can be reduced for any surface by minimizing some other measure of the size of the argument to ρ_R. For example, in the one-dimensional case, the probability that R contains an inflection, an extremum or a self-shadow is directly proportional to the size of the interval $[p_0, p_1]$ where $p_0 = \min_x p(x)$ and $p_1 = \max_x p(x)$. Thus, by minimizing the size of $[p_0, p_1]$, the probability of a surface inconsistency, given by $\rho_R([p_0, p_1])$, can be reduced. For example, in Figure 6.2 this means that surface (a) is more consistent than surface (b), and that surface (c) is more consistent than surface (d), since in both cases, the range of $[p_0, p_1]$ is reduced.

The problem may be considered in the following manner. With every point on the surface, associate an orientation, and hence a pair of partial derivatives, $f_x = p, f_y = q$. Thus, each point on the surface may be mapped to a point in a space spanned by p and q axes, the gradient space

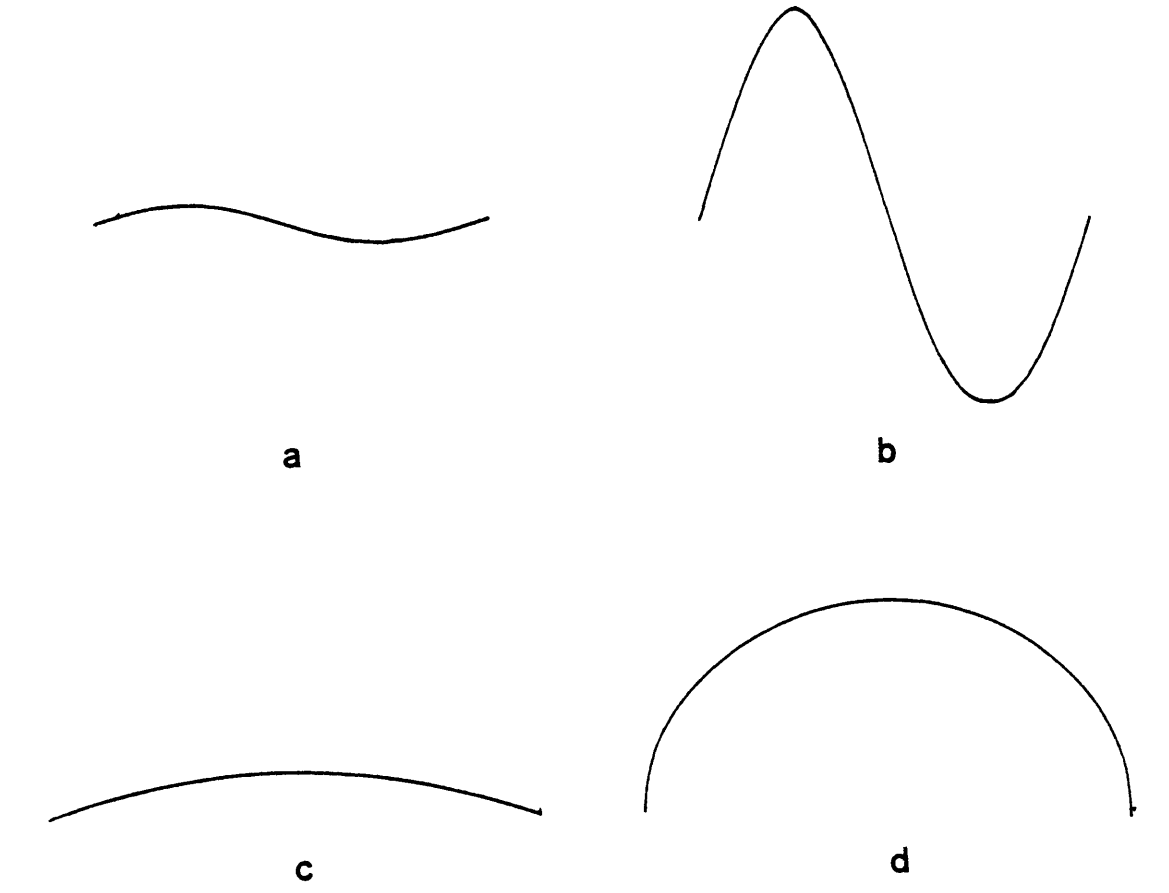


Figure 6.2. Consistency of surfaces. Surface (a) is more consistent with the zero-crossing information than surface (b). Similarly, surface (c) is more consistent than surface (d).

(Huffman, 1971; Mackworth, 1973; Horn, 1977). Hence, with each surface patch, one may associate a neighbourhood of $p - q$ space by mapping the p and q values associated with each point on the surface into gradient space. This neighbourhood will be referred to as the $p - q$ neighbourhood spanned by the surface patch.

Thus, for each patch of surface, the probability function ρ_R assigns a value based on the neighbourhood of $p - q$ space spanned by the surface patch. The observation about the monotonicity of probability measures implies that although the explicit form for the probability function may not be known, its value can be reduced by minimizing the size of the neighbourhood of $p - q$ space. Thus, rather than using the probability measure ρ_R, a measure of the $p - q$ neighbourhood spanned by a

surface patch will be used as the functional Θ. Then, the minimization of this functional will result in a reduction of the probability of surface inconsistency, and the surface defined in this way will form a reasonable approximation of the original underlying surface.

Of course, there are some constraints on the minimization of the size of the $p - q$ neighbourhood. For example, each surface patch cannot be minimized in isolation. To see this, consider a one-dimensional surface. Between any two adjacent zero-crossing points, the minimization of the neighbourhood in $p - q$ space would result in a single point, corresponding to a planar surface between the known depth values. The problem with this simple method of reducing surface inconsistency is that it does not account for the interaction of surface patches. In particular, such a method would result in a piecewise planar surface approximation. For any three consecutive zero-crossing points, such a method would introduce a discontinuity in surface orientation at the middle zero-crossing. This is unacceptable since it is required that the surface be second differentiable. Thus, there are some constraints on the manner in which the neighbourhoods in $p - q$ space are minimized.

In fact, in the nomenclature of gradient space, the problem may be phrased as follows. Between any two adjacent zero-crossings, one can define a patch of the surface f (actually a curve in the case of a cylindrical surface). This surface curve $f(x)$ can be mapped into a parametric curve in gradient space, $p(x)$. If f is continuous and bounded, then so is p and the support of p, defined as the set of all values of $p(x)$ as x traces over the surface patch f, is bounded and connected. The goal is to minimize the extent of the support of p for each section of the surface, subject to the following constraints:

(1) The support of $p(x)$ must contain the point p_0 corresponding to the orientation of the plane between the known depth points. (Due to the mean value theorem.)

(2) The integral of $p(x)$ over its support must equal the difference in depth between the zero-crossings.

(3) If f_1, f_2 are adjacent patches of the surface, then the supports of $p_1(x), p_2(x)$ must overlap.

(4) At one of the known points x', belonging to surface patches f_1 and f_2, the corresponding gradient values must be the same, $p_1(x') = p_2(x')$.

For a general two-dimensional surface, a similar set of constraints holds.

Thus, the computational argument has led to the following problem. We know from the analysis

of the images, that changes in image intensity occur at certain positions in the image, marked by the zero-crossings of the convolved image. At all other points in the image, no such intensity changes are evident. For each surface portion, we can associate a measure of the probability of that surface implying a zero-crossing in the convolved image intensities. This probability function has as its σ-algebra, the algebra of all neighbourhoods of gradient space. Since between zero-crossings, there are no other zero-crossings, this gives a measure on the inconsistency of this particular surface portion. To choose the least inconsistent surface, we seek to reduce this probability, as measured over all portions of the surface.

6.1.2 The Problem is Well-Defined

There exists standard matchmatical machinery for comparing surfaces, by defining a functional from the space of surfaces to the real numbers. In this section, that machinery is developed. For further details, see for example Rudin (1973).

A vector space V over a field Φ is a set, whose elements are called vectors, in which two operations, addition and scalar multiplication, are defined with the following algebraic properties.

(1) To every pair of vectors v and w correponds a vector $v + w$ such that

$$v + w = w + v \qquad \text{and} \qquad v + (w + u) = (v + w) + u;$$

V contains a unique vector $\mathbf{0}$, called the origin, such that $v + 0 = v$ for every $v \in V$; and to each $v \in V$ corresponds a unique vector $-v$ such that $v + (-v) = \mathbf{0}$.

(2) To every pair (α, v) with $\alpha \in \Phi$ and $v \in V$ corresponds a vector αv such that

$$1v = v, \qquad \alpha(\beta v) = (\alpha\beta)v,$$

and such that the distributive laws hold,

$$\alpha(v + x) = \alpha v + \alpha w, \qquad (\alpha + \beta)v = \alpha v + \beta v.$$

If V is a vector space, $W \subset V$, $U \subset V$, $v \in V$, and $\lambda \in \Phi$, the following notations are often used:

$$\begin{aligned} v + W &= \{v + w : w \in W\} \\ v - W &= \{v - w : w \in W\} \\ W + U &= \{w + u : w \in W, u \in U\} \\ \lambda W &= \{\lambda w : w \in W\}. \end{aligned}$$

FORMULATING THE SURFACE CONSISTENCY CONSTRAINT

A set $W \subset V$ is a *subspace* of V if

$$\alpha W + \beta W \subset W, \qquad \text{for all scalars} \quad \alpha, \beta.$$

A set $W \subset V$ is called *convex* if

$$tW + (1-t)W \subset W, \qquad 0 \leq t \leq 1.$$

A vector space V is called a *normed space* if there exists a function, called the *norm*, from the vector space to the nonnegative reals, $\Theta: V \mapsto \Re^+$, such that

(a) $\Theta(v+w) \leq \Theta(v) + \Theta(w) \qquad \forall v, w \in V$

(b) $\Theta(\alpha v) = |\alpha| \Theta(v) \qquad v \in V, \alpha$ scalar

(c) $\Theta(v) > 0 \qquad$ if $\quad v \neq 0.$

Condition (a) is called the triangle inequality and is the generalization of the observation in Euclidean space that the length of one side of a triangle is never larger than the sum of the other two sides.

A *semi-norm* on V is a real-valued function p on V such that

(a) $p(v+w) \leq p(v) + p(w)$

(b) $p(\alpha v) = |\alpha| p(v).$

Condition (a) is called the subadditivity property.

Lemma: A number of trivial facts are true of semi-norms.

(a) $p(0) = 0$

(b) $p(v) \geq 0$

(c) $\mathcal{N} = \{v: p(v) = 0\}$ is a subspace of V.

Proof: Since $p(\alpha v) = |\alpha| p(v)$, (a) is true, by setting $\alpha = 0$. The subadditivity implies that

$$p(v) = p(v - w + w) \leq p(v - w) + p(w).$$

Thus, $p(v) - p(w) \leq p(v - w)$. Similarly, $p(w) - p(v) \leq p(w - v)$. Since $p(v - w) = p(w - v)$, it follows that

$$|p(v) - p(w)| \leq p(v - w).$$

If $w = 0$ then $|p(v)| \leq p(v)$ which implies $p(v) \geq 0$.

Finally, consider $\mathcal{N} = \{v{:}p(v) = 0\}$, and let $v, w \in \mathcal{N}$. Clearly, $p(\alpha v + \beta w) \geq 0$. But $p(\alpha v + \beta w) = |\alpha| p(v) + |\beta| p(w) = 0$. Thus, $\alpha v + \beta w \in \mathcal{N}$ and $\mathcal{N}$ is a subspace. ∎

Any space with a semi-norm defined on it can be associated with an equivalent normed space. To see this, let W be a subspace of a vector space V. For every $v \in V$, let $\pi(v)$ be the coset of W that contains v,

$$\pi(v) = v + W.$$

These cosets are elements of a vector space V/W called the *quotient space of* V *modulo* W. In this space, addition is defined by

$$\pi(v) + \pi(w) = \pi(v + w)$$

and scalar multiplication is defined by

$$\alpha\pi(v) = \pi(\alpha v).$$

The origin of the space V/W is $\pi(0) = W$. Thus, π is a linear map of V onto V/W with W as its null space, called the *quotient map of* V *onto* V/W.

The important point about quotient spaces is the following. Suppose p is a semi-norm on a vector space V. Let

$$\mathcal{N} = \{v{:}p(v) = 0\}.$$

This is a subspace, by the above discussion. Let π be the quotient map from V onto $V/\mathcal{N}$, and define a mapping $p'{:}V/\mathcal{N} \mapsto \Re$,

$$p'(\pi(v)) = p(v).$$

If $\pi(v) = \pi(w)$ then $p(v - w) = 0$. Since $|p(v) - p(w)| \leq p(v - w)$, then $p'(\pi(v)) = p'(\pi(w))$ and p' is well-defined on $V/\mathcal{N}$. It is straightforward to show that p' is a norm on $V/\mathcal{N}$.

The point of the previous discussion is that now all facts about normed spaces apply to semi-normed spaces, to within a factor of the null space $\mathcal{N}$, since the quotient space defined by a semi-norm is always a normed space.

Given a sequence of vectors, $\{v_n\}$ in a vector space V and a norm Θ on that space, the sequence is said to be *Cauchy* if for all $\epsilon > 0$ there is an integer M such that

$$\Theta(v_m - v_n) < \epsilon \qquad \forall m, n > M.$$

For a sequence of vectors $\{v_n\}$ to converge, there must exist some vector v such that for every $\epsilon > 0$, there exists some integer M such that for all $m > M$, $\Theta(v - v_m) < \epsilon$. If every Cauchy sequence converges in a normed space, then the norm is said to be *complete.* A norm is said to satisfy the *parallelogram law* if

$$\Theta^2(v + w) + \Theta^2(v - w) = 2\Theta^2(v) + 2\Theta^2(w).$$

We shall rely on a simple but important property concerning certain types of norms.

First, note that given a set of functions, $\mathcal{F}$, defined on some space V, the set $\mathcal{F}$ also forms a vector space. To see this, let $f, g \in \mathcal{F}$ and define the vector $f + g$ by

$$(f + g)(v) = f(v) + g(v).$$

Similarly, if α is an element of the scalar field, define the vector αf by

$$(\alpha f)(v) = \alpha f(v).$$

Using these definitions of addition and scalar multiplication, it can easily be shown that $\mathcal{F}$ forms a vector space.

Theorem: Suppose there exists a complete norm Θ on a space of functions H, which satisfies the parallelogram law. Then, every nonempty closed convex set $E \subset H$ contains a unique x of minimal norm.

Proof:(See for example Rudin, 1973) The parallelogram law states

$$\Theta^2(v + w) + \Theta^2(v - w) = 2\Theta^2(v) + 2\Theta^2(w).$$

We let

$$d = \inf\{\Theta(v) : v \in E\}.$$

Choose $v_n \in E$ such that $\|v_n\| \mapsto d$. By the convexity of E, we know that $\frac{1}{2}(v_n + v_m) \in E$ and so $\|v_n + v_m\|^2 \geq 4d^2$. If v and w are replaced in the definition of the parallelogram law by v_n and v_m, then the right hand side tends to $4d^2$. But $\|v_n + v_m\|^2 \geq 4d^2$, so one must have $\|v_n - v_m\|^2 \mapsto 0$ to preserve the equality. Thus, $\{v_n\}$ is a Cauchy sequence in H. Since the norm is complete, the sequence must converge to some $v \in E$, with $\Theta(v) = d$.

To prove the uniqueness, if $v, w \in E$ and $\Theta(v) = d, \Theta(w) = d$ then the sequence $\{v, w, v, w, \ldots\}$ must converge, as we just saw. Thus $v = w$ and the element is unique. ∎

Corollary 1: If Θ is a semi-norm, then the same holds true, except that now the element is unique up to an element of the null space of the semi-norm,

$$\mathcal{N} = \{f \in H : \|f\| = 0\}.$$

Corollary 2: Let X be a vector space of "possible" functions on $\Re^2$ and let

$$U = \{f \in X \mid f(x_i, y_i) = F_i \quad i = 1, \ldots, N\}$$

so that U is the set of functions which interpolate the known data $\{F_i\}$. Let Θ be a semi-norm, which measures the "inconsistency" of a function $f \in X$ — f is "better" than g if $\Theta(f) < \Theta(g)$. If Θ is a complete semi-norm and satisfies the parallelogram law, then there exists a unique (to within a function of the null space of Θ) surface $s \in U$ which is least inconsistent and interpolates the data. Hence the interpolation problem is well-defined.

Proof: Clearly U is convex since for any $f, g \in U$,

$$(\lambda f + (1 - \lambda)g)(x_i, y_i) = (\lambda + 1 - \lambda)F_i = F_i$$

for any data point f_i. Furthermore, U is closed, since if $f_n \in U$ and $f_n \mapsto f$, then $f(x_i, y_i) = F_i$ and $f \in U$. Then the previous theorem states that U has a unique (to within an element of the null space) element of minimal norm, which is exactly the desired "least inconsistent" surface. ∎

6.1.3 The Physical Meaning of the Criteria

In the previous section, a set of mathematical criteria were developed that ensure a unique solution to the interpolation problem. The criteria were that a complete semi-norm which satisfies the

parallelogram law should be defined over a space of possible functions such that the more consistent a surface f is with the zero-crossing information, the smaller the value of the semi-norm $\Theta(f)$ should be. Although these criteria are well founded mathematically, one would also like to understand what they mean in terms of the physical behaviour of the surfaces.

There are essentially four criteria involved. The first is that any semi-norm must be linear in the sense of

$$\Theta(af) = |a|\Theta(f)\,.$$

What does this mean in terms of the surfaces? Consider some surface, interpolating a set of known values. For each section of this surface, there is a neighbourhood of $p - q$ space spanned by it. We have seen that the consistency of the surface is proportional to the size of this neighbourhood. If the surface is scaled by some factor a, the values of p and q over this portion of the surface will also be scaled, and hence the entire $p - q$ neighbourhood will be scaled. As a consequence, the consistency of the surface should also be scaled by the same factor, so that

$$\Theta(af) = |a|\Theta(f)\,.$$

The second criterion is that the triangle inequality hold,

$$\Theta(f + g) \leq \Theta(f) + \Theta(g)\,.$$

Consider two surfaces f and g and let the difference between them be a third surface $\epsilon = f - g$. It is clear that the size of the $p - q$ neighbourhood spanned by any section of f must be bounded by the size of the $p - q$ neighbourhood spanned by g plus the size of the $p - q$ neighbourhood spanned by ϵ. Hence, by the properties of a probability measure, the triangle inequality must hold.

The other two mathematical conditions are that the semi-norm be complete and satisfy the parallelogram law. Both of these conditions are required to ensure that the semi-norm has a unique minimum over a convex subset of the space. That is, these conditions will guarantee a unique solution to the interpolation problem.

Thus, the criteria on the interpolation problem, originally motivated from a mathematical argument, can be seen to have an intuitive basis in the desired shapes of the surfaces.

6.1.4 The Space of Functions

We have seen that the computational problem is well-defined. There are two factors still to consider. The first is the choice of a vector space of functions which will constitute the set of possible surfaces. The second is to determine the semi-norm on that space which will measure the desired notion of "least inconsistent" as defined by the theorems of Chapter 5. The question of which space of functions to use is considered first.

An inner product is a function $\mu: V \times V \mapsto \Re$ written $\mu(v, w) = (v, w)$, satisfying

(1) $(y, x) = (x, y)$

(2) $(x + y, z) = (x, z) + (y, z)$

(3) $(\alpha x, y) = \alpha(x, y) \quad \alpha \in \Re$

(4) $(x, x) \geq 0$

(5) $(x, x) = 0$ if and only if $x = 0$.

A real vector space V is called an *inner product space* if there is an inner product defined on it. Note that if condition (5) does not hold, then the function is called a semi-inner product, and the space is a semi-inner product space. If μ is an inner product, then $\Theta(v) = \mu(v, v)^{\frac{1}{2}}$ is a norm on the space. This can easily be shown; a more difficult criterion to check is the triangle inequality.

To see this, note the following. It is easy to determine that

$$0 \leq \Theta^2(\lambda v + w) = |\lambda|^2 \Theta^2(v) + 2\mu(v, w)\lambda + \Theta^2(w).$$

If $v \neq 0$, take

$$\lambda = \frac{-\mu(v, w)}{\Theta^2(v)}.$$

Then the above expression becomes

$$0 \leq \Theta^2(\lambda v + w) = \Theta^2(w) - \frac{|\mu(v, w)|^2}{\Theta^2(v)}.$$

Hence,

$$|\mu(v, w)| \leq \Theta(v)\Theta(w).$$

To show the triangle inequality, note that

$$\begin{aligned} \Theta^2(v + w) &= \Theta^2(v) + \Theta^2(w) + 2\mu(v, w) \\ &\leq \Theta^2(v) + \Theta^2(w) + 2\Theta(v)\Theta(w) \\ &= (\Theta(v) + \Theta(w))^2. \end{aligned}$$

Thus, every inner product space can be converted to a normed space. If the resulting normed space is complete, then it is called a *Hilbert space.*

Lemma: If X is a semi-Hilbert space of possible surfaces, and Θ is a semi-inner product norm, then there exists a unique (to within an element of the null space of Θ) surface in X which minimizes the norm Θ over all surfaces.

Proof: By the definition of Hilbert space, the norm is complete. It is easy to show that it satisfies the parallelogram law from the definition of $\Theta(v) = \mu(v, v)^{\frac{1}{2}}$. Thus, if the space of functions considered is a Hilbert space, then, by the previous theorem, the interpolation problem is guaranteed to have a unique minimal solution. Recall that the null space of Θ is given by

$$\mathcal{N} = \{f \in X : \Theta(f) = 0\}.$$

If the space of functions is a semi-Hilbert space, then the problem has a minimal element which is unique up to an element of the null space, $\mathcal{N}$. ∎

Thus, in the interpolation problem, the choice for the space of functions is clear. However, as one might expect, there are several possible definitions of the pseudo norm, Θ. In most cases, these will give very similar surfaces, and in the next section I shall outline a number of them.

6.1.5 Possible Functionals

In this section, we consider norms Θ, seeking complete, parallelogram semi-norms where possible, since then we are guaranteed that the solution is unique to within the null pace. However, it is important to stress that there may be several viable alternatives, and that the choice may not be critical. In the next section, when an algorithm for solving the computational problem is created, an algorithmic method which applies to any semi-norm will, in fact, be created.

The computational theory argued that the functional should measure the "inconsistency" or "likelihood" of the surface. We have observed that it is the range of gradient values between zero-crossings that matters. The attempt here is to define a measure based on this. The measure should be in a form which allows the constraints on the problem easily to be expressed. Also, as far as possible, the measure should be a semi-norm on a semi-Hilbert space, in order to ensure a unique solution. There are many possible forms for this measure, and several of them will be considered in light of these conditions.

6.1.5.1 Case 1: One Dimension

Example 1.1: In the one-dimesional case, the range of gradient values is related to the total arc-length of the curve:

$$\Theta_1(f) = \int \left(1 + f_x^2\right)^{\frac{1}{2}} dx$$

However, this is not a semi-norm and there is no straightforward way of transforming it into one.

Example 1.2: One can also use the curvature of the curve:

$$\Theta_2(f) = \left\{ \int \frac{f_{xx}^2}{\left(1 + f_x^2\right)^3} dx \right\}^{\frac{1}{2}}$$

Although this is perhaps the most "natural" definition of a functional, it is not a semi-norm, and hence it is considered to be unacceptable. To see why it is not a semi-norm, consider the following. If f is in the space of surfaces, then

$$\begin{aligned} \Theta_2(\alpha f) &= \left\{ \int \frac{\alpha^2 f_{xx}^2}{\left(1 + \alpha^2 f_x^2\right)^3} dx \right\}^{\frac{1}{2}} \\ &= |\alpha| \left\{ \int \frac{f_{xx}^2}{\left(1 + \alpha^2 f_x^2\right)^3} dx \right\}^{\frac{1}{2}} \\ &\neq \alpha \Theta(f). \end{aligned}$$

This condition will be true only if $f_x \equiv 0$. While this is certainly far too restrictive a condition to place on the possible surfaces, it does suggest a possible alternative.

Example 1.3: A third choice is the quadratic variation of the gradient, which may be measured by:

$$\Theta_3(f) = \left\{ \int f_{xx}^2 dx \right\}^{\frac{1}{2}}$$

Note that it is a close approximation to the curvature of the curve, provided that f_x is small. Θ_3 is a semi-norm, so the surface which minimizes this norm will be unique to within an element of the null space of the semi-norm. The null space of Θ_3 is the set of all linear functions:

$$\mathcal{N} = \text{span}\{1, x\}$$

where

$$\text{span}\{v_1, \ldots, v_m\} = \{a_1 v_1 + \ldots + a_m v_m \mid a_1, \ldots, a_m \in \Re\}.$$

Not only does this form of the functional satisfy the mathematical criteria of a complete, parallelogram semi-norm, it has a strong relationship to the "natural" form $\Theta_2(f)$, since the restriction of f_x small is acceptable. Those cases in which f_x is not negligible correspond to situations in which the surface is rapidly curving away from the viewer, corresponding to occluding boundaries. Such occurances will be rare in an image. Moreover, between such points, the surface will satisfy the restriction and the above semi-norm is well-suited to the interpolation problem.

6.1.5.2 Case 2: Two Dimensions

To each of the examples of the one-dimensional case, there is an analogue in the two-dimensional case.

Example 2.1: One possibility is to measure the total surface area which is related to the range of surface orientation. This is known as Plateau's problem. The functional for this case is given by

$$\Theta_4(f) = \int\int \left(1 + f_x^2 + f_y^2\right)^{\frac{1}{2}} dxdy$$

Example 2.2 As in the one-dimensional case, a second possibility is to measure the curvature of the surface. The curvature of a surface is usually measured in one of two ways.

For any point on the surface, consider the intersection of the surface with a plane containing the normal to the surface at that point. This intersection defines a curve, and the curvature of that curve can be measured as the arc-rate of rotation of its tangent. For any point, there are infinitely many normal sections, each defining a curve. Two of these sections may be distinguished. If the normal plane is rotated slightly, a new normal section is defined, and the curvature of this curve at the point in question can be measured. As the normal section is rotated through 2π radians, all possible normal sections will be observed. There are two sections of particular interest, that which has the maximum curvature and that which has the minimum. It can be shown that the directions of the normal sections corresponding to these sections are orthogonal. These directions are the *principal directions* and the curvatures of the normal sections in these directions are the *principal curvatures*, denoted κ_a and κ_b. It can be shown that the curvature of any other normal section is defined by the principal curvatures.

There are two standard methods for describing the curvature of the surface, in terms of the principal curvatures. One is the first (or mean) curvature of the surface

$$J = \kappa_a + \kappa_b.$$

The other is the second or Gaussian curvature of the surface

$$K = \kappa_a \cdot \kappa_b.$$

For a surface defined by $\{x, y, f(x, y)\}$, these curvatures are given by

$$J = \frac{\partial}{\partial x}\frac{f_x}{\sqrt{1+f_x^2+f_y^2}} + \frac{\partial}{\partial y}\frac{f_y}{\sqrt{1+f_x^2+f_y^2}}$$

and

$$K = \frac{f_{xx}f_{yy} - f_{xy}^2}{\left(1+f_x^2+f_y^2\right)^2}.$$

Thus, there are two possibilities for the functional. One is to measure the first (or mean) curvature of the surface,

$$\Theta_5(f) = \left\{\int\int J^2\,dxdy\right\}^{\frac{1}{2}}$$

$$= \left\{\int\int \frac{\left(f_{xx}(1+f_y^2) + f_{yy}(1+f_x^2) - 2f_x f^y f_{xy}\right)^2}{\left(1+f_x^2+f_y^2\right)^3}\,dxdy\right\}^{\frac{1}{2}}.$$

As in the one-dimensional case, this is not a semi-norm, since

$$\Theta_5(af) = |a|\left\{\int\int \frac{\left(f_{xx}(1+a^2f_y^2) + f_{yy}(1+a^2f_x^2) - 2af_x af^y f_{xy}\right)^2}{\left(1+a^2f_x^2+a^2f_y^2\right)^3}\,dxdy\right\}^{\frac{1}{2}}$$

$$\neq |a|\Theta_5(f).$$

However, if f_x and f_y are assumed to be small, then it is closely approximated by a semi-norm. In this case, consider

$$\Theta_6(f) = \left\{\int\int \left(\nabla^2 f\right)^2 dxdy\right\}^{\frac{1}{2}}$$

This is a semi-norm, with null space consisting of all harmonic functions.

A second possibility for reducing curvature is to reduce the second or Gaussian curvature,

$$\Theta_7(f) = \left\{\int\int K^2\,dxdy\right\}^{\frac{1}{2}}.$$

By an argument similar to the above, it can be shown that this is not a semi-norm.

Example 2.3: As in the one-dimensional case, one can also consider the quadratic variation. The quadratic variation in $p = f_x$ is given by

$$\int\int \left(p_x^2 + p_y^2\right) dxdy$$

and the quadratic variation in $q = f_y$ is given by

$$\int\int \left(q_x^2 + q_y^2\right) dxdy$$

If the surface is twice continuously differentiable, then $p_y = q_x$, and by combining these two variations, one gets the quadratic variation:

$$\Theta_8(f) = \left\{\int\int \left(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2\right) dxdy\right\}^{\frac{1}{2}}$$

Again, as in the one-dimensional case, this is a complete semi-norm which satisfies the parallelogram law. Hence, the space of interpolation functions has an element of minimal norm, which is unique up to an element of the null space, where the null space is the set of all linear functions:

$$\mathcal{N} = \text{span}\{1, x, y\}$$

Duchon (1972, 1973) refers to the surfaces which minimize this expression as *thin plate splines* since the expression Θ_8 relates to the energy in a thin plate forced to interpolate the data.

Note that it is also possible to measure the quadratic variation in surface orientation, rather than the quadratic variation in gradient. However, the quadratic variation in orientation is not a semi-norm.

From this mathematical analysis, it seems that the best candidate for the semi-norm Θ is that given by the quadratic variation in gradient, as indicated above. In any case, in the next chapter, a method of solving this computational problem is developed, which is independent of the form of the semi-norm Θ. An example of the implementation of a particular semi-norm will then be considered.

6.2 Subjective Contours

Although our interest lies with the interpolation of surfaces, it is interesting to note the applicability of the computational problem of surface interpolation to a related problem, that of subjective contours. Subjective contours arise when the visual system fills in the gap between distinct edges

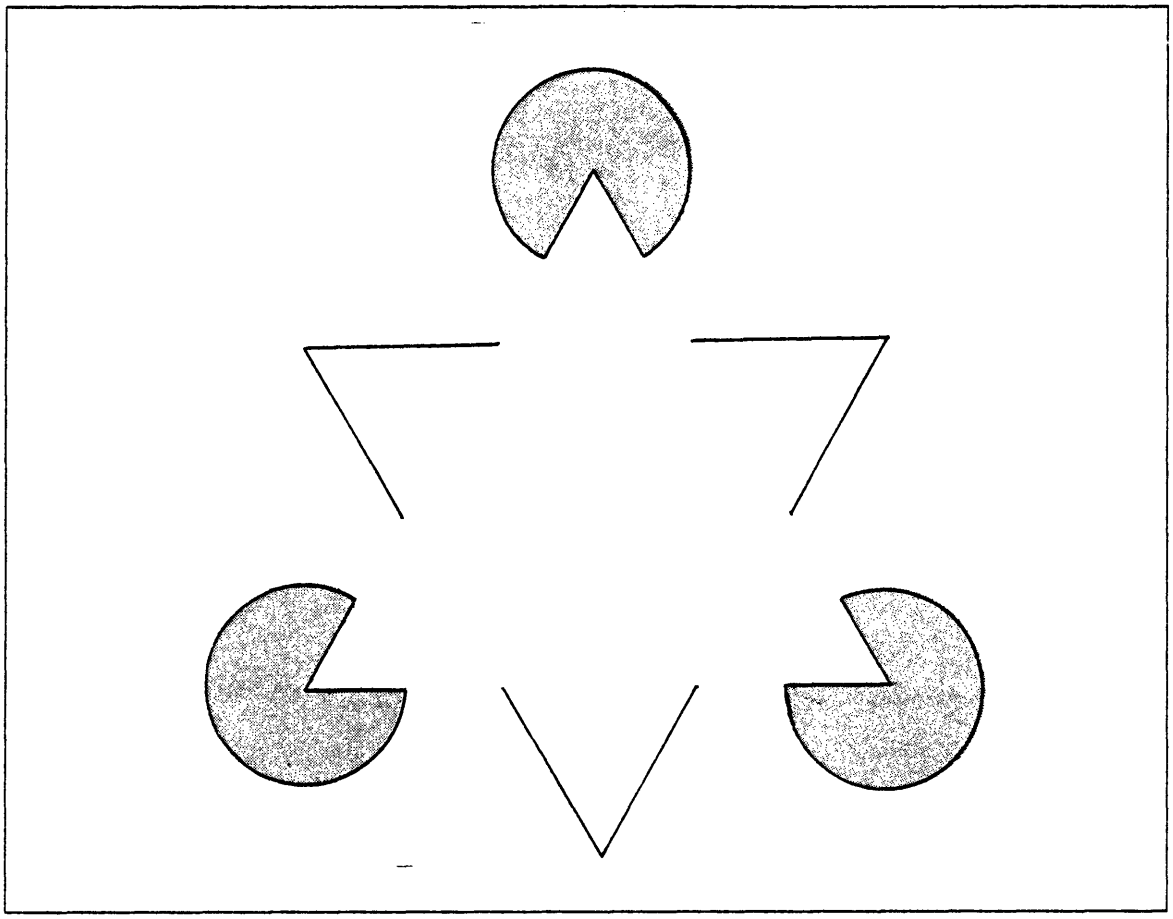


Figure 6.3. Subjective Contours. The perception associated with the figure is of a triangle occluding a set of circles. Although the sides of the triangle are not explicit in the figure, the visual system fills in the gaps to form subjective contours.

in the image (Brigner and Gallagher 1974, Coren 1972, Coren and Theodor 1975, Gregory 1972, Gregory and Harris 1974, Kanizsa 1976, Schumann 1904). An example is shown in Figure 6.3 (after Kanizsa, 1976).

Much of the interest in the literature has been concerned with the triggering conditions, under which the perception of the contour will occur. A second question is that of the shape of the subjective contour, and the previous discussion is relevant to this question. Ullman (1976) has suggested that the form of the subjective contour is given by the curve which minimizes

$$\int \frac{f_{xx}^2}{(1+f_x^2)^3}\,dx.$$

The previous discussion suggests that this is not the most optimal form for the functional which measures the consistency of curves. Rather, it suggests that an alternative would be to use the curve

which minimizes

$$\int f_{xx}^2 dx.$$

This is an approximation to the first case, valid when f_x is small. This suggests that if the difference in orientation of the endpoints of the subjective contour is small, then the perception of the subjective contour should be strong, while if the difference in orientation is large, the approximation is less valid and the perception should be weak.

A forthcoming paper (Brady, Grimson and Langridge, 1980) investigates these questions in more detail.

6.3 Relevance to the Human System

The initial intention was to develop a computational theory of the interpolation problem which was valid for general mathematical situations. However, since the processes which provide the input to the interpolation process are all parts of the human visual system, one must also consider the relevance of the interpolation problem to the human visual system.

A method for constructing more complete representations than those given, for example, by the Marr-Poggio stereo algorithm, has been developed. The question of whether the human system constructs such a specific representation is now considered.

There are two ways of considering this question. The first is by asking what the representation of the surfaces is used for, and whether such utilization requires a complete representation. Marr and Nishihara (1978) view the $2\frac{1}{2}$-D sketch as an intermediate representation, combining information about the visible surfaces from several independent sources. As such, it is used by higher level processes to extract three-dimensional descriptions of objects, descriptions better suited to perceptual tasks such as recognition. A complete specification could certainly aid the extraction of descriptions, simply because it makes information explicit and hence easier to access. However, there is no compelling evidence that such a representation is necessary for higher level processes. Thus, we must seek evidence elsewhere.

The second way of considering the form of the representation in the human system is to test it psychophysically. The issue here is basically how specific and explicit a representation is created. Thus, in some sense, the output of the stereo algorithm constitutes a complete representation of the surfaces to within a particular resolution. The point of interest here is whether the human system

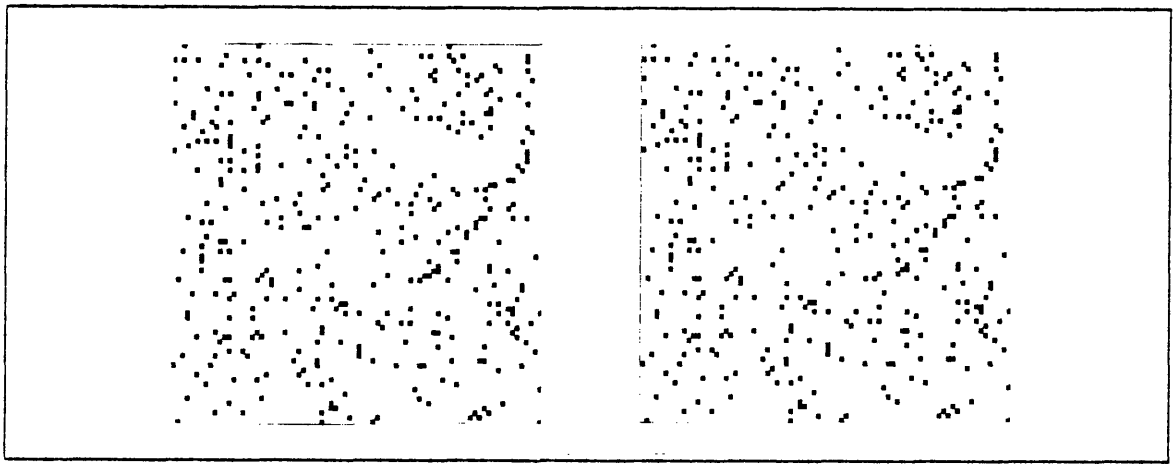


Figure 6.4. A Sparse Random Dot Stereogram. Althought explicit disparity is only available at the edges of the dots, a vivid impression of two planar surfaces is obtained.

requires and constructs a more specific, finer resolution representation of the surfaces.

6.3.1 Psychophysics

Although the question of one-dimensional interpolation of subjective contours has been studied in the psychological literature, the question of two-dimensional interpolation of subjective surfaces has had less attention. It would seem that this theoretical formulation of the problem may be precise enough to be tested, and future work in investigating the psychophysical aspects of subjective surfaces will be useful in substantiating the theory.

Some initial psychophysical evidence is available. Figure 6.4 illustrates a sparse, 5% density random dot stereogram. The impression obtained upon viewing this stereogram is one of two distinct planes, sharply separated in depth. Yet, by any theory, explicit disparity information is available only along the edges of the dots, which cover a very small portion of the total area. Hence, it would seem on the basis of this stereogram that some type of filling-in, or interpolation of surface information, is taking place in the visual system. Particularly noteworthy in this case are the strong subjective contours between the two planes.

The role of the filling-in process can be investigated in more detail by the method illustrated in Figure 6.5. The object in this stereogram is a half cylinder, lying below a reference plane. The density of the dots which lie on the cylinder fades gradually from 10% to zero in the center of the image. In this manner, a gap is created in the cylinder, without any sharp changes in dot density (and

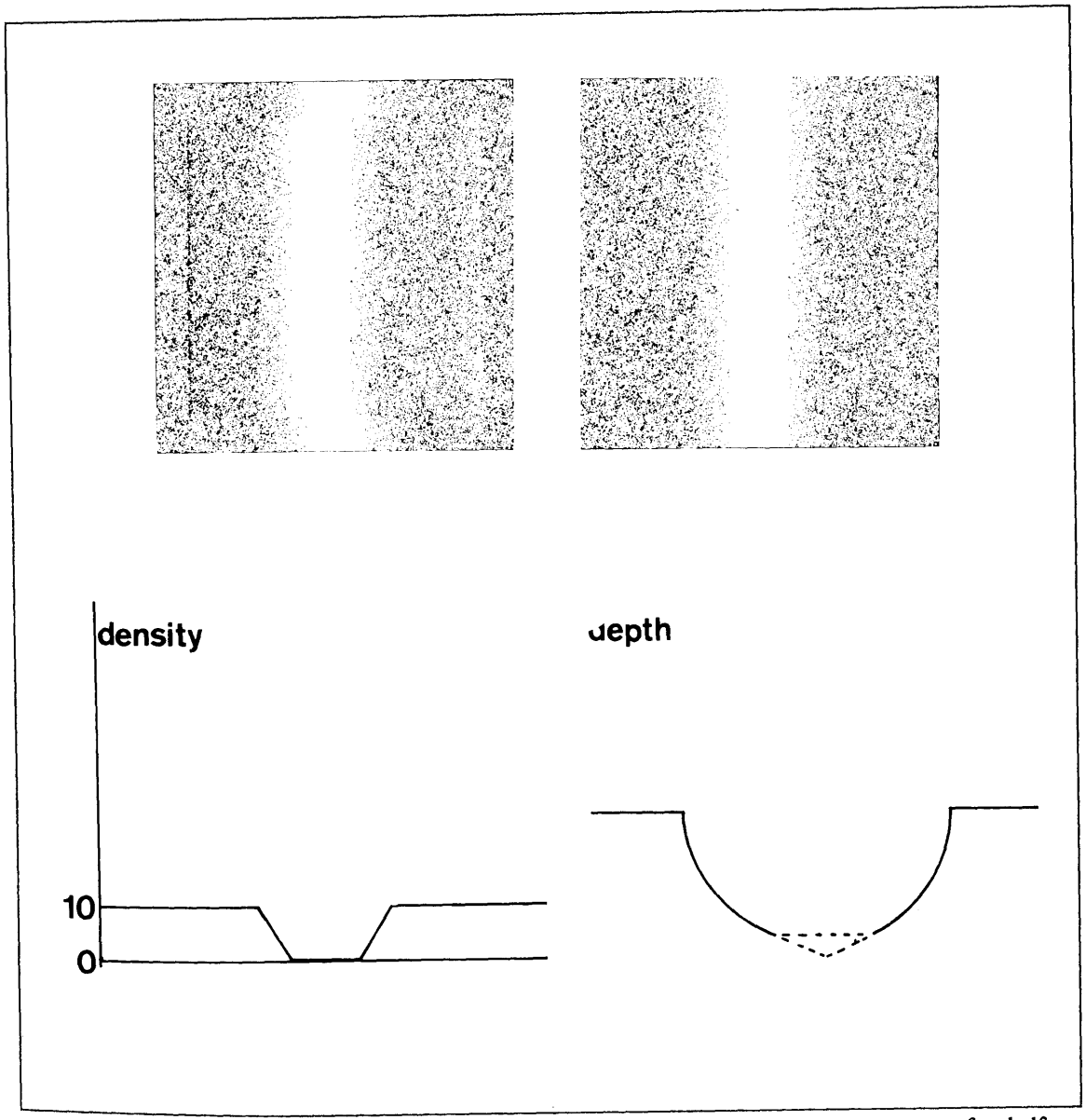


Figure 6.5. The Half Cylinder. The top pair of images is a random dot stereogram of a half cylinder. The bottom left figure shows the dot density used to create the stereogram. The bottom right figure shows the depth values of a cross-section of the stereogram. The dotted lines indicate the perception which would be expected for linear interpolation and tangential interpolation. The actual perception is of a smooth cylinder.

thus without any suggestion of an object occluding the cylinder). The perception of this stereogram obtained by most viewers is of a complete cylinder, thereby indicating again that some type of filling-in is taking place. More interestingly, one can carefully test the attributes of the filling-in process by inserting a stereoscopic probe into the center of the blank region. Three important observations are

made from this testing, and each relates to a possible form of the filling-in process. The first concerns the case in which the interpolation of the surface is strictly linear between the known disparity values. For the cylinder, this would correspond to a planar truncation of the cylinder, between the lowest visible points, also indicated in the figure. Not only is this inconsistent with the perception of the stereogram, but when a probe is inserted in the position corresponding to this situation, it is perceived as lying above the surface of the cylinder. The second case is one in which the interpolation of the surface is a linear extension of the tangents of the surface through the bottom-most visible points. For the cylinder, this would correspond to the planar extension of the cylinder, terminating in a sharp intersection of the planes, as indicated in the figure. Again, this is inconsistent with the perception of the stereogram, and when a probe is inserted at this position, it is perceived to lie below the surface of a now transparent cylinder. Thus, an upper and a lower bound on where the surface appears to lie has been established. When the probe is placed in positions intermediate to the previous two, the situation is less clear. For most such positions, the probe appears roughly to lie on the surface of the cylinder, with the strongest perception for the case of the probe lying exactly on the cylinder.

The conclusion I draw from this experiment is that some type of interpolation, or filling-in, is taking place, but that it seems to be rather imprecise, although it does seem roughly to preserve the curvature of the object involved. Note that in the case of the linear extension of the disparities and in the case of the tangential extension of the disparities, both of which were seen to be incorrect, such extensions introduce sharp discontinuities in the surface orientation of the surface. This is consistent with the theorems relating surface shape and zero-crossings. In particular, since there are no explicit zero-crossings within the gap in the cylinder, no surface which would introduce surface inflections is acceptable, because of the probability of that surface requiring additional zero-crossings.

CHAPTER 7

CONSTRAINED OPTIMIZATION

In the previous section, a computational problem was derived. It stated that to each possible surface which interpolated the known data, a real number could be assigned via a functional Θ. This functional measured the consistency of the surface with respect to the zero-crossings of the convolved image — the smaller the value, the less inconsistent the surface. In this manner, the least inconsistent surface is that which minimizes this functional.

A set of mathematical conditions were also outlined. These were imposed on the functional, in order to ensure that the problem was well defined. In particular, the functional was required to be a particular type of semi-norm and the space of all possible functions was required to be a Hilbert space. In this chapter, the problem of creating an explicit algorithm to solve this computational problem is considered. There are two distinct classes of methods which could be used - global and local. Before outlining possible methods, the type of method best suited to this problem will be considered.

7.1 The Role of Algorithmic Criteria

As mentioned in the first chapter, an essential problem for any computational theory is to determine the implicit assumptions used by the visual system to perform the computation. Such assumptions are valid assumptions about the environment which are explicitly incorporated into the computation. Ullman's rigidity assumption in visual motion perception (1979), Marr and Hildreth's

condition of linear variation and spatial coincidence assumption (1979), and Marr and Poggio's assumptions of uniqueness and continuity (1979) are three examples. Such assumptions help to constrain the computational theory. As such they may be considered as computational constraints on the problem.

There is a second set of criteria which may be applied to any theory and, more importantly, to any algorithm for a theory. These deal with the requirement of biological feasibility, and are important if one is to describe a model of the human system. They will be termed *algorithmic criteria.* Ullman (1979b) has listed a number of such criteria which should apply to any biologically feasible algorithm. A similar set is briefly sketched here.

Rapidity

In most circumstances, both for humans and for machines, a visual system serves as a major input device to a problem solving system that must interact with its environment. This makes it critical that it perform its computations as rapidly as possible.

Parallelism

The need to process large amounts of input data in short amounts of time implies the use of computations which can be implemented in a parallel manner, using a large number of interconnected processors.

Local-Support

If the number of processors involved in the computation is large, it becomes infeasible to connect each one to all of the others. Rather, there should only be local connections between the processors. Here, "local" means not only that the number of connections be small, but also that since the information being processed has as an underlying coordinate system, a two-dimensional plane, the connections should also be local in a spatial sense. If the support of a function, defined on a two-dimensional grid, is the set of points on the grid which contribute in a non-trivial manner to the computation of the function, then our requirement is that the processors implementing our computation must have local support.

Simplicity

The next requirement concerns the complexity of the individual processors. If the computation can be made local only at the expense of requiring each processor to compute some complex

function requiring a long time interval to complete, then this is not desirable. Although the method of implementation will affect the amount of time needed to perform a computation, in general, we shall favour processes in which each individual processor performs a simple computation. Note that the global computation performed by the parallel network may be complex; it is only necessary that the individual processors be simple.

One final consideration, though not as critical as the first four, concerns the *uniformity* of the processors. If it is possible, an algorithm which utilizes parallel networks of identical processors will be favored over other algorithms. However, such a requirement is not crucial.

Although the original motivation for such requirements on the algorithm arises from consideration of the human visual system, and requirements of biological feasability, they could apply equally well to other types of image processing systems. As such, they are taken as general criteria for the computations to be investigated, regardless of whether the algorithm designed serves as a model of the human system. So in designing algorithms to solve a particular visual process, the first step is to seek a method that solves the problem. Having done so, one can then consider its applicability in light of the criteria outlined above, and possible modifications to the algorithm in order to satisfy those criteria.

7.2 Methods of Solution

In this section, several possible methods of solution will be outlined, before considering a particular algorithm appropriate to the computational problem. The review article of Schumaker (1976) contains more information about possible methods of solving the interpolation and approximation problems.

Many of the methods studied in the literature are global methods, where the value of the surface at any point depends all the known values. The discussion above implies that local methods, where the value of the surface at any point depends only on the nearby known points, are of particular interest. In general, global interpolation methods can be made local by partitioning the domain D into subdomains and patching together the interpolation functions for each subdomain. That is, to define the surface function f, one defines the surface in local patches, f_i over the subdomain D_i.

Note also that many of the methods for interpolation which are to be found in the literature require regular grids of known data points. Clearly, the Marr-Poggio algorithm, as well as other visual processes, such as Ullman's structure from motion, supplies data at scattered points. Hence, the

concentration will be on methods applicable to such data. However, note that there is the possibility of a two stage process — a first stage to construct a surface g based on the scattered data, and then a second stage using the regular data of g for the construction of a smoother surface f.

7.2.1 Partitions

The most common subdomains for defining surface patches are rectangles (for gridded data) and triangles (for scattered data). When dealing with scattered data, one must consider methods for triangulating the domain into subdomains. The major problem with irregular or scattered data is that the triangularization of the field is a difficult problem. Moreover, the triangularization is not unique and in many cases (such as this) it is difficult to avoid long, thin triangles, which are poorly suited for interpolation. Some algorithms for triangularization are given in Cavendish (1974) and Lawson (1972).

7.2.2 Piecewise Linear Interpolation

The simplest method for defining a local interpolating surface is to construct each surface patch $f_i(x, y)$ to be of the form $a_1 + a_2x + a_3y$. The data at the corners of the triangle are sufficient to determine the coefficients for that piece of f. This results in a piecewise linear surface which is globally continuous. The major problem with this method is that no higher order smoothness will be attained. Thus, along each side of a triangle, the resulting surface will be discontinuous in first and all higher order derivatives. This is clearly unacceptable for this situation.

Examples of the use of this method for data fitting can be found in Lawson (1972) and Whitten and Koelling (1974).

7.2.3 Polynomial Interpolation

The theory of finite dimensional interpolation is well known (for example, see Davis, 1963.) In general, it works as follows:

Let $\{\phi_i\}_1^N$ be a set of N functions on a domain D. Then

$$f(x, y) = \sum_{j=1}^{N} a_j \phi_j(x, y)$$

will satisfy the constraints $f(x_i, y_i) = F_i, i = 1, \ldots, N$ if and only if $\{a_j\}_1^N$ is a solution of the linear

system

$$\sum_{j=1}^{N} a_j \phi_j(x_i, y_i) = F_i \qquad i = 1, \ldots, N,$$

which has a unique solution if and only if it is nonsingular.

Standard choices for the ϕ_j are polynomials in x, y. Problems with such a choice include guaranteeing that the system is nonsingular, or even when it is, guaranteeing that it is not ill-conditioned, and the fact that polynomials tend to exhibit a considerable oscillatory behaviour, yielding unacceptable undulating surfaces, likely to introduce inflections in the image intensities inconsistent with the zero-crossings.

In general, polynomial interpolation has not been applied to scattered data. Some treatment, however, can be found, for example in Ciarlet and Raviart (1971, 1972a, 1972b, 1972c), Guenther and Roetman (1970), Kunz (1957), Prenter (1975), Steffensen (1927), Thacher (1960), Thacher and Milne (1960), and Whaples (1958).

Polynomial interpolation is somewhat simplified, in the case of known data on a grid. By this, the following is meant. Let H be the rectangle $[a, b] \times [c, d]$ and let

$$a = x_0 < x_1 < \ldots < x_{k+1} = b$$

$$c = y_0 < y_1 < \ldots < y_{l+1} = d$$

Let F be defined on H and let the known values be

$$F_{ij} = F(x_i, y_j) \quad i = 0, \ldots, k+1 \quad j = 0, \ldots, l+1.$$

As noted, however, gridded data is only applicable in a two-stage process. Gridded polynomial fitting has been well studied in the literature, see for example, Prenter (1975) or Steffensen (1927).

If we do not require that the surface exactly interpolate the known data, but only approximate it, then the polynomial method may be extended to that of polynomial least squares fitting.

The general theory of discrete least-squares fitting is very well known. To review, suppose that $\{\phi_j\}_1^n$ are n given functions on D. Let

$$\Phi(a) = \sum_{i=1}^{N} \left(\sum_{j=1}^{n} a_j \phi_j(x_i, y_i) - F_i \right)^2$$

where $a = (a_1, \ldots, a_n)^T$ is an arbitrary vector in $\Re^n$. The problem is to find an a' such that

$$\Phi(a') = \min_a \Phi(a).$$

Then

$$f(x, y) = \sum_{j=1}^{n} a'_j \phi_j(x, y)$$

is the discrete least-squares approximation of $\{F_i\}_1^N$. Usually, $n \ll N$.

There are several methods for solving this problem. For example, if the $\{\phi_j\}_1^n$ are orthonormal relative to the inner product

$$(\phi, \psi) = \sum_{i=1}^{N} \phi(x_i, y_i)\psi(x_i, y_i)$$

then the solution is given by

$$f(x, y) = \sum_{j=1}^{N} F_j \phi_j(x, y).$$

One can also solve this problem by using normal equations or by using general matrix methods on $Aa = F$, where

$$F = (F_1, \ldots, F_N)^T$$
$$A = (\phi_j(x_i, y_i))$$

Polynomial discrete least-squares fitting has been widely studied for scattered data. For example, see Cadwell and Williams (1962), Crain and Bhattacharyya (1967), Whitten (1970, 1972).

Polynomial least squares methods can be considered as multi-dimensional regression (see Effroymson, 1960).

The method of weighted least squares can also be used. In this case, rather than using the function Φ, one uses

$$\Phi(a) = \sum_{i=1}^{N} w_i \left(\sum_{j=1}^{n} a_j \phi_j(x_i, y_i) - F_i \right)^2$$

where $a = (a_1, \ldots, a_n)^T$ is an arbitrary vector in $\Re^n$. The problem is again to find an a' such that

$$\Phi(a') = \min_a \Phi(a).$$

Then

$$f(x, y) = \sum_{j=1}^{n} a'_j \phi_j(x, y)$$

is the discrete least-squares approximation of $\{F_i\}_1^N$, (Pelto, Elkins and Boyd, 1968).

Since discrete least-squares fitting can be carried out with any finite set of functions, many authors have used tensor products of splines (Hayes and Halliday 1974, Whitten 1971).

Also, rather than using least-squares fitting, one can use ℓ_1 or ℓ_∞ approximation. In other words, given $\{\phi_j\}_1^n$ defined on D, the vector a' is sought such that

$$\Phi(a) = \sum_{i=1}^{N} \left| \sum_{j-1}^{n} a_j \phi_j(x_i, y_i) - F_i \right|$$

is minimized or such that

$$\Phi(a) = \max_{1 \le i \le N} \left| \sum_{j=i}^{n} a_j \phi_j(x_i, y_i) - F_i \right|$$

is minimized. Both of these approximations can be solved as linear programming problems (Rabinowitz 1968, 1970, Rosen 1971).

All of the above discussion applies to the general global case. In the local case, if functions $\{\phi_j(x, y)\}_1^N$ are constructed with the property that

$$\phi_j(x_i, y_i) = \delta_{ij} \qquad i, j = 1, \ldots, N,$$

then these functions can be constructed as pyramids in such a way that the function ϕ_j has support only on the triangles surrounding the data point (x_i, y_i). These functions are usually referred to as Lagrangian functions.

The Lagrangian approach to local interpolation is similar to the finite element method, which is concerned with the solution of an operator equation in the form of a linear combination of a set of functions (called elements) with the above property. The functions need not be restricted to polynomials, but may be rational functions or more complicated functions. In fact, it is possible to construct Lagrangian functions (or elements) with small support but higher global smoothness.

There are a great many papers in the finite-element literature concerned with defining convenient smooth elements. Included are the books of Aziz (1972), de Boor (1975), Strang and Fix (1973), Whiteman (1973), and the papers of Barnhill, Birkhoff and Gordon (1973), Barnhill and Gregory (1972, 1975), Barnhill and Mansfield (1974), Birkhoff and Mansfield (1974), Bramble and Zlamal (1970), Goel (1968), Hall (1969), Mitchell (1973), Mitchell and Phillips (1972), Nicolaidis (1972, 1973), Zenisek (1970), Zienkiewicz (1970), Zlamal (1968, 1970, 1974).

It should be noted that the construction of elements with higher-order smoothness becomes increasingly difficult. For example, Mansfield (1976) notes that to get an element with support on a triangle and achieve global continuity, it is necessary to use polynomials of degree at least 5. Akima (1978) gives an algorithm for accomplishing this.

7.2.4 Shepard's Method

Shepard (1968) has developed a method specifically intended for the case of scattered data. In brief, the method consists of the following.

Let ρ be a metric in the plane. Given a point (x, y), let

$$r_i = \rho((x, y), (x_i, y_i)) \quad i = 1, \ldots, N.$$

Let $0 < \mu < \infty$. Then Shepard's interpolation formula is given by:

$$f(x, y) = \begin{cases} \sum_{i=1}^{N} \frac{F_i}{r_i \mu} \Big/ \sum_{i=1}^{N} \frac{1}{r_i \mu}, & \text{if all} \quad r_i \neq 0, \\ F_i, & \text{if some} \quad r_i = 0. \end{cases}$$

To determine the value of the surface at a point, this method essentially weights all the data points according to their distance from that point.

The problem with Shepard's method is that it is very sensitive to the value of μ. For $0 < \mu \leq 1$, the surface f has cusps at the data points. For $1 < \mu$, f has flat spots at the data points. Thus, to avoid cusps, one needs $1 < \mu$. But if μ is relatively large, then the surface is very flat near the data points, and very steep in between. In this situation this is undesirable. Note that the usual value for μ is 2 (see Poeppelmeir, 1975 and Shepard, 1968).

Two other problems with Shepard's method are first, if N is large, a great deal of calculation is required to compute the surface, and second, the weights are based on distance, but not on direction, although this can be corrected (Shepard, 1968).

Of course, this method is a global one. Shepard modifies the method to make it local in the following manner.

Fix $0 < R$, and define

$$\psi(r) - \begin{cases} \frac{1}{r}, & 0 < r \leq \frac{R}{3}, \\ \frac{27}{4R}\left(\frac{r}{R} - 1\right)^2, & \frac{R}{3} < r \leq R, \\ 0, & R < r. \end{cases}$$

This function is continuously differentiable, and vanishes identically for $f > R$. Now define

$$f(x, y) = \begin{cases} \sum_{i=1}^{N} F_i[\psi(r_i)]^\mu \Big/ \sum_{i=1}^{N} [\psi(r_i)]^\mu, & \text{if all} \quad r_i \neq 0, \\ F_i, & \text{if some} \quad r_i = 0. \end{cases}$$

This function interpolates the values F_i at the data points, and elsewhere the values are weighted averages of the data values which lie at points within a distance R of the point.

7.2.5 Quasi-interpolants

One way of creating local methods of interpolation and approximation is to apply global methods of small partitions of the domain. There are also direct local methods which construct an approximate surface without solving a system of equations.

Let $\mathcal{F}$ be a linear space of functions on D, and $\{\lambda_i\}_1^N$ be a set of linear functionals on $\mathcal{F}$. Let $\{\phi_i\}_1^N$ be a prescribed set of functions on D. Then we are interested in approximation schemes of the form:

$$QF(x, y) = \sum_{i=1}^{N} \lambda_i F \phi_i(x, y).$$

This is a surface fitting problem, where the data are given by $F_i = \lambda_i F$ for $i = 1, \ldots, N$. If the ϕ_i have support on small subsets of D, and if each λ_i has support on the same set, then the above computation is local. For example, if we let λ_i be point evaluation at (x_i, y_i) and $\phi_i(x, y)$ be a function with support in a neighbourhood of (x_i, y_i) then the approximation formula is given by

$$QF(x, y) = \sum_{i=1}^{N} F_i \phi_i(x, y).$$

This is similar to the Lagrange form of interpolation, but unless the ϕ_i satisfy $\phi_j(x_i, y_i) = \delta_{ij}$, QF will not be an interpolant. As a consequence, they are called quasi-interpolants. The main problem

with this method involves a careful choice of $\{\phi_i\}_1^N$ in order to ensure appropriate accuracy and smoothness. Usually this excludes cases of scattered data. For an example of this method, see Lyche and Schumaker (1975).

Other examples of local approximation schemes include Babuska (1970), de Boor and Fix (1973), Fix and Strang (1969), Fredrickson (1971).

7.2.6 Spline Interpolation

Let X be a linear space of "smooth" functions defined on the domain D. Let

$$U = \{f \in X : f(x_i, y_i) = F_i, \quad i = 1, \ldots, N\}$$

so that U is the set of smooth functions which interpolate the data. Now let Θ be a functional on X which measures the smoothness of a function $f \in X$ — the smaller $\Theta(f)$ is, the smoother f is.

Consider the minimization problem of finding $s \in U$ such that $\Theta(s) = \inf_{u \in U} \Theta(u)$, assuming that such an s exists. Then s will be the "smoothest" interpolant, and in view of the similarity with classical spline approximation, s is called a spline function interpolating F. For a general theory of spline interpolation, see Laurent (1972). The general method used in solving such minimization problems globally is to define a functional Θ, frequently as a pseudo-norm, and then to construct a reproducing kernel K on $D \times D$, so that the surface is explicitly given by the form

$$s(x, y) = \sum_{i=1}^{N} a_i K((x, y); (x_i, y_i)) + \sum_{i=1}^{d} b_i p_i(x, y)$$

where the p_i are a basis for the null space of the norm. The coefficients $\{a_i\}$ and $\{b_i\}$ can be determined by solving a system of linear equations. Examples of this method include Atteia (1966a, 1966b, 1970), Duchon (1975, 1976), Mansfield (1971, 1972a, 1972b, 1974), Schaback (1973, 1974), Thomann (1970a, 1970b).

Spline approximation over gridded data has been extensively studied, with a wide body of literature. Here, the general method is to use the product of B-splines over each rectangular surface patch, together with a set of boundary conditions. Most common is bicubic spline interpolation. Examples include Ahlberg, Nilson and Walsh (1965, 1967), Birkhoff and de Boor (1965), Birkhoff and Garabedian (1960), de Boor (1962), Koelling and Whitten (1973), Spath (1969) and Theilheimer and Starkweather (19619).

Many generalizations of the spline problem have been investigated. Some of the methods include Arthur (1974, 1975), Birkhoff, Schultz and Varga (1968), de Boor (1973), Delvos (1975a, 1975b), Delvos and Schempp (1975), Delvos and Schlosser (1974), Fisher and Jerome (1974, 1975), Haussman (1974), Haussman and Munch (9173), Manteanu (1973a, 1973b), Nielson (1970, 1973), Ritter (1969, 1970), Sard (1973, 1974), Schultz (1969a, 1969b), Spath (1971) and Zavialov (1973, 1974a, 1974b).

If we do not require exact interpolation, but only surface approximation, spline methods are also applicable.

Let X be a linear space of "smooth" functions, and Θ be a functional on X that measures the smoothness of an element in X. Let E be a functional on X which measures how well a function fits the data. Then the problem is to find $s \in X$ such that

$$\rho(s) = \inf_{u \in X} \rho(u)$$

where

$$\rho(f) = \Theta(f) + E(f),$$

if such an s exists. As in the case of spline interpolation, the general theory of spline approximation may be found in Laurent (1972). The methods used to find solutions to this problem globally are similar to those used to solve the global spline interpolation problem — constructing a reproducing kernel and solving a system of linear equations to determine the coefficients of the resulting equation for the surface. Examples of this method include Duchon (1975, 1976) and Pivovarova and Puknacheva (1975).

Local spline interpolation has been widely used, (Birkhoff and de Boor, 1965; Birkhoff and Garabedian, 1960; de Boor, 1962; Ahlberg, Nilson and Walsh, 1967; Whitten and Koelling, 1974; Arthur, 1974, 1975; Birkhoff, Schultz and Varga, 1968; Delvos, 1975; Delvos and Kosters, 1975; Schultz, 1969a, 1969b; and many more). In general, given a partition of the domain into regions (defined by a set of knots), a spline function is a collection of functions, each defined on a different portion of the partition, which are pieced together to form the surface and which have continuous derivatives up to some order r. A standard example of a one-dimensional spline is the cubic spline, in which the individual functions are cubic polynomials, and the functions have identical zeroth, first and second order derivatives at the knots. In two dimensions, one can use the product of cubic splines,

known as bicubic splines. There are, of course, many other possible functions which can be used besides cubics.

There are a number of problems associated with using splines. The first case concerns the use of interpolating splines, in which the knots are chosen to be the known data points. In order to compute such splines, it is necessary to know the directions of the tangents of the surface at the extreme points. This may not always be possible. As well, if the data are noisy, the interpolating splines usually show strong oscillations, which are not acceptable.

One way around this problem is to chose a partition in which the knots are variable, rather than fixed at the known points. Ideally, one would like to leave them as free parameters to be chosen during an optimization problem. However, this turns out to be a very unpleasant mathematical and computational problem. Whereas the curve fitting problem is linear for fixed knots, the case of variable knots is nonlinear. There are local descent methods for finding the curve iteratively. However, the computational effort required for splines of higher order than linear is prohibitive. (For a further discussion of the use of spline approximations and its problems, see Pavlidis, 1977.)

One final comment about the use of splines concerns the idea of spline blending. These methods are used for constructing surfaces in the case of gridded data. They interpolate not only function values at isolated points, but also along the grid lines themselves. Of course, in our situation, the fact that we are given scattered data makes such methods of little use. Examples of this method include Coons (1967), Barnhill and Reisenfeld (1974) (including many references), Earnshaw and Yuille (1971), Forrest (1972a, 1972b), Ferguson (1964) and Hosaka (1969).

7.3 Mathematical Programming

Of the various possible local methods of solution outlined, most of them seem unsuited to our particular problem, either because of requirements on the form of the input (gridded data) or because the methods do not satisfy the criteria of biological feasibility. There is one other method of solution, the method of nonlinear programming. This method is chosen to be applied to our problem, in part because the problem is viewed as one of constrained optimization. Ullman (1979b) has shown that many problems of relaxation and constrained optimization can be solved by local processes of a type suitable to meet the algorithmic criteria. Indeed, a method very similar to that outlined here was used by Ullman in solving the motion correspondence problem (Ullman, 1979a). However, the difference

between the computational theory and the algorithm for solving it is again stressed. Although a particular algorithm for solving the computational theory is illustrated in this chapter, there may be many other possible methods of solution. Some will be better suited to certain purposes than others. This does not reflect, however, on the relevance of the computational theory, which holds true independent of the particular algorithm used to implement it.

Briefly, the relevant aspects of nonlinear programming are outlined (for a more complete developement, see for example Luenberger, 1973), and then applied to the interpolation of surfaces. Two results will be relied upon. the Kuhn-Tucker theorem states that under certain conditions, finding a constrained optimum is equivalent to finding the saddle-point of an associated Lagrangian. The Arrow-Hurwicz system of equations then shows that finding the saddle-pont of a Lagrangian can be accomplished by a system of equations which meets our set of algorithmic constraints.

7.3.1 Nonlinear Programming

The general nonlinear programming problem is of the following form. Let x be an n-vector, $\mathrm{x} = \{x_1, x_2, \ldots, x_n\}$, g(x) be an p-vector, $\mathrm{g(x)} = \{g_1(\mathrm{x}), \ldots, g_p(\mathrm{x})\}$, and h be an m-vector, $\mathrm{h(x)} = \{h_1(\mathrm{x}), \ldots, h_m(\mathrm{x})\}$.. Let $f(\mathrm{x})$ be any function. The problem is then:

$$\begin{array}{ll} \text{minimize} & f(\mathrm{x}) \\ \text{subject to} & \mathrm{h(x)} = 0, \qquad \mathrm{g(x)} \leq 0. \end{array}$$

The problem of finding the vector x′ which maximizes $f(\mathrm{x})$ subject to these constraints h, g is known as a *problem of constrained optimization.*

Some definitions will aid the discussion that follows. If f is a function with continuous first partial derivatives, the *gradient* of f is the vector

$$\nabla f(\mathrm{x}) = \left[\frac{\partial f(\mathrm{x})}{\partial x_1}, \ldots, \frac{\partial f(\mathrm{x})}{\partial x_n}\right].$$

If f has continuous second partial derivatives, the *Hessian* of f at x is the $n \times n$ matrix denoted $\nabla^2 f(\mathrm{x})$ or F(x) and defined as

$$\mathrm{F(x)} = \left[\frac{\partial^2 f(\mathrm{x})}{\partial x_i \partial x_j}\right].$$

For a vector-valued function, $\mathrm{f} = (f_1, \ldots, f_m)$ the first derivative is defined as the $m \times n$ matrix

$$\nabla \mathrm{f(x)} = \left[\frac{\partial f_i(\mathrm{x})}{\partial x_j}\right].$$

If f has continuous second partial derivatives, it is possible to define the m Hessians $F_1(x), \ldots, F_m(x)$ corresponding to the m component functions.

Also, given any $\lambda = [\lambda_1, \ldots, \lambda_m]$ the real-valued function λf has gradient equal to $\lambda \nabla f(x)$ and Hessian, denoted $\lambda F(x)$, equal to

$$\lambda F(x) = \sum_{i=1}^{m} \lambda_i F_i(x).$$

A point x that satisfies all the constraints is said to be *feasible.* If additionally $g(x) < 0$, it is said to be *strictly feasible.* An inequality constraint $g_i(x) \leq 0$ is said to be *active* at a feasible point x if $g_i(x) = 0$.

Let x' be a point satisfying the constraints

$$h(x') = 0, \qquad g(x') \leq 0$$

and let J be the set of indices j for which $g_j(x') = 0$. Then x' is said to be a *regular point* of the constraints if the gradient vectors $\nabla h_i(x')$, $\nabla g_j(x')$, $1 \leq i \leq m, j \in J$ are linearly independent.

Consider first the first-order necessary conditions for a point to be a local minimum point subject to the constraints. (For a proof, and further development see Kuhn and Tucker, 1951; Arrow, Hurwicz and Uzawa, 1958; or Luenberger, 1973.)

Theorem (Kuhn-Tucker Conditions): Let x' be a relative minimum point for the problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \qquad g(x) \leq 0 \end{array}$$

and suppose x' is a regular point for the constraints. Then there is a vector $\lambda \in E_m$ and a vector $\mu \in E_p$ with $\mu \geq 0$ such that

$$\nabla f(x') + \lambda \nabla h(x') + \mu \nabla g(x') = 0$$
$$\mu g(x') = 0,$$

where $\lambda \nabla h(x')$ is the scalar product of the two vectors. ∎

The vectors λ and μ are usually called *Lagrangian multipliers.* It is convenient to introduce the *Lagrangian* associated with the constrained problem, defined as

$$l(x, \lambda, \mu) = f(x) + \lambda h(x) + \mu g(x).$$

The necessary conditions can then be expressed in the form

$$\nabla_x l(x, \lambda, \mu) = 0$$
$$\nabla_\lambda l(x, \lambda, \mu) = 0$$
$$\nabla_\mu l(x, \lambda, \mu) = 0$$

the last two of these being simply a restatement of the constraints.

One can also develop necessary and sufficient second order conditions on the problem.

Second-Order Necessary Conditions. Suppose the functions f, g, h have continuous second partial derivatives and that x' is a regular point of the constraints. If x' is a relative minimum point for the problem, then there is a $\lambda \in E_m, \mu \in E_p, \mu \geq 0$ such that

$$\nabla f(x') + \lambda \nabla h(x') + \mu \nabla g(x') = 0$$
$$\mu g(x') = 0,$$

hold and such that

$$L(x') = F(x') + \lambda H(x') + \mu G(x')$$

is positive semidefinite on the tangent subspace of the active constraints at x'. ∎

Second-Order Sufficiency Conditions. Let f, g, h have continuous second partial derivatives. Sufficient conditions that a point x' satisfying the constraints be a strict relative minimum point of the problem is that there exist $\lambda \in E_m, \mu \in E_p$ such that

$$\mu \geq 0$$
$$\mu g(x') = 0$$
$$\nabla f(x') + \lambda \nabla h(x') + \mu \nabla g(x') = 0$$

and the Hessian matrix

$$L(x') = F(x') + \lambda H(x') + \mu G(x')$$

is positive definite on the subspace

$$M' = \{y : \nabla h(x')y = 0, \nabla g_j(x')y = 0 \quad \text{for all} \quad j \in J\}$$

where

$$J = \{j : g_j(x') = 0, \mu_j > 0\}.$$

∎

7.3.2 The Arrow-Hurwicz Gradient Method

The main point about the Kuhn-Tucker conditions is that they relate the solution of a constrained optimization problem to the position of a saddle-point of the associated Lagrangian. As a result, it may be possible to consider methods for computing this optimum solution.

There are many methods for finding solutions to constrained optimization problems. In order to find one that meets the criteria of local-support, one may take advantage of the particular form of the function to be optimized. In particular, let

$$l(\mathrm{x}, \mu) = f(\mathrm{x}) + \mu \mathrm{g}(\mathrm{x})$$

be the Lagrangian associated with a constrained optimization problem involving only inequality constraints. Consider the system of difference equations defined by

$$\mathrm{x}(t+1) = \max\left[0, \mathrm{x}(t) - \rho \nabla_{\mathrm{x}} l(\mathrm{x}(t), \mu(t))\right]$$
$$\mu(t+1) = \max\left[0, \mathrm{y}(t) + \rho \nabla_{\mu} l(\mathrm{x}(t), \mu(t))\right]$$

with an initial position $\{\mathrm{x}(0), \mu(0)\}$ such that

$$\mathrm{x}(0) \geq 0, \quad \mathrm{y}(0) \geq 0,$$

where ρ is a given positive number, and $\nabla_{\mathrm{x}} l$, $\nabla_{\mu} l$ are the partial derivatives of l with respect to x and μ, respectively:

$$\nabla_{\mathrm{x}} l(\mathrm{x}, \mu) = \nabla_{\mathrm{x}} f(\mathrm{x}) + \mu \nabla_{\mathrm{x}}(\mathrm{x})$$
$$\nabla_{\mu} l(\mathrm{x}, \mu) = \mathrm{g}(\mathrm{x})$$

Such a system will be defined to be **-stable* if the following condition is satisfied:

For any initial position $\{\mathrm{x}(0), \mu(0)\} \geq 0$ *and any positive number* $\epsilon > 0$, *there exists a positive number* $\rho_0 > 0$ *such that, for the solution* $\{\mathrm{x}(t), \mu(t)\}$ *of this system with* $\rho \leq \rho_0$, *there is an integer* t_0 *with the properties*

$$V[\mathrm{x}(t+1), \mu(t+1)] \leq V[\mathrm{x}(t), \mu(t)], \quad 0 \leq t < t_0,$$

and

$$V[\mathrm{x}(t), \mu(t)] \leq \epsilon \quad t \geq t_0,$$

where

$$V(\mathrm{x}, \mu) = \min\{\, |\mathrm{x} - \mathrm{x}'|^2 + |\mu - \mu'|^2 \,\}$$

The minimization is over the set of all μ' *such that* (x', μ') *is a saddle-point of* $l(\mathrm{x}, \mu)$ *for some* x'.

A function f on a convex set Ω is defined to be *convex* if, for every $x_1, x_2 \in \Omega$ and every $\alpha, 0 \leq \alpha \leq 1$, there holds

$$f(\alpha x_1 + (1-\alpha)x_2 \leq \alpha f(x_1 + (1-\alpha)f(x_2).$$

If, for every $0 < \alpha < 1$ and $x_1 \neq x_2$, there holds

$$f(\alpha x_1 + (1-\alpha)x_2 < \alpha f(x_1 + (1-\alpha)f(x_2),$$

then f is said to be *strictly convex*. One can then prove the following theorem, which describes a method for computing the optimum vector, based on the Kuhn-Tucker conditions.

Theorem (Arrow-Hurwicz): Suppose that

(a) $f(x), g_1(x), \ldots, g_p(x)$ **are convex functions in** $x \geq 0$ **and have continuous partial derivatives,**

(b) **there is a feasible vector** x^0,

(c) f **is strictly convex in** x.

Then the Arrow-Hurwicz system of difference equations for the problem of minimizing $f(x)$ **subject to the conditions** $g(x) \leq 0$, **is *-stable. (For a proof, see Arrow and Hurwicz, 1956; Arrow, Hurwicz and Uzawa 1958).** ∎

This theorem, together with the Kuhn-Tucker theorem, indicates that to find an optimum vector for the constrained optimization problem, one need only find a saddle-point of the Lagrangian, via the Arrow-Hurwicz system of equations. Further, Uzawa states that condition (c) implies that the optimum solution is uniquely determined. Thus, by using these two theorems, and the statement of the interpolation problem, a basis for an algorithm is formed. However, I stress again the fact that this method of implementation is independent of the particular form of the semi-norm Θ, and gives a general method for implementing any such semi-norm.

7.4 The One-Dimensional Case

To illustrate the method, we return to the example of interpolating cylindrical surfaces. The problem is:

$$\text{Minimize} \left\{ \int f_{xx}^2 \, dx \right\}^{\frac{1}{2}}$$

subject to the constraints that the surface pass through a given set of points.

A number of observations about this particular problem are important:

(1) The computation takes place on a uniform grid, rather than continuously. Thus, the variables of the computation are the surface values at the grid points. These variables are defined as $f_i = f(x_i)$. As a consequence, the integral in the semi-norm becomes a finite sum. In addition, the derivative operator must be converted to a finite difference.

$$\begin{aligned} f_{xx}(x_i) &= f(x_{i-1}) - 2f(x_i) + f(x_{i+1}) \\ &= f_{i-1} - 2f_i + f_{i+1}. \end{aligned}$$

(2) The minimization of

$$\left\{ \int f_{xx}^2 \, dx \right\}^{\frac{1}{2}}$$

is equivalent to the minimization of

$$\int f_{xx}^2 \, dx$$

so that the square root may be removed.

These factors combine to redefine the problem as:

$$\text{Minimize } \sum_{i=1}^{n-1} (f_{i-1} - 2f_i + f_{i+1})^2 .$$

The constraints on the problem are as follows. For some set S of grid points, the depth is known, and the surface should pass through these points. This condition may be relaxed slightly, by only requiring that the surface pass near those points, that is:

$$|f_i - c_i| < \epsilon$$

or

$$\begin{aligned} \epsilon - (f_i - c_i) &< 0 \\ \epsilon - (c_i - f_i) &< 0 \end{aligned} \qquad \forall i \in S.$$

where the c_i are the known depth values.

There are two reasons for this relaxation of the boundary conditions. One is that the stereo program only gives disparities to within a certain degree of accuracy. (Currently, this is one pixel. See the discussion for possible improvements.) Thus, one should not expect the surface to pass exactly through the points, but rather, it should pass within some distance ϵ of the points. The second reason

is that there may be some noise in the stereo output, (i.e. there may be some points assigned incorrect matches, and thus incorrect disparities.) By only requiring that the surface pass near the known depth points, the effects of such noise on the surface are diminished somewhat.

The Lagrangian for this problem is given by:

$$l(\mathrm{f},\mu) = \sum_{i=1}^{n-1} \left(f_{i-1} - 2f_i + f_{i+1}\right)^2 + \sum_{i\in S} \mu_i \cdot (\epsilon - f_i + c_i) + \mu_i' \cdot (\epsilon - c_i + f_i).$$

The partial derivatives of l are:

$$l_{f_i} = \begin{cases} 2f_0 - 4f_1 + 2f_2 + a & i = 0 \\ -4f_0 + 10f_1 - 8f_2 + 2f_3 + a & i = 1 \\ 2f_{i-2} - 8f_{i-1} + 12f_i - 8f_{i+1} + 2f_{i+2} + a & 2 \le i \le n-2 \\ 2f_{n-3} - 8f_{n-2} + 10f_{n-1} - 4f_n + a & i = n-1 \\ 2f_{n-2} - 4f_{n-1} + 2f_n + a & i = n \end{cases}$$

where

$$a = \begin{cases} \mu_i' - \mu_i & i \in S \\ 0 & i \notin S \end{cases}$$

Similarly

$$l_{\mu_i} = \epsilon - f_i + c_i$$
$$l_{\mu_i'} = \epsilon - c_i + f_i$$

Thus, the system of Arrow-Hurwicz equations becomes:

$$f_i(t+1) = \max\left[0, f_i(t) - \rho l_{f_i}(\mathrm{f}(t), \mu(t))\right]$$
$$\mu_i(t+1) = \max\left[0, u_i(t) + \rho l_{\mu_i}(\mathrm{f}(t), \mu(t))\right]$$
$$\mu_i'(t+1) = \max\left[0, \mu_i(t) + \rho l_{\mu_i'}(\mathrm{f}(t), \mu(t))\right]$$

with the partial derivatives $l_{f_i}, l_{u_i}, l_{u_i'}$ as given above.

7.5 The Two-Dimensional Case

For the two-dimensional case, consider the problem of:

$$\text{Minimize} \left\{ \int\int \left(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2\right) dxdy \right\}^{\frac{1}{2}}$$

subject to the constraints that the surface pass within some distance of a given set of points. As in the one-dimensional case, the computation may be reduced to:

THE TWO-DIMENSIONAL CASE

$$\text{Minimize} \int\int \left(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2\right) dxdy$$

and applies to a discrete grid. As a consequence, it is necessary to convert to a discrete equivalent of the differential operators. For ease of representation, the operators are shown as the set of coefficients as they would appear over the grid points of the uniform grid. (Zeros are assumed for all points without a coefficient.)

One discrete equivalent of f_{xx} is given by:

$$1 \quad -2 \quad 1$$

and is denoted f_{xx}^{ij}.

One discrete equivalent of f_{yy} is given by:

$$\begin{matrix} 1 \\ -2 \\ 1 \end{matrix}$$

and is denoted f_{yy}^{ij}.

One discrete equivalent of f_{xy} is given by:

$$\begin{matrix} -\frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & -\frac{1}{4} \end{matrix}$$

and is denoted f_{xy}^{ij}.

If $S = \{(x, y) \mid f(x, y)$ *is known* $\}$, then the Lagrangian for the two-dimensional case is:

$$l(\mathrm{f}, \mu) = \sum_i \sum_j \left\{ \left(f_{xx}^{ij}\right)^2 + 2\left(f_{xy}^{ij}\right)^2 + \left(f_{yy}^{ij}\right)^2 \right\}$$
$$+ \sum_{(i,j)\in S} \left\{ \mu_{ij} \cdot (\epsilon - f_{ij} + c_{ij}) + \mu'_{ij} \cdot (\epsilon - c_{ij} + f_{ij}) \right\}.$$

Now the partial derivatives of l are given as follows.

(1) Along the outermost rows and columns, the following are computed.

(a) In the corners, use operators of the following form. For the lower left corner, the coefficients are:

$$\begin{matrix} 1\frac{7}{8} & 0 & \frac{1}{8} \\ -4 & 0 & 0 \\ 4\frac{1}{8}+\alpha & -4 & 1\frac{7}{8} \end{matrix}$$

For the other corners, appropriately rotated versions of this operator are used.

(b) For the elements one removed from the corner, use operators of the following form. For the lower left corner, the coefficients are:

$$\begin{matrix} 1\frac{7}{8} & 0 & \frac{1}{8} \\ -8 & 0 & 0 \\ 12\frac{1}{8}+\alpha & -4 & 1\frac{7}{8} \\ -4 & 0 & 0 \end{matrix}$$

For the other locations, appropriately rotated versions of this operator are used.

(c) Elsewhere along the outermost rows and columns use operators of the following form. For the left column, the coefficients are:

$$\begin{matrix} 1\frac{7}{8} & 0 & \frac{1}{8} \\ -8 & 0 & 0 \\ 14\frac{1}{8}+\alpha & -4 & 1\frac{3}{4} \\ -8 & 0 & 0 \\ 1\frac{7}{8} & 0 & \frac{1}{8} \end{matrix}$$

For the other outermost rows and columns, appropriately rotated versions of this operator are used.

(2) Along the second outermost rows and columns, the following are computed.

(a) In the corners, use operators of the following form. For the lower left corner, the coefficients are:

$$\begin{matrix} 0 & 1\frac{7}{8} & 0 & \frac{1}{8} \\ 0 & -8 & 0 & 0 \\ -4 & 20\frac{1}{8}+\alpha & -8 & 1\frac{3}{4} \\ 0 & -4 & 0 & 0 \end{matrix}$$

For the other corners, appropriately rotated versions of this operator are used.

(b) Elsewhere, along the second rows and columns, use operators of the following form. For the lower left corner, the coefficients are:

$$\begin{matrix} 0 & 1\frac{7}{8} & 0 & \frac{1}{8} \\ 0 & -8 & 0 & 0 \\ -4 & 22\frac{1}{4}+\alpha & -8 & 1\frac{3}{4} \\ 0 & -8 & 0 & 0 \\ 0 & 1\frac{7}{8} & 0 & \frac{1}{8} \end{matrix}$$

For the other second outermost rows and columns, appropriately rotated versions of the this operator are used.

(3) Elsewhere, the coefficients of the operator are given by:

$$\begin{matrix} \frac{1}{8} & 0 & 1\frac{3}{4} & 0 & \frac{1}{8} \\ 0 & 0 & -8 & 0 & 0 \\ 1\frac{3}{4} & -8 & 24\frac{1}{2}+\alpha & -8 & 1\frac{3}{4} \\ 0 & 0 & -8 & 0 & 0 \\ \frac{1}{8} & 0 & 1\frac{3}{4} & 0 & \frac{1}{8} \end{matrix}$$

The variable α is given by

$$\alpha = \begin{cases} \mu'_{ij} - \mu_{ij} & (i,j) \in S \\ 0 & (i,j) \notin S \end{cases}$$

Finally, the partial derivatives with respect to the Lagrangian variables are given by:

$$l_{\mu_{ij}} = \epsilon - f_{ij} + c_{ij}$$
$$l_{\mu'_{ij}} = \epsilon - c_{ij} + f_{ij}$$

Thus, an Arrow-Hurwicz system of equations has been defined for the value f_{ij} of the surface at the grid points, with Lagrangian constraint variables u_{ij}, u'_{ij} at the known depth points. Provided the step size ρ is small enough, this system is guaranteed to converge. Moreover, in the case of f being strictly convex, the system has a unique solution, (Arrow, Hurwicz and Uzawa, 1958).

CHAPTER 8

THE INTERPOLATION ALGORITHM

In the previous chapter, the Arrow-Hurwicz method for constrained optima was outlined. We can now turn to the design of an explicit algorithm to perform the interpolation of surfaces, over a uniform grid.

We stated the form of semi-norm to use:

$$\Theta(f) = \left\{ \int \int \left(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2 \right) dx dy \right\}^{\frac{1}{2}}.$$

For this particular semi-norm, the steps run as follows.

(0) Determine the starting position for the iteration.

(1) Convolve the current approximation of the surface (except for the two outermost rows) with a mask given by

$$\begin{array}{ccccc} \frac{1}{8} & 0 & 1\frac{3}{4} & 0 & \frac{1}{8} \\ 0 & 0 & -8 & 0 & 0 \\ 1\frac{3}{4} & -8 & 24\frac{1}{2} & -8 & 1\frac{3}{4} \\ 0 & 0 & -8 & 0 & 0 \\ \frac{1}{8} & 0 & 1\frac{3}{4} & 0 & \frac{1}{8} \end{array}$$

Set the outer row by convolving with corner masks of the form

$$\begin{matrix} 1\frac{7}{8} & 0 & \frac{1}{8} \\ -4 & 0 & 0 \\ 4\frac{1}{8} & -4 & 1\frac{7}{8} \end{matrix}$$

and its equivalent forms; by using, for elements one removed from the corner, masks of the form

$$\begin{matrix} 1\frac{7}{8} & 0 & \frac{1}{8} \\ -8 & 0 & 0 \\ 12\frac{1}{8} & -4 & 1\frac{7}{8} \\ -4 & 0 & 0 \end{matrix}$$

and its equivalent forms; and by using, elsewhere along the outermost rows, masks of the form

$$\begin{matrix} 1\frac{7}{8} & 0 & \frac{1}{8} \\ -8 & 0 & 0 \\ 14\frac{1}{8} & -4 & 1\frac{3}{4} \\ -8 & 0 & 0 \\ 1\frac{7}{8} & 0 & \frac{1}{8} \end{matrix}$$

and its equivalent forms.

Set the second row by convolving with the corner masks of the form

$$\begin{matrix} 0 & 1\frac{7}{8} & 0 & \frac{1}{8} \\ 0 & -8 & 0 & 0 \\ -4 & 20\frac{1}{8} & -8 & 1\frac{3}{4} \\ 0 & -4 & 0 & 0 \end{matrix}$$

and its equivalent forms; and by using, elsewhere along the row, masks of the form

$$\begin{matrix} 0 & 1\frac{7}{8} & 0 & \frac{1}{8} \\ 0 & -8 & 0 & 0 \\ -4 & 22\frac{1}{4} & -8 & 1\frac{3}{4} \\ 0 & -8 & 0 & 0 \\ 0 & 1\frac{7}{8} & 0 & \frac{1}{8} \end{matrix}$$

and its equivalent forms.

(2) To the convolved values from step (1), add the values $\mu'_{ij} - \mu_{ij}$ at the known depth points, $(i, j) \in S$.

(3) Scale the entire grid of convolved values from step (2) by the unit step size ρ.

(4) Add the scaled convolution values of step (3) to the current approximation of the surface.

(5) Threshold the values obtained in step (4), in order to remove any negative entries and replace them with **0**.

This completes the process of updating the approximation of the surface at the grid points. To update the Lagrangian multipliers μ_{ij}, μ'_{ij}, the following steps are performed.

(6) For each known grid point, $(i, j) \in S$, compute:

$$l_{\mu_{ij}} = \epsilon - f_{ij} + c_{ij}$$
$$l_{\mu'_{ij}} = \epsilon - c_{ij} + f_{ij}$$

(7) Scale the values obtained in step 6 by the unit step size ρ and subtract from the current values of the multipliers μ_{ij}, μ'_{ij}.

(8) Threshold the values obtained in step (7), in order to remove any negative entries and replace them with **0**.

Having created an explicit algorithm, one can now consider how well it fits the algorithmic criteria and how well the results of applying it meet the computational constraints. The algorithmic criteria are considered first.

Parallel: Clearly, the repetitive structure of the computation allows for a straightforward parallel implementation, composed of simple processors positioned at each grid point. This is one of the most important algorithmic constraints, since it allows the computation to be performed very quickly.

Simple: Given a parallel implementation, it is clear that the operations performed at each point of the grid, involving only a few additions and multiplications, are very simple.

Uniform: Given a parallel implementation of the computation, consider the computations performed by each individual processor. Except for the edges of the image, the processes are almost identical for each grid point. The only difference is the extra addition performed at grid points corresponding to known depth points.

Local-support: This is one of the strongest requirements of the algorithm, since it reduces the amount of communication, and hence the number of connections, needed by each processor.

The proposed algorithm clearly satisfies this constraint, since each processor must access at most 14 different values, which are all localized spatially on the grid about the processor.

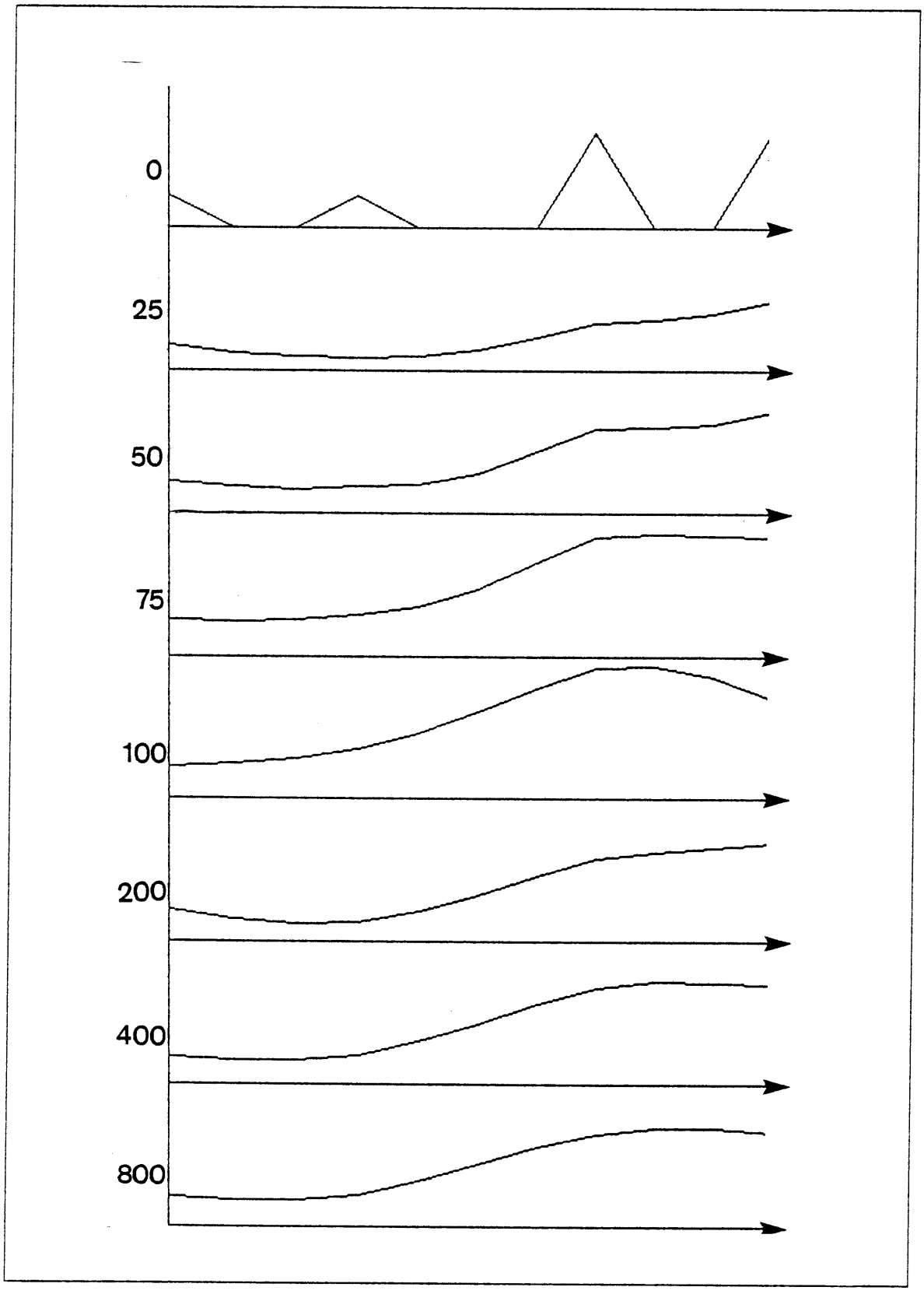
Rapid: The main factor in determining the rapidity of an algorithm is the speed at which a reasonable approximation of the surface may be extracted. Since the process is iterative, a major factor contributing to the speed of the algorithm is the rate of convergence of the iteration. This is particularly true in this case, since the *cost* of an individual step may be regarded as small; that is, the time needed to perform additions and multiplications is small.

The rate of convergence of the iteration is usually defined as the rate at which the iteration moves from a given initial position to its next position. While this will determine the rate at which the iteration converges, it does not exactly account for the factor of interest, namely, how long does it take the algorithm converge to an acceptable approximation of the surface?

Tthe speed at which the algorithm attains a reasonable surface approximation is critically dependent on both the initial starting position and the unit step size ρ, which determines the rate of convergence. To illustrate the effect of these factors on the speed of the algorithm, a one-dimensional example of interpolation is considered. For this example, there are four known depth points, at $x = 0, 3, 7, 10$.

Perhaps the most critical factor affecting the speed of the algorithm is the initial position. To illustrate this, two cases are considered. In the first, the starting position for the iteration is found by setting the surface to the known depth values at those grid points where such values occur, and by setting the surface to 0 elsewhere. For a step size of $\rho = 0.05$ and a constraint parameter of $\epsilon = 0.05$ the state of the algorithm after various iterations is illustrated in Figure 8.1.

Figure 8.1. Interpolation from Zeroed Starting Position. The top figure shows the initial starting position for the surface, and the lower figures show the surface approximations after 25, 50, 75, 100, 200, 400 and 800 iterations.


0
25
50
75
100
200
400
800

Several observations are noteworthy. The first is that the number of iterations required to converge to within some reasonable distance of the limiting surface (the surface to which the algorithm tends as the number of iterations becomes large) is enormous — on the order of 400. Second, the iterative surface approximations will oscillate about the limiting surface. If the step size ρ is too large, the oscillation is undamped and the process does not converge. Provided that ρ is small enough, the oscillation is damped and the iterative surface approximations will converge to the limiting surface. The envelope spanning the iterative approximations is very broad in the case under consideration, as can be seen by the approximations after 25, 50, 75 and 100 iterations. As a consequence, a large number of iterations must pass before the envelope of the approximations is reasonably bounded about the limiting surface. Thus, if one wishes to be able to halt the iteration at any point and be guaranteed a good approximation to the surface, one must allow at least 400 iterations to pass.

As a second case, the initial position is altered as follows. For each point x_i on the grid, a neighbourhood of some size is considered, (in the example shown, the neighbourhood was 5 elements). For this neighbourhood, a least squares fit of the known depth points in the neighbourhood is performed. The value of the least squares fit at the point x_i is assigned as the value of the starting position for that point. This operation is performed for all points on the grid, thereby generating the initial starting point for the algorithm. For the example being considered, the initial starting point generated in this manner is shown in Figure 8.2. Given this starting position, it can be seen that the number of iterations required to reach a reasonable approximation to the surface drops drastically. After only 10 iterations, the basic shape of the limiting surface is evident. Furthermore, the envelope of the iterative approximations is much smaller, so that at any stage, the current approximation is close to the limiting surface, and forms an acceptable approximation.

If the step size ρ is decreased, the rate of convergence, of course, also decreases. This is shown in Figure 8.3, where ρ has been decreased by an order of magnitude. The speed of the algorithm is also decreased by roughly an order of magnitude, requiring 100 iterations to reach the state which the previous example reached in 10 iterations.

Figure 8.2. Interpolation of Least-Squares Initial Position. The top figure shows the initial position of the surface, which has been formed by a least-squares fitting of the known data points. The lower figures show the surface approximation after 5, 10, 20, 50, 100, 200 and 400 iterations.

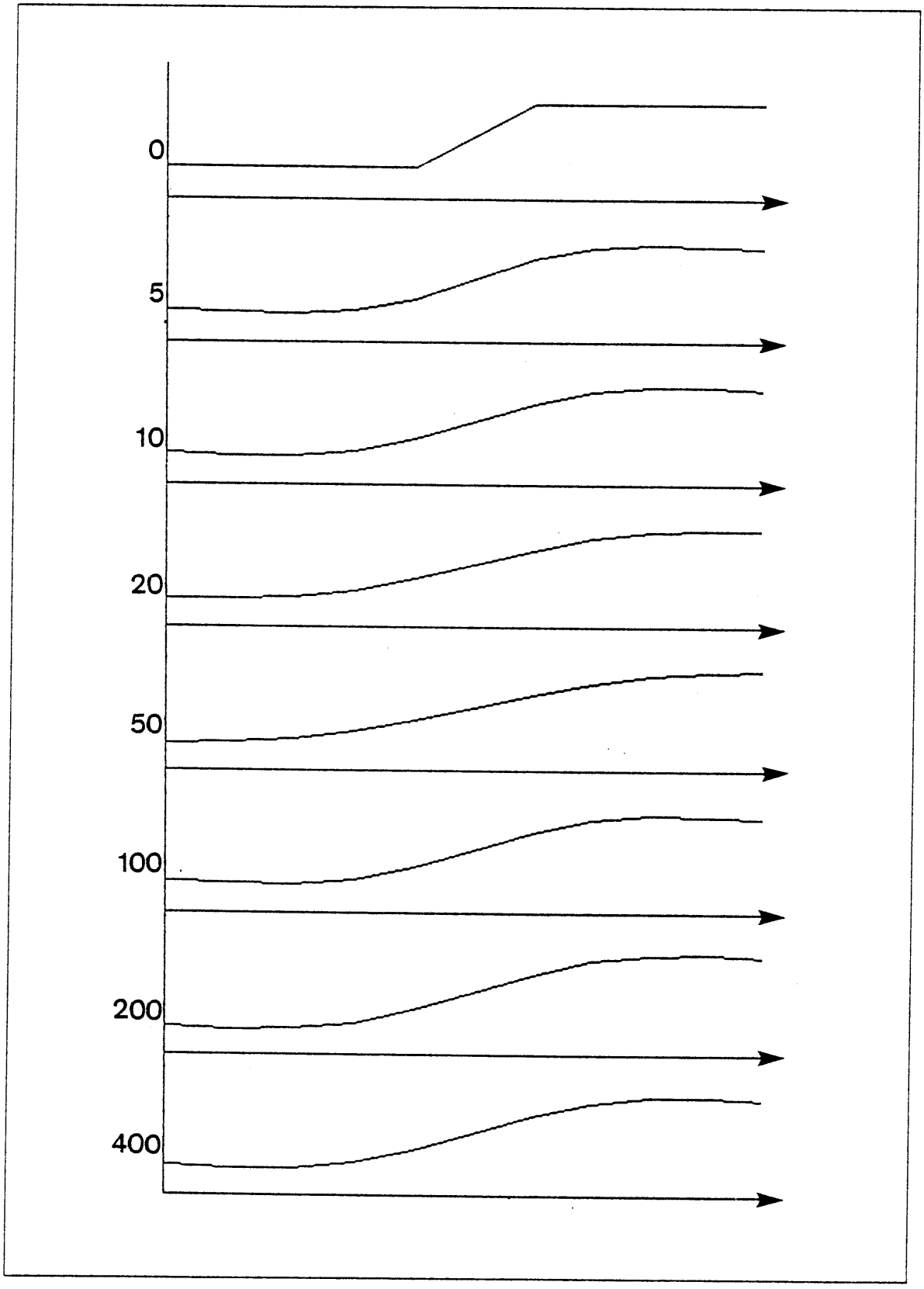
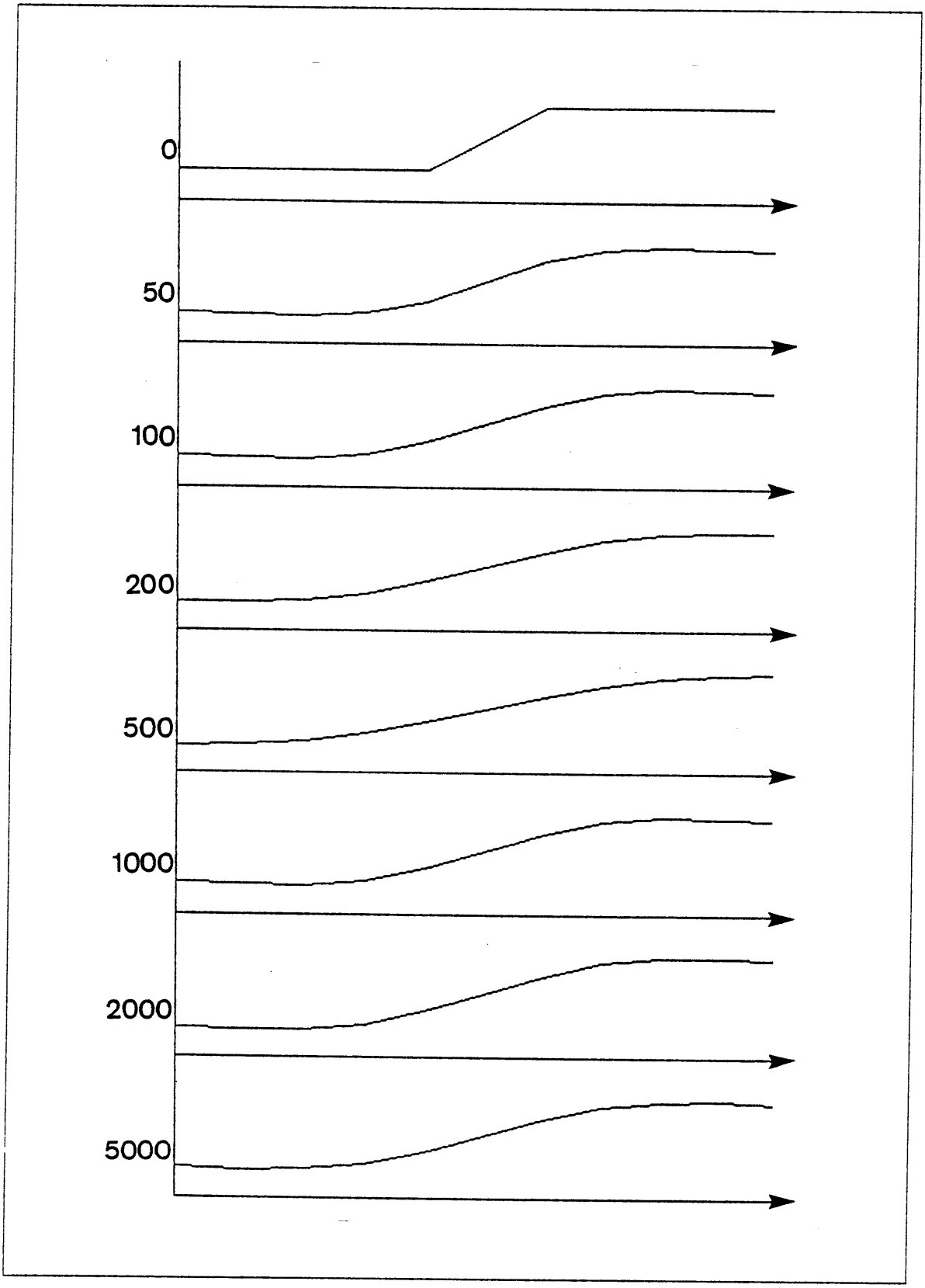

0
5
10
20
50
100
200
400

Figure 8.3. Interpolation with Decreased Step Size. The top figure shows the initial surface approximation. The bottom figures show the surface approximations after different numbers of iterations. In this case, the step size ρ has been decreased by a factor of 10 from that of Figure 8.2.


0
50
100
200
500
1000
2000
5000

My conclusion from this set of examples is that given a reasonable starting position, the algorithm is rapid, with an acceptable surface approximation obtained within very few iterations.

Thus, one final addition may be made to the proposed algorithm:

(0) For each point of the grid, take a neighbourhood of some size about the point, and perform a least squares fit through the known points in the neighbourhood. At the point under consideration, assign as a starting value, the value associated with that point in the least squares fit.

Thus, the proposed algorithm is consistent the algorithmic criteria of parallelism, rapidity, local-support, uniformity and simplicity.

8.1 Performance

We now turn to the consideration of how well the algorithm satisfies the computational constraint of containing a minimum amount of bending, while passing through a set of known points.

We have already seen mathematical arguments supporting the computational problem set out in Chapter 6. As in the case of the stereo theory, by creating and implementing an explicit algorithm for the theory, its adequacy can be tested. To do this, a set of examples of applying the algorithm are considered.

The first examples, shown in Figures 8.4-8.8, involve a set of synthetically generated boundary conditions. In each case, the boundary conditions of the known depth points are shown along two sets of grid lines, as well as the interpolated surface constructed by the program. The examples include a cylinder, portions of a sphere, and portions of some hyperbolic paraboloids. The examples all demonstrate the capability of the algorithm to fill in smooth surfaces between the known points.

It is worthwhile noting the effect of the constraint parameter, ϵ, on the shape of the surface. To illustrate this, the one-dimensional example is again considered. In Figure 8.2, the constraint parameter was set to $\epsilon = 0.05$, forcing a strong constraint at each known point. In Figures 8.9 and 8.10, ϵ is relaxed to 0.2 and to 1.0. The relaxing of the constraints can be seen to alter the shape of the surface. Thus, depending on the confidence associated with the depth points, a small ϵ will be needed to enforce a good approximation to the surface. I shall return to this point in Chapter 9.

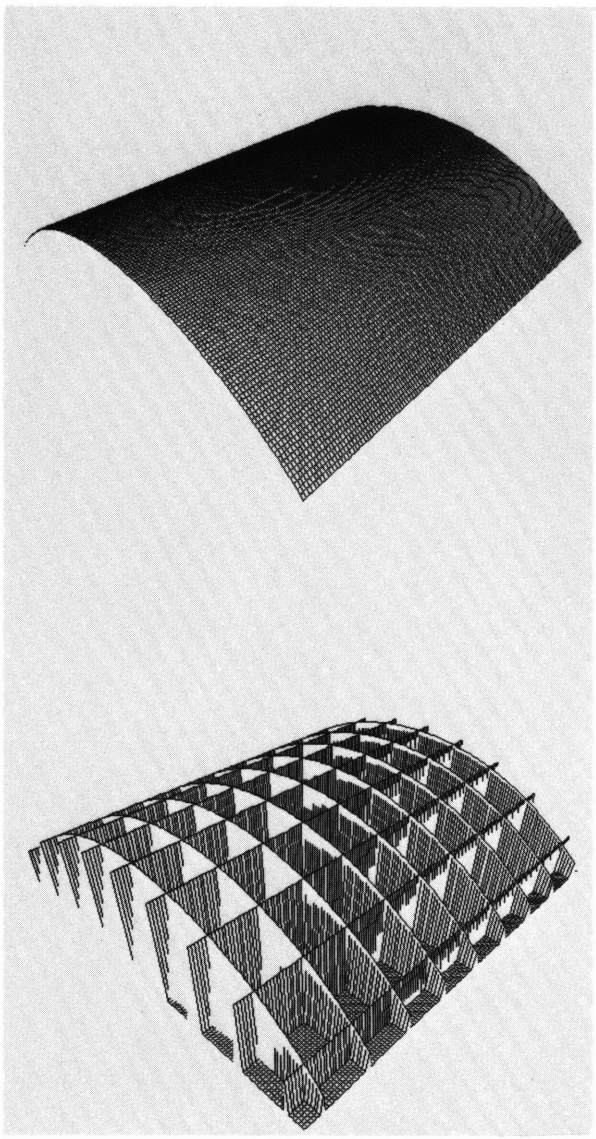


Figure 8.4. Interpolated Cylinder. The lower figure shows the boundary conditions of a cylinder. The upper figure shows the interpolated surface.

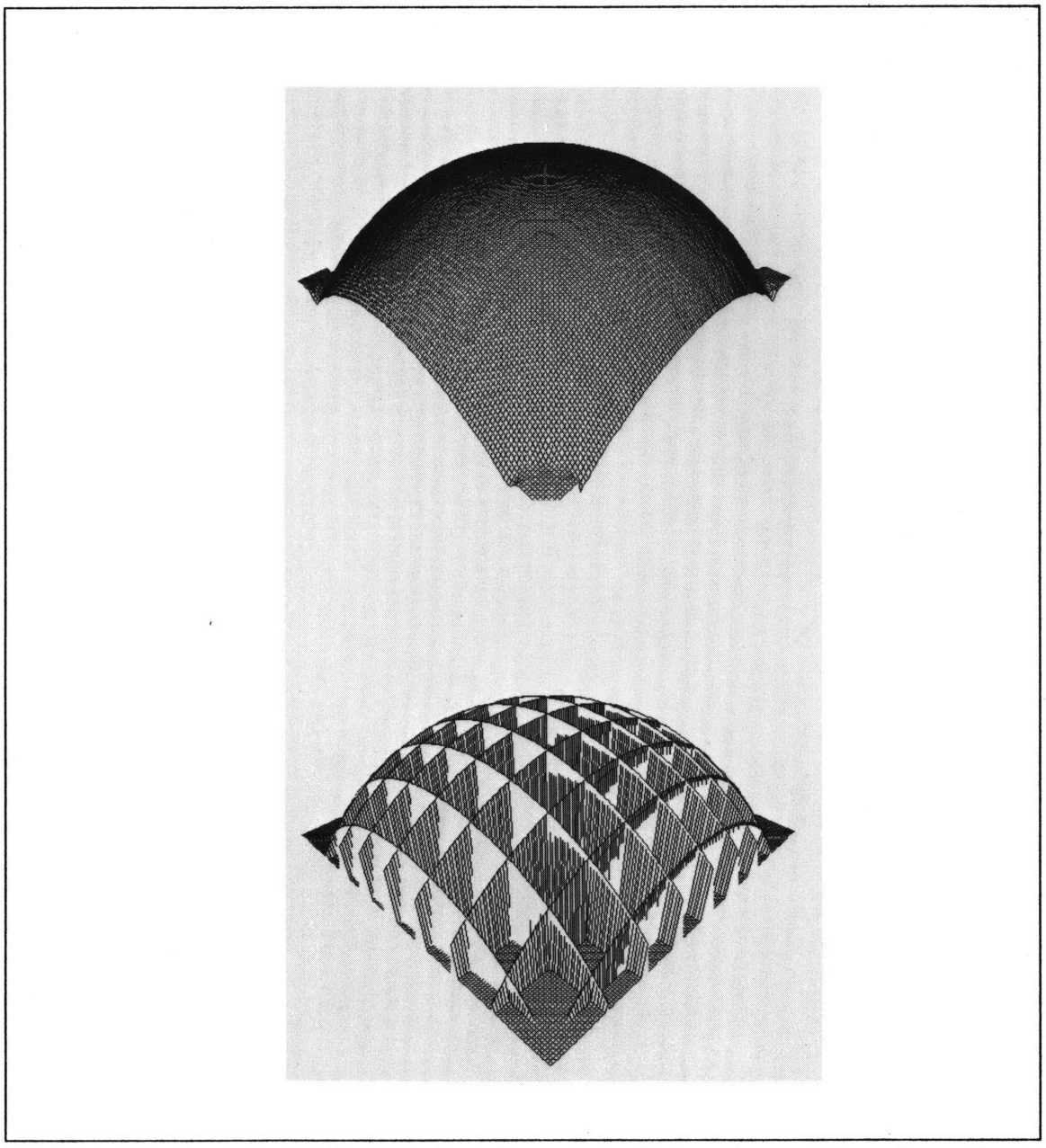


Figure 8.5. Interpolated Sphere. The lower figure shows the boundary conditions of a portion of a sphere. The upper figure shows the interpolated surface.

8.2 Discontinuities

Note that this implementation of the interpolation algorithm does not account for one of the main factors necessary in a description of surface shape — surface discontinuities. In the next chaper,

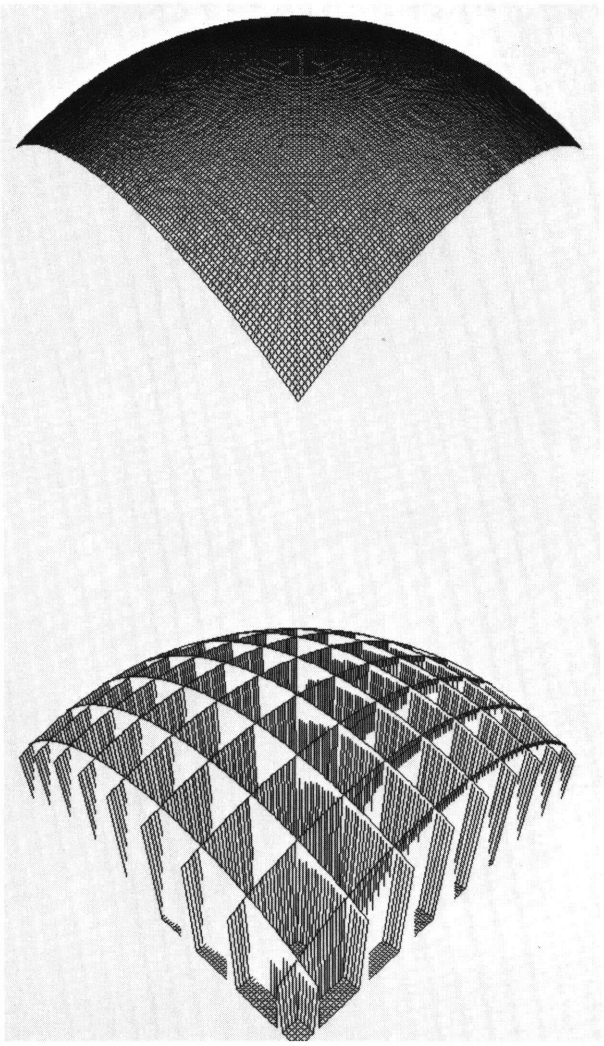


Figure 8.6. Interpolated Sphere. The lower figure shows the boundary conditions of portion of a sphere. The upper figure shows the interpolated surface.

the effects of omitting this component from the algorithm will be discussed, and possible methods for detecting surface discontinuities and occluding contours will be considered.

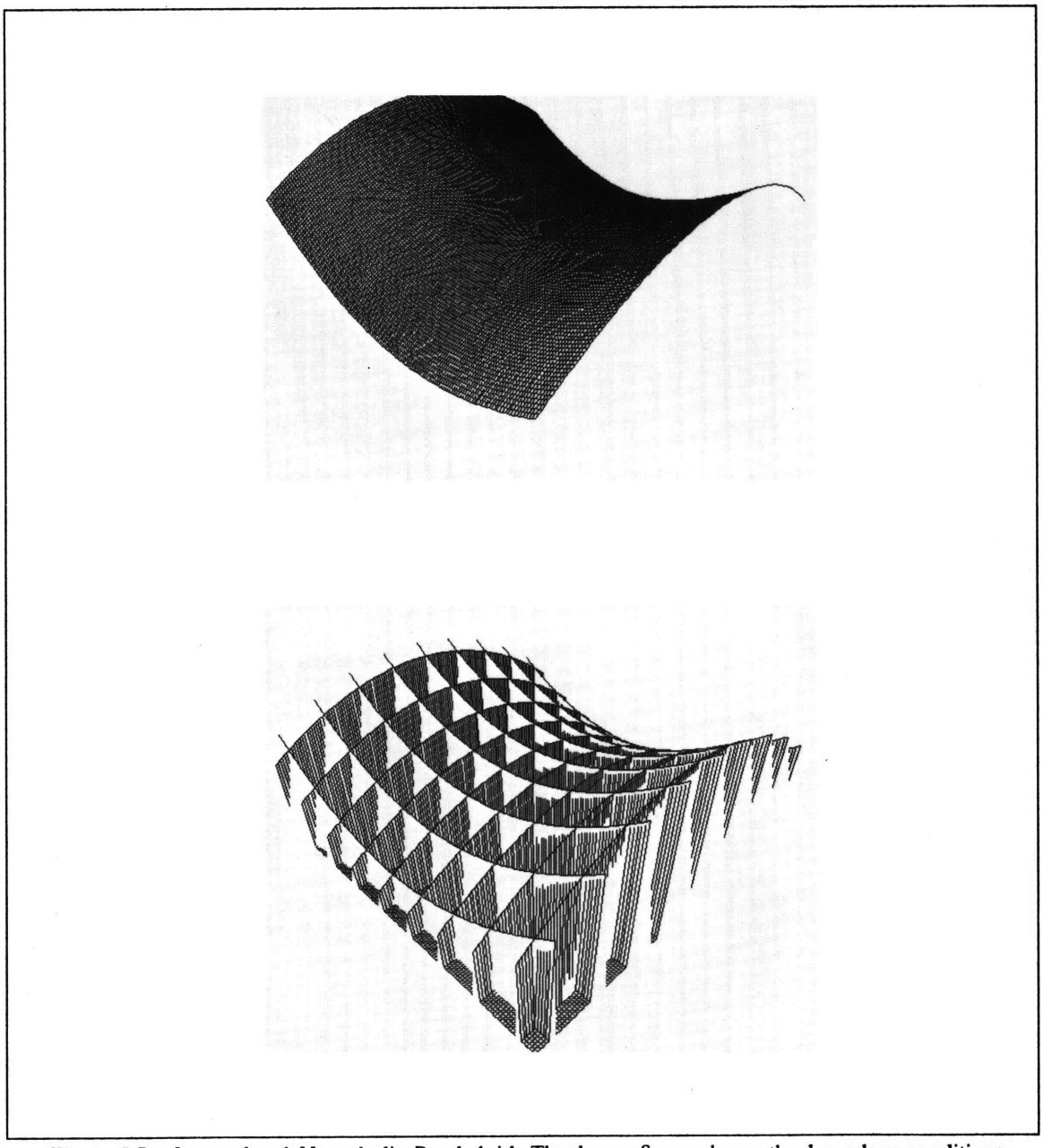


Figure 8.7. Interpolated Hyperbolic Parabaloid. The lower figure shows the boundary conditions of a hyperbolic parabaloid. The upper figure shows the interpolated surface.

8.3 Putting It All Together

Finally, all the pieces needed to accomplish our goal — the explicit reconstruction of three-dimensional surfaces, in this case from stereo images — have been assembled. The steps are:

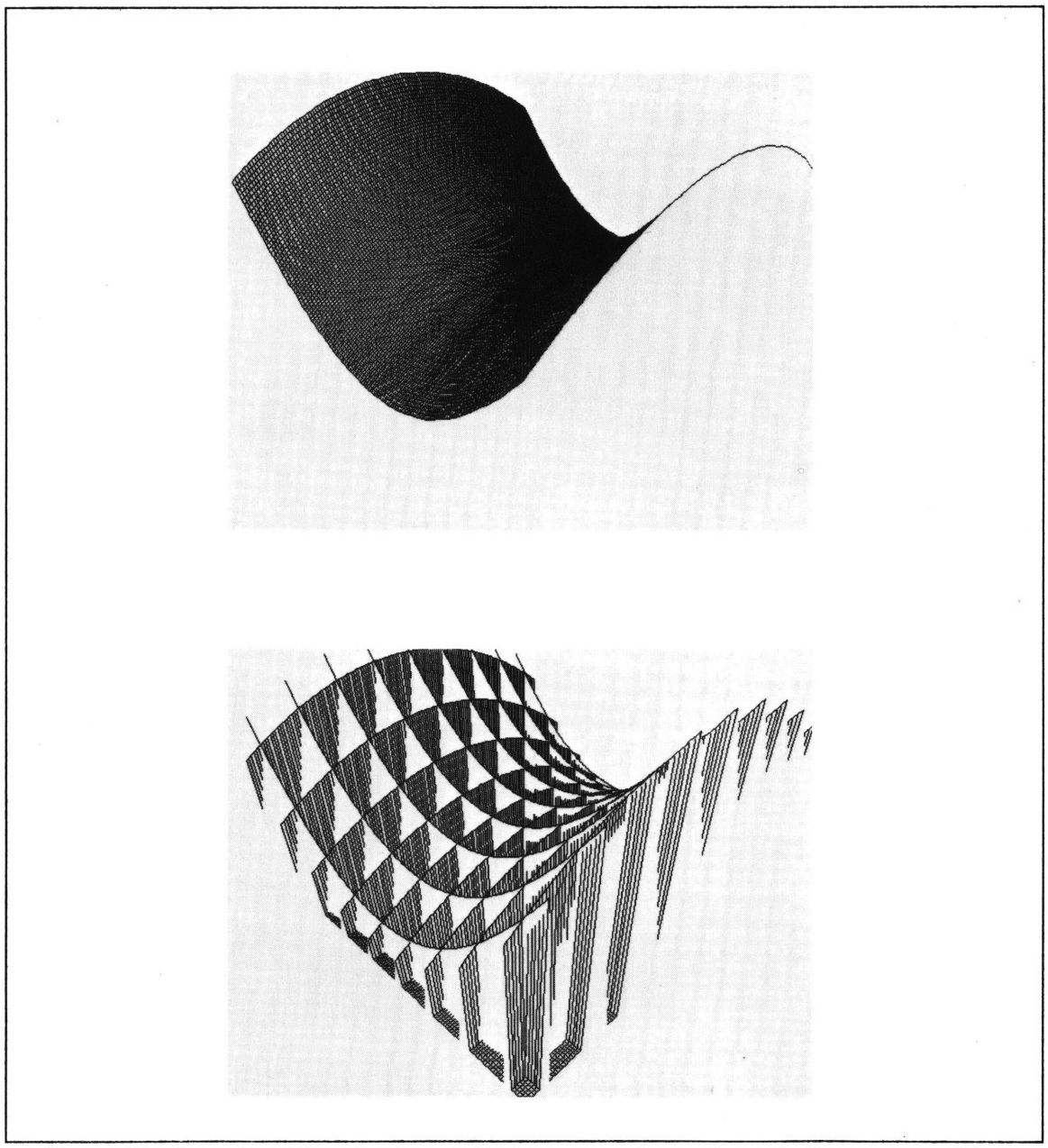


Figure 8.8. Interpolated Hyperbolic Parabaloid. The lower figure shows the boundary conditions of a hyperbolic parabaloid. The upper figure shows the interpolated surface.

(1) Process the images with $\nabla^2 G$ at several scales and obtain zero-crossing descriptions of the changes in the images.

(2) Use the Marr-Poggio stereo matcher to obtain disparity information about the zero-crossings.

Figure 8.9. Interpolation with Relaxation of the Constraint Parameter. The conditions are similar to Figure 8.2, save that ϵ is relaxed to 0.2.

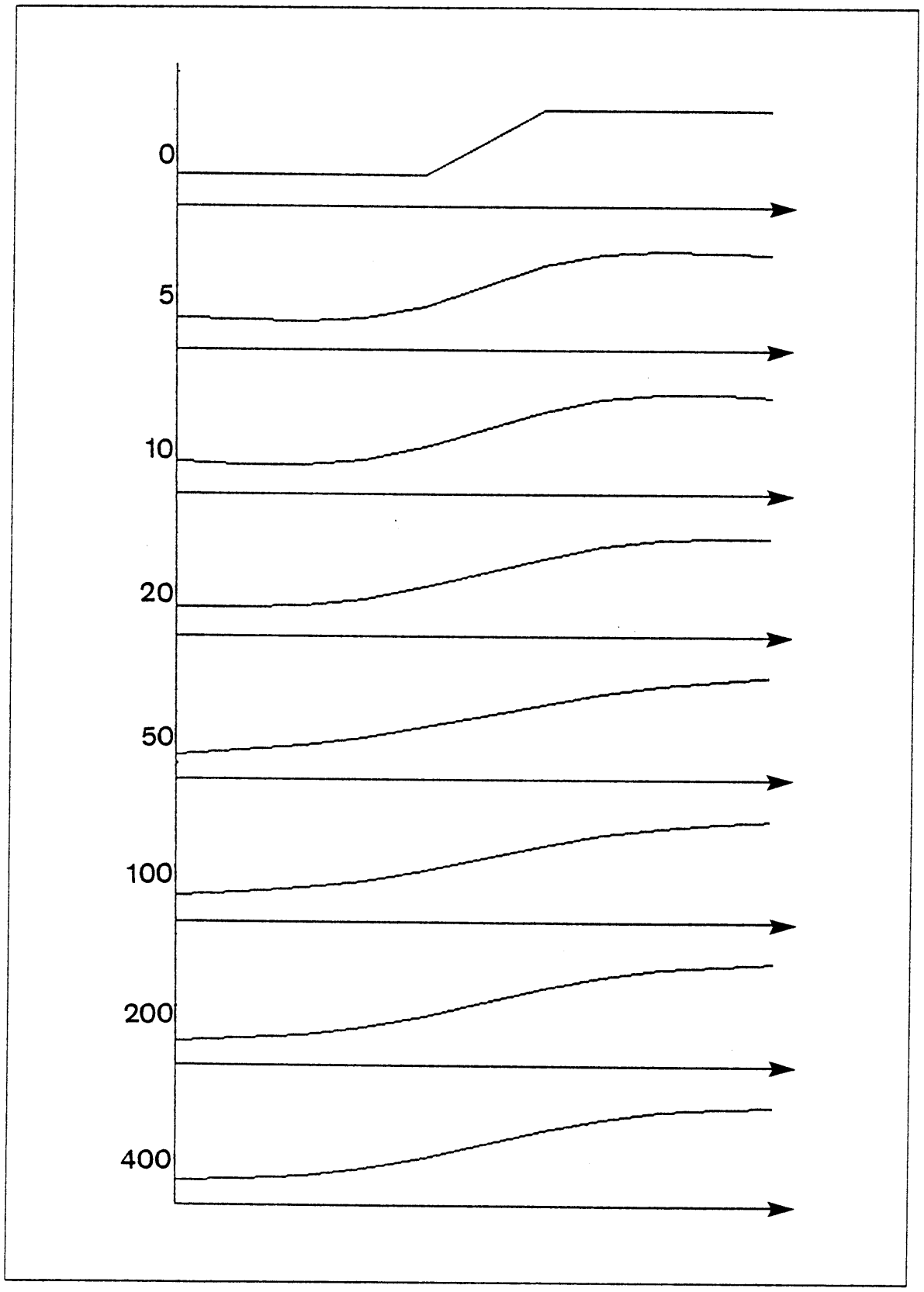
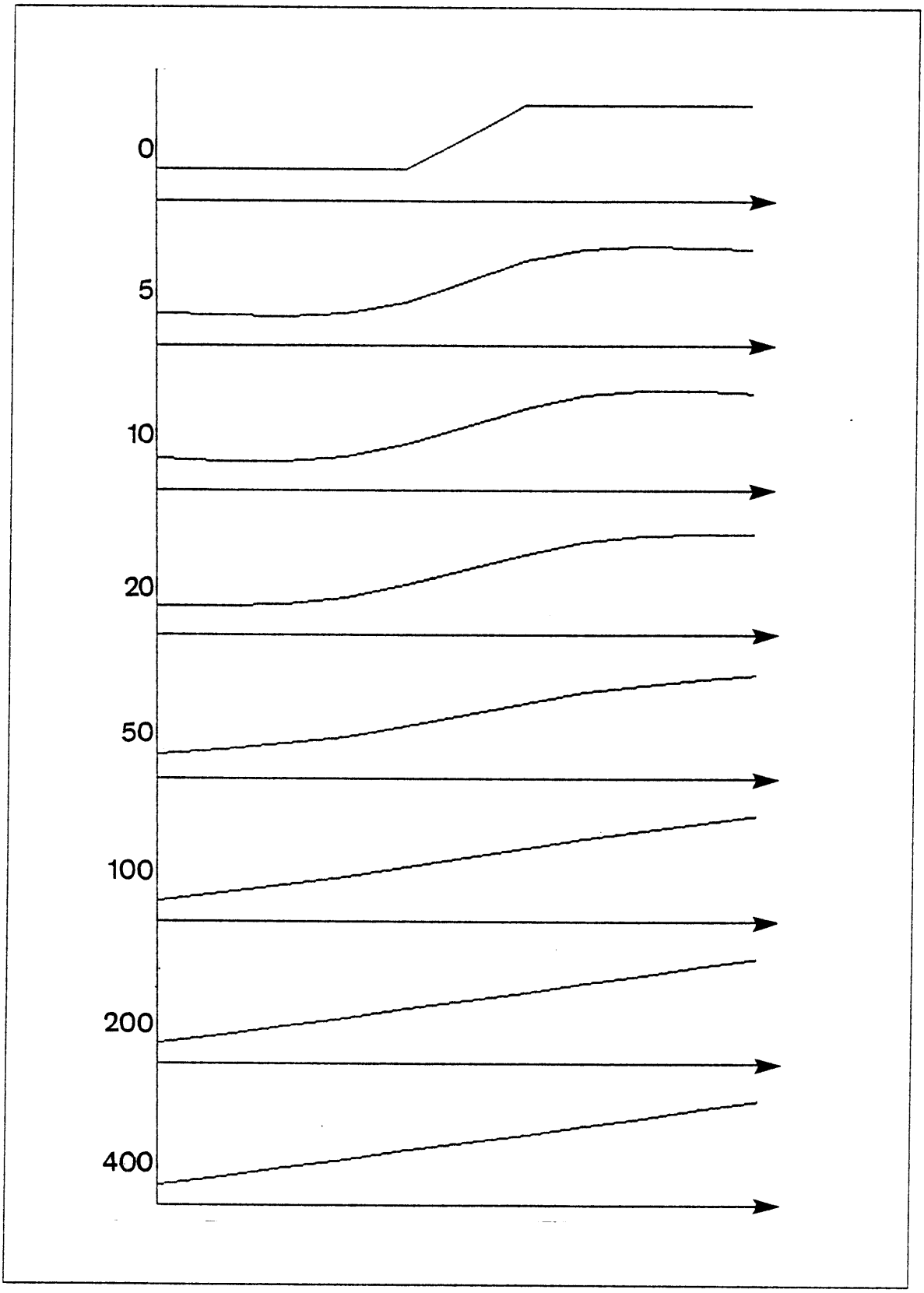

0
5
10
20
50
100
200
400

Figure 8.10. Interpolation with Relaxation of the Constraint Parameter. The conditions are similar to Figure 8.2, save that ϵ is relaxed to 1.0.


0
5
10
20
50
100
200
400

(3) Apply the Arrow-Hurwicz system of equations corresponding to minimizing

$$\int\int \left(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2\right) dxdy$$

subject to the condition that the surface f pass through the points obtained in step (2).

The complete system is illustrated on some examples. The first is the wedding cake random dot pattern of Figure 4.3. We have already seen the performance of the Marr-Poggio stereo theory on this example. The interpolated surface obtained by the algorithm is shown in Figure 8.11. The planar structure of the wedding cake is evident. Because discontinuities are not made explicit, a certain amount of smoothing takes place across the edges of the planes.

The second example is the painted coffee jar of Figure 3.4. The results of the Marr-Poggio stereo theory were shown in that figure and the interpolated surface is shown in Figure 8.12. The general shape of the jar and the background are clearly visible. Because of the resolution of the grid on which the interpolation takes place, a certain amount of local distortion of the surface is evident. A second observation is that the shape of the bottle is somewhat square compared with the actual shape. This is in part because of the lack of explicit discontinuities. Since the discontinuities are not made explicit, information from the background can affect the shape of the bottle, and in attempting to fit a smooth surface to both objects, the edges of the bottle are driven upward, resulting in a square shape.

As a third example, a portion of the Martian surface is considered, using stereo photographs obtained by the Viking lander. The interpolated surface is shown in Figure 8.13. Interestingly, although the monocular images give the impression of a flat plane running off to the horizon, the program finds a set of distinct ridges in depth, with places of sharp discontinuity in depth. When fused, the images yield the same impression. It is also of interest to note that the total range of disparities, from the horizon to the foreground, found by the program is on the order of 200 picture elements.

In the next chapter, possible refinements of the method are discussed.

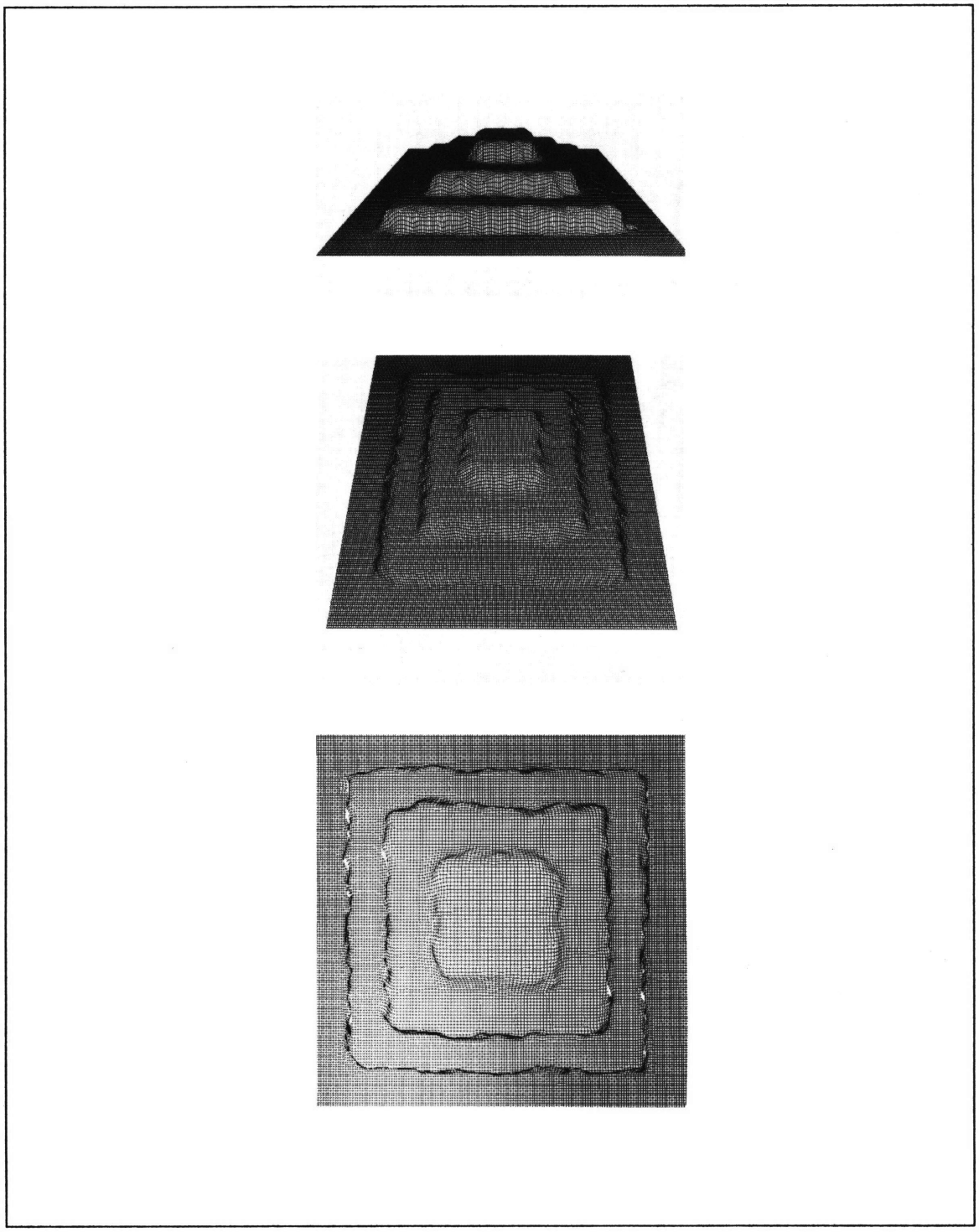


Figure 8.11. The Interpolated Wedding Cake. Three views of the interpolated surface are shown. The four distinct planes are clearly visible. Note that although the discontinuities in the surfaces have not been explicitly accounted for, they are clearly visible in the figures.

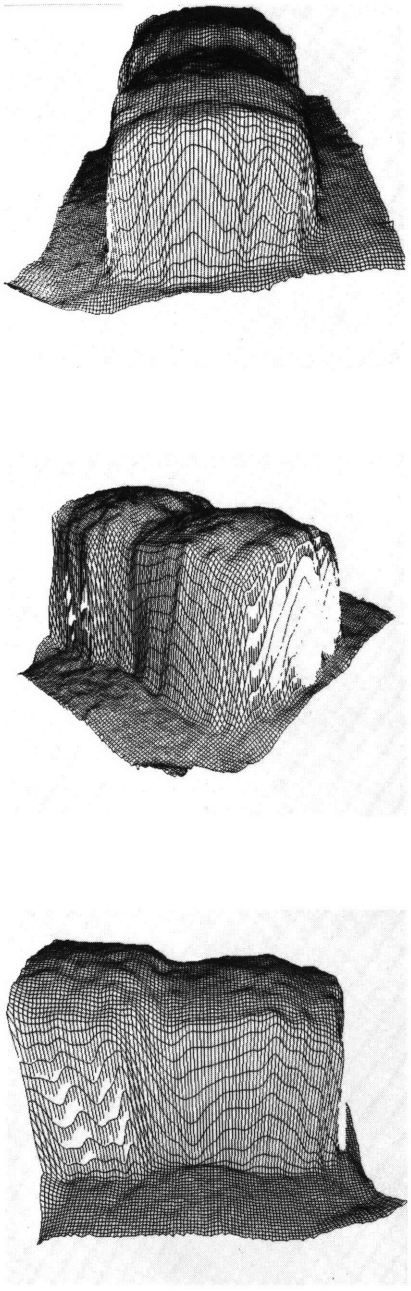


Figure 8.12. The Interpolated Coffee Jar. Three views of the interpolated surface are shown. The general shape of the jar and the background are clearly visible. Because the resolution of the grid on which the interpolation takes place is the same as that of the original images, a certain amount of high frequency distortion of the interpolated surface is evident. This is discussed in Chapter 9.

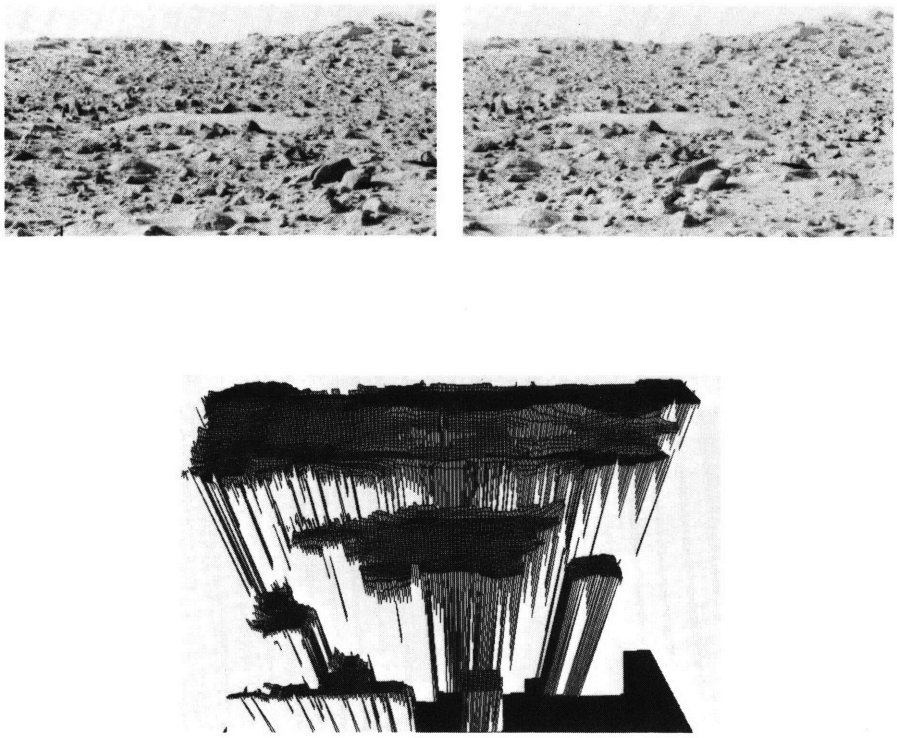


Figure 8.13. The Interpolated Martian Surface. The top pair of images are the stereo pair. The lower figure shows the interpolated disparity array. The height of a point in the array corresponds to the distance to the point in the image. The total range of disparity in this image is roughly 200 pixels. Some sections of the foreground were not matched by the stereo algorithm, and have not been interpolated. It is interesting to note that the disparity map contains a series of sharp breaks in disparity, corresponding to occluding hills in the image. These breaks are not evident in the monocular images, yet are clearly visible when the two images are fused.

CHAPTER 9

ANALYSIS AND REFINEMENTS

In the previous chapter, a demonstration of the achievementof the original goal — the design of a method for constructing, from a pair of stereo images, a complete surface specification — was presented, by demonstrating the effectiveness of the algorithm on a set of images. However, there are still aspects of the algorithm which can be improved In this chapter, possible refinements to the method are discussed. Some of these have been implemented and tested, others form possible avenues of future work.

9.1 Discontinuities

One of the implicit assumptions of the interpolation algorithm is that the pieces of surface are in fact pieces of a single surface. Of course, for almost all images, this will not be the case. What alterations are necessary in order to account for the existence of several surfaces within a scene? Does the lack of explicit discontinuities in the surface representation has an important effect on it?

One of the problems associated with the failure to make surface discontinuities explicit is that information about the shape of one surface can affect the shape of an adjacent surface. This is illustrated in the following example. A set of known depth points is given in Figure 9.1. Intuitively, the most likely surface to fit through these points would be a pair of planes with a discontinuity in depth between them. However, the requirement of a smooth surface to fit through these points results in a warping and rippling of the surface that is undesirable. Thus, it seems that the lack of explicit

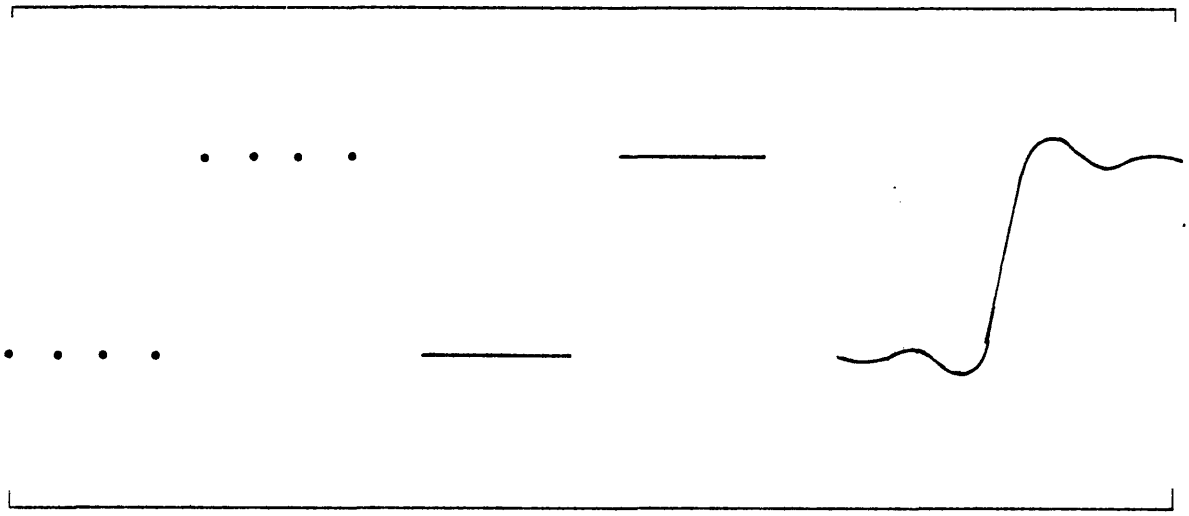


Figure 9.1. Discontinuities in the Surfaces. The left figure shows a set of known data points. Intuitively, the correct reconstructed surface would be a pair of planes, with a discontinuity between them, as shown in the middle figure. If the interpolation algorithm attempts to reconstruct a surface through the boundary points, without a discontinuity, the result is as shown in the right figure. The sharp change in depth results in a rippling of the surface.

discontinuities can affect the shapes of the interpolated surfaces in an unacceptable manner.

In order to make discontinuities explicit, there are several questions to ask about the process. How do are the discontinuities detected? Where are they placed in the representation? When does the detection of discontinuities take place relative to the interpolation process?

Two possible methods of detecting the discontinuities are discussed in turn.

9.1.1 Occlusions in the Stereo Algorithm

Consider the geometry indicated in Figure 9.2. There are regions of the left image which will not have a corresponding part in the right image, and vice versa. Consequently, any zero-crossings in this portion of one image will have no counterpart in the other image, and the stereo program should not assign any match to such zero-crossings. Hence, one proposal would be a mechanism for detecting occlusions by searching for portions of the image which contain unmatched zero-crossings. Then, the interpolation can be restricted to take place only over those sections of the image which are bounded by zero-crossings with known disparity values.

This method would detect the discontinuities before the interpolation, since it uses stereo information directly to detect the occlusions. A problem with the method is that it will not detect all

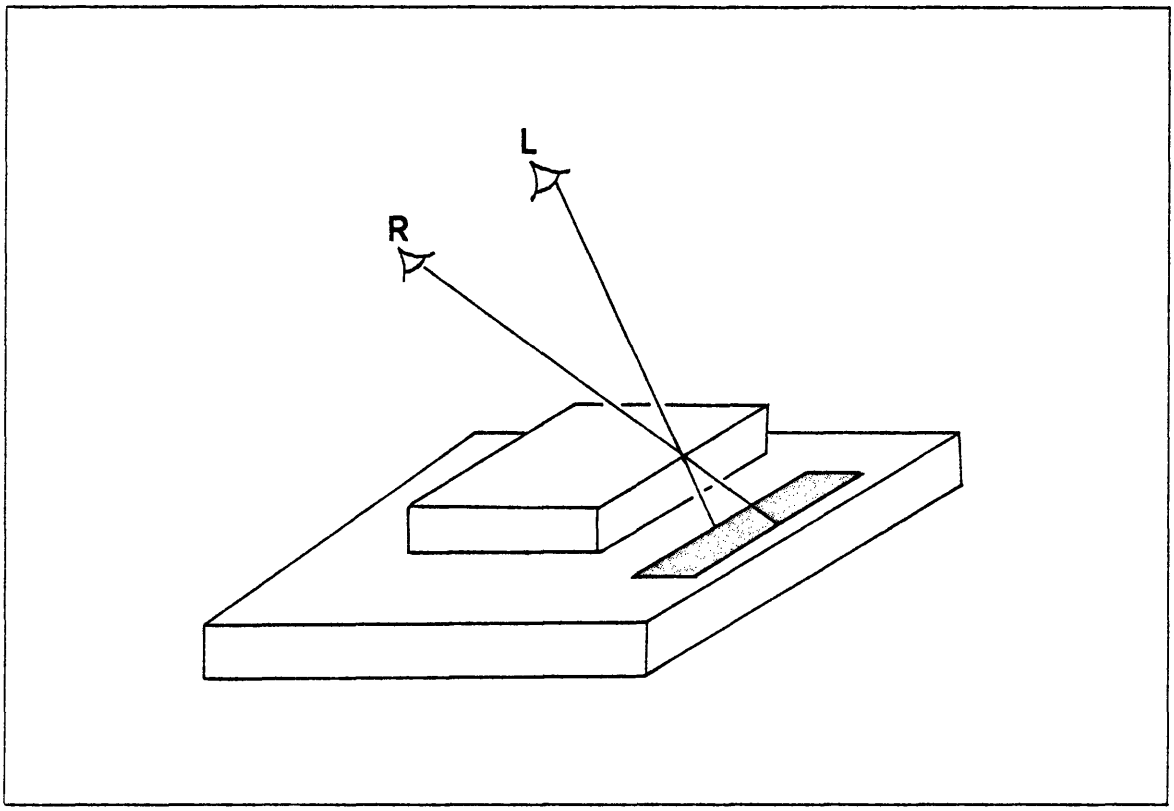


Figure 9.2. Occlusions. The upper surface occludes portions of the lower surface in each eye. These portions are different for the two eyes. The cross-hatched area of the lower surface indicates the region of the surface visible to the left eye, but not to the right.

discontinuities, only those in the horizontal direction, because a discontinuity can occur in the vertical direction without causing an occlusion. Hence, any method for detecting discontinuities which relies only on the unmatched zero-crossings will be incomplete.

9.1.2 The Primal Sketch Revisited

Perhaps the simplest method is again to use the ideas inherent in the primal sketch. Recall that the primal sketch created descriptions of points in the image associated with inflections in intensity, for a range of resolutions. Since the image intensities may be considered as a type of three-dimensional surface, the primal sketch operators essentially detect discontinuities in the image intensities for a range of resolutions. Thus, one could apply the same type of analysis to the detection of surface discontinuities, where now the surface on which the operators apply is the reconstructed depth surface, rather than the intensity surface.

It is worth noting that not only should the operators be of the form used in the extraction of the primal sketch, but that it may also be useful to use a range of operators, as in the primal sketch. Recall that the reason for using multiple zero-crossing detectors was that surface changes, and hence intensity changes, could take place over a wide range of scales. This is still true in the case of surface descriptions, such as have been constructed for the coffee jar or the wedding cake. Thus, surface discontinuities corresponding to occluding edges will frequently tend to correspond to large surface changes, while internal surface discontinuities, due to a warping of the surface, will tend to correspond to small surface changes. By using a range of $\nabla^2 G$ operators, one can extract both occluding contour discontinuities, as well as ripples or warpings of the surface itself.

Note that this method would require that the surface interpolation had already taken place, before it could be applied. Since one of the algorithmic criteria from Chapter 7 was that all algorithms be rapid, one must consider the consequence of detecting discontinuities after the interpolation of the surfaces. There are two main reasons for the explicit detection of discontinuities. One is that such an explicit representation of this information will allow higher level processes, such as recognition, or axis extraction, to operate more easily, since the process serves to make implicit information explicit. However, a second reason is to create more accurate surface representations, by cancelling out the type of effect illustrated in Figure 9.1. If the process used to isolate discontinuities takes place after interpolation, and if the interpolation process requires the discontinuities to improve the interpolated surface approximation, one must propose a two pass interpolater. Thus, the major question is whether such a two pass process will affect our constraint of rapid algorithms. Fortunately, the answer is no. This follows from two facts. The first is that the number of iterations required to achieve an acceptable surface approximation is very small, on the order of 10. Thus, even if the interpolation process is required to run twice, the number of iterations needed is still small. Further, since the surface approximation obtained without explicitly accounting for the discontinuities is very close to the limiting surface except in the areas of the discontinuities (that is, any effects of the discontinuities are quickly damped out as one moves across the surface), the initial starting position for the second pass of the interpolation algorithm is very close to the limiting surface, and only a few iterations will be needed to refine the surface approximation.

9.1.3 Interpolation Over Occluded Regions

Even though occluded regions of the image can only be viewed from one eye, the human system still associates a depth value with these regions. This has an interesting implication for the interpolation algorithm. For most occluded regions, the only depth information available is at the edges of the occluded region. Psychophysical experiments have shown that the occluded region is always perceived at the depth of the lower surface. Thus, in figure 9.2, the occluded region would be perceived at the level of the lower surface. Note that this is consistent with the physics of the situation, since if the occluded region were perceived at the level of the upper surface, then it should in fact be visible to the right eye, and this is not the case.

This observation suggests that when an occlusion is detected, it is explicitly located along the occluding boundary corresponding to the edge of the nearer object. This allows the occluded region itself to be associated with the lower surface, and the interpolation algorithm will fill in surface values for the occluded region from this lower surface.

This raises an interesting psychophysical prediction. The psychophysical literature has examined the case of planar surfaces and their occlusions, as in Figure 9.2. If the interpolation method developed here is given an explicit discontinuity along one edge of the occluded region, it will correctly fill in the region as an extension of the lower plane. Of more interest is the case in which the occluded region is not planar. For example, consider a cylindrical object, such as that of Figure 6.4. If the interpolation algorithm is given this type of input, it will fill in the occluded portions of the surfaces as a smooth continuation of the curved cylinder. If the interpolation algorithm is correct, then this predicts that the surface perception for human observers in this situation should also be that of a smooth cylinder. This has not yet be tested psychophysically.

9.2 Noise Removal

Although in general the Marr-Poggio stereo algorithm is very good at matching zero-crossings correctly (especially for random dot patterns), incorrect disparity values may sometimes be assigned to regions of the image. Since the surface interpolator explicitly attempts to fit a surface through all the disparity points, such noise points can affect the shape of the surface approximation. Indeed, the effect of these noise points can spread over a noticable portion of the surface, before the nearby disparity values can damp out its effect. Thus, it would be preferable to remove these noise points,

or at least neutralize their effect on the approximated surface shape. One possibility is that if a two pass interpolator is used, as suggested in the previous section, the detection of surface discontinuities will isolate such noise points from the rest of the surface, and the second pass of the interpolator will adjust the surface approximation to remove the influence of the noise points on the first pass approximation. Certainly this will be true for noise points with disparity values far removed from the correct values. And for noise points whose disparity values are only slightly different from the correct surface disparities, the difference does not really matter. However, the final result would be that the noise points, while being isolated from the rest of the correct surface, would still remain in the final surface description. It would be preferable to completely remove such points.

Is it possible to identify and remove such noise points from the disparity map? If the noise points are isolated spatially, then it is possible to identify such points as undesirable. This follows from the form of the primal sketch operators. The case to consider is that in which one must distinguish between a set of noise points in a disparity map and a small object separated in depth from the rest of the scene. For the small object, there is a minimum size of zero-crossing contour which the operator will yield about the object. Hence, if the number of zero-crossing points which differ significantly from their neighbours is less than this minimum, one may conclude that the points are noise, and thus remove them. This will result in an improved surface approximation.

9.3 Acuity

It can be seen from the example of the interpolated coffee jar in Figure 8.12, that the interpolated surface contains a bumpy quality which clearly is not consistent with the original object. How can this be explained? The cause of the effect is the fact that the disparity values are specified to within only a pixel. This yields a fairly coarse disparity map which results in the bumps observed in the interpolated coffee jar. Hence, one method of removing the bumps would be to improve the accuracy of the disparities obtained by the algorithm. Note that some improvement in disparity accuracy is necessary if the algorithm is to be consistent with the human system, since a pixel corresponds to roughly 27 seconds of arc, while humans are capable of stereo acuity to a resolution of 2 — 10 seconds (Howard, 1919; Woodburne, 1934; Berry, 1948; Tyler, 1977).

In order to account for finer disparity values, it is necessary to localize the zero-crossing to a better accuracy than has been done so far. Since the convolution values are only specified at each pixel,

one method for more accurately specifying the zero-crossing positions is to interpolate between the known convolution values (Crick, Marr and Poggio 1980, Marr, Poggio and Hildreth 1979, Hildreth 1980). Perhaps the simplest method is to rely on the observation of Hildreth that for most cases, even a simple linear interpolation will give extremely accurate localization of the zero-crossings. The addition of finer resolution depth information may improve the performance of the algorithm.

This example also raises a question of scale. Depending on the application of the surface specification, different amounts of resolution may be required. For example, if the ulimate goal of the surface specification is to obtain a rough idea of the position and shape of the surfaces in a scene, the spatial resolution at which surface information must be made explicit may not be critical. In this case, the known data from the stereo algorithm may be sampled at a coarser resolution, before the interpolation takes place. This will result in a smoother surface approximation, although some of the finer detail may be removed. An example of the coffee jar interpolated at a coarser spatial resolution is shown in Figure 9.3. It can be seen that the overall surface is smoother than in Figure 8.12. Further, although the reconstructed surface is less exact in terms of fine variation of the surface shape, the overall shape of the bottle is still preserved in this interpolation.

9.4 Retinal Mappings

The operators used to create the symbolic image descriptions matched by the Marr-Poggio stereo matcher were derived in part from evidence about the human visual system (Wilson and Bergen 1979). One aspect of this evidence is that the size of each operator scales with eccentricity, being larger in the periphery than in the fovea. However, in the development of the stereo implementation, the operations were assumed to be uniform for each size mask. A modification of the implementation, to be more consistent with the human evidence, is now considered.

There are two possible modifications that could be made. First, the size of the convolution masks, as well as the size of the matching neighbourhoods, should explicitly increase with eccentricity. For a parallel implementation, such as the human brain, there is no great difficulty in imposing this variance in size on the individual processors. However, for a serial computer implementation, such a variation in operator size introduces several practical difficulties.

There is an alternative implementation (Schwartz, 1977) which avoids these difficulties. Suppose that the retinal images are transformed, mapping them to a representation with coordinate axes given

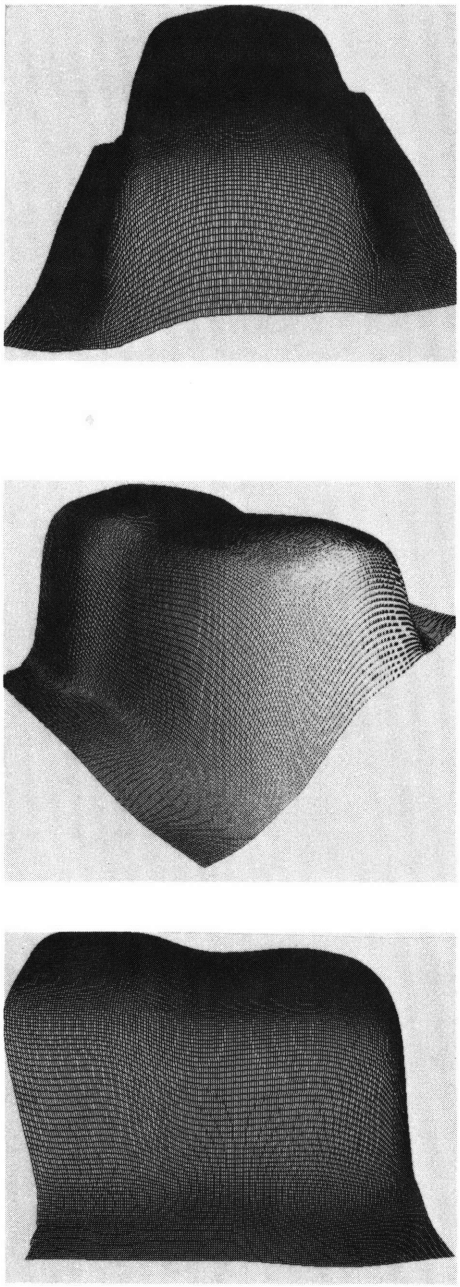


Figure 9.3. Acuity Surfaces. The coffee jar of Figure 8.12 is illustrated. In this case, the spatial resolution of the interpolation has be reduced, resulting in a smoother overall surface, although some fine surface detail is lost.

by orientation relative to the fovea, θ, along one axis, and log(radius) along the other axis. Such a mapping has the feature that in the new representation, the application of a uniformly sized operator over the entire transformed image is equivalent to the application of a varying sized operator over the entire original image. Since the variation in mask size with eccentricity affects both the extraction of the convolutions, and the size of the matching neighbourhood, it is necessary to perform the stereo matching in this transformed representation, before mapping back to a normal coordinate system. However, note that in this transformed representation, direction is only preserved locally. In the original implementation, given the location of a zero-crossing in one image, the location of the matching zero-crossing was performed by searching a horizontal neighbourhood about the corresponding location in the other image. In the transformed image, one cannot simply apply the same matching algorithm since the horizontal direction in the original image does not correspond to the horizontal direction in the transformed image. Rather, for each value of θ, the orientation of the matching neighbourhood must be changed from horizontal to something else, depending on the exact value of θ. The matching algorithm can be modified to account for this.

Once the matching has been performed in this transformed representation, the resulting disparity values can be transformed back to the original retinal coordinate system. This method has not yet been tested.

9.5 Multiple Representations

One final point concerns the question of the resolution of the depth representation. In the previous sections, we have outlined a method which creates a single disparity map, interpolates it, and then finds discontinuities and inflections in depth using a range of $\nabla^2 G$ operators. Rather than combining the depth information from the multiple stereo channels into a single representation, only to later create multiple descriptions of the inflections in depth, at varying resolutions, one could consider maintaining separate levels of description. In this manner, one would interpolate the disparity information obtained at each scale, using the output of each stereo channel. Then the inflections in depth could be detected for each such interpolated surface, by an appropriate $\nabla^2 G$ operator. This would result in a representation of the visible surfaces in which multiple levels of scale are explicitly maintained. This may prove to be useful to higher level visual modules which require depth descriptions at different resolutions.

CHAPTER 10

REFERENCES

Ahlberg, J.H., Nilson, E.N. and Walsh, J.L. *The Theory of Splines and Their Applications*, Academic Press, New York, 1967.

Ahlberg, J.H., Nilson, E.N. and Walsh, J.L. "Extremal, orthogonality, and convergence properties of multidimensional splines," *J. Math. Anal. Appl.* 11 (1965), 27-48.

Akima, H. "A method of bivariate interpolation and smooth surface fitting for irregualrly distributed data points," *ACM Trans. on Math. Software* 4, 2 (1978), 148-159.

Arrow, K.J. and Hurwicz, L. "Reduction of constrained maxima to saddle-point problems," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability ed. J. Neyman. Berkeley and Los Angeles: University of California Press*, 1956, 1-20.

Arrow, K.J., Hurwicz, L., and Uzawa, H. *Studies in Linear and Non-linear Programming*, Stanford University Press, Stanford, Ca., 1958.

Arthur, D.W. "Multivariate spline functions, I. Construction, properties, and computation," *J. Approximation Theory* 12 (1974), 396-411.

Arthur, D.W. "Multivariate spline functions, II. Best error bounds," *J. Approximation Theory* 15 (1975), 1-10.

Atteia, M. Etude de certains noyaux et theorie des fonctions (spline) en Analysis Numerique, Univ. of Grenoble, 1966a.

Atteia, M. "Existence et determination des fonctions splines a plusieurs variables," *C. R. Acad. Sci. Paris* 262 (1966b), 575-578.

Atteia, M. "Fonctions (spline) et noyaux reproduisants D'Aronszajn-Bergman," *Rev. Francaise*

Informat, Recherche Operationnelle 4E Annee R-3 (1970), 31-43.

Aziz, A.K. (ed) *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*, Academic Press, New York, 1972.

Babuska, I. "Approximation by hill functions," *Comment. Math. Univ. Carolinae* 11, 4 (1970), 787-811.

Bajcsy, R. and Lieberman, L. "Texture gradients as a depth cue," *Computer Graphics and Image Processing* 5 (1976), 52-67.

Barnhill, R.E. Birkhoff, G. and Gordon, W.J. "Smooth interpolation in triangles," *J. Approximation Theory* 8 (1973), 114-128.

Barnhill, R.E. and Gregory, J.A. Blending function interpolation to boundary data on triangles, Department of Math. Brunel Univ.,, TR 14, 1972.

Barnhill, R.E. and Gregory, J.A. "Compatible smooth interpolation in triangles," *J. Approximation Theory* 15 (1975), 214-225.

Barnhill, R.E. and Mansfield, L. "Error bounds for smooth interpolation in triangles," *J. Approximation Theory* 11 (1974), 306-318.

Barnhill, R.E. and Riesenfeld, R.F. *Computer Aided Geometric Design*, Academic Press, New York, 1974.

Berry, R.N. "Quantitative relations among vernier, real depth, and stereoscopic depth acuities," *J. Exp. Psychol.* 38 (1948), 708-721.

Birkhoff, G. and de Boor, C. "Piecewise polynomial interpolation and approximation," *Approximation of Functions, H.L. Garabedian, ed., Elsevier, Amsterdam* (1965), 164-190.

Birkhoff, G. and Garabedian, H. "Smooth surface interpolation," *J. Math. Phys.* 39 (1960), 258-268.

Birkhoff, G. and Mansfield, L. "Compatible triangular finite elements," *J. Math. Anal. Appl.* 47 (1974), 531-553.

Birkhoff, G., Schultz, M.H. and Varga, R.S. "Piecewise Hermite interpolation in one and two variables with applications to partial differential equations," *Numer. Math.* 11 (1968), 232-256.

de Boor, C. "Bicubic spline interpolation," *J. Math. and Phys.* 41 (1962), 212-218.

de Boor, C. "On calculating with B-splines," *J. Approximation Theory* 6 (1972), 50-62.

de Boor, C. (ed) *Mathematical Aspects of Finite Elements in Partial Differential Equations*, Academic Press, New York, 1975.

de Boor, C. and Fix, G.J. "Spline approximation by quasi-interpolants," *J. Approximation Theory* 8

(1975), 19-45.

Braddick, O. Multiple matching in stereopsis, unpublished MIT report, 1978.

Brady, J.M., Grimson, W.E.L. and Langridge, D.J. "Shape encoding and subjective contours," (1980), to appear.

Bramble, J. and Zlamal, M. "Triangular elements in the finite element method," *Math. Comp.* **24** (1970), 809-820.

Brigner, W.I. and Gallagher, M.B. "Subjective contour: Apparent depth or simultaneous brightness contrast?," *Perceptual and Motor Skills* **38** (1974), 1047-1053.

Cadwell, J.H. and Williams, D.E. "Some orthogonal methods of curve and surface fitting," *Comput. J.* **4** (1961).

Campbell, F.W. and Robson, J.G. "Applications of Fourier analysis to the visibility of gratings," *J. Physiol. (Lond.)* **197** (1968), 551-556.

Cavendish, J. C. "Automatic triangulation of arbitrary planar domains for the finite element method," *Int. J. for Numer. Methods in Engineering* **8** (1974), 697-696.

Ciarlet, P.G. and Raviart, P.A. "Interpolation de Lagrange dans $\Re^n$," *C. R. Acad. Sci. Paris Ser.* **A273** (1971), 578-581.

Ciarlet, P.G. and Raviart, P.A. "General Lagrange and Hermite interpolation in $\Re^N$ with applications to finite element methods.," *Arch. Rational Mech. Anal.* **46** (1972a), 177-199.

Ciarlet, P.G. and Raviart, P.A. "Interpolation de Lagrange sur des elements finis courbes dans $\Re^n$," *C. R. Acad. Sci. Paris Ser.* **A274** (1972b), 640-643.

Ciarlet, P.G. and Raviart, P.A. "Interpolation theory over curved elements with applications to finite element methods," *Comp. Meth. Appl. Mech. Eng.* **1** (1972c), 217-249.

Clarke, P. G. H., Donaldson, I. M. L. and Whitteridge, D. "Binocular mechanisms in cortical areas I and II of the sheep," *J. Physiol. (Lond.)* **256** (1979), 509-526.

Coons, S.A. Surfaces for computer-aided design of space forms, MIT Project MAC, Tr 41, 1967.

Coren, S. "Subjective contours and apparent depth," *Psychological Review* **79**, 4 (1972), 359-367.

Coren, S. and Theodor, L.H. "Subjective contour: the inadequacy of brightness contrast as an explanation," *Bulletin of the Psychonomic Society* **6** (1975), 87-89.

Cornsweet, T.N. *Visual Perception*, Academic Press, New York, NY., 1970.

Crain, I.K. and Bhattacharyya, B.K. "Treatment of non-equispaced two-dimensional data with a digital computer," *Geoexploration* **5** (1967), 173-194.

REFERENCES

Crick, F.H.C., Marr, D. and Poggio, T. "An information processing approach to understanding the visual cortex," *in The Cerebral Cortex, N.R.P.* (1980).

Davis, J.C. *Interpolation and Approximation*, Blaisdell, New York, NY., 1963.

Delvos, F.J. "On surface interpolation," *J. Approximation Theory* 15 (1975), 129-137.

Delvos, F.J. and Kosters, H.W. "On the variational characterization of bivariate interpolation methods," *Math. Z.* 145 (1975), 129-137.

Delvos, F.J. and Schempp, W. "Sard's method and the theory of spline systems," *J. Approximation Theory* 14 (1975), 230-243.

Delvos, F.J. and Schlosser, K.H. "Das tensorproduktschema von spline systemen," *in Splinefunktionen, K. Bohmer, G. Meinardus, and W. Schempp, (eds.), Bibliographisches Institut, Zurich* (1974), 59-74.

Dev, P. "Perception of depth surfaces in random-dot stereograms: a neural model," *Int. J. Man-Machine Studies* 7 (1975), 511-528.

Duchon, J. Fonctions-spline du type plaque mince en dimension 2, Univ. of Grenoble, Rpt. 231, 1975.

Duchon, J. Fonctions-spline a energie invariante par rotation, Univ. of Grenoble, Rpt. 27, 1976.

Earnshaw, J.L., and Yuille, I.M. "A method of fitting parametric equations for curves and surfaces to sets of points defining them approximately," *Computer Aided Design* 3 (1971), 19-22.

Effroymson, M.A. "Multiple regression analysis," *in Mathematical Methods for Digital Computers, A. Ralston and H. Wilff (eds.) Wiley, New York* (1960), 191-203.

Fender, D. and Julesz, B. "Extension of Panum's fusional area in binocularly stabilized vision," *J. Opt. Soc. Am.* 57 (1967), 819-830.

Ferguson, J. "Multivariable curve interpolation," *J. Assoc. Comput. Mach.* 11 (1964), 221-228.

Fisher, S.D. and Jerome, J.W. "Elliptic variational problems in L^2 and L^∞," *Indiana J. Math.* 23 (1974), 685-698.

Fisher, S.D. and Jerome, J.W. "Spline solutions to L^1 extremal problems in one and several variables," *J. Approximation Theory* 13 (1975), 73-83.

Forrest, A.R. "On Coons and other methods for the representation of curved surfaces," *Computer Graphics and Image Processing* 1 (1972a), 341-359.

Forrest, A.R. "Interactive interpolation and approximation by Bezier polynomials," *Computer Journal* 15, 1 (1972b), 71-79.

Frisby, J. P. and Clatworth, J. L. "Learning to see complex random-dot stereograms," *Perception* 4

(1975), 173-178.

Fix, G. and Strang, G. "Fourier analysis of the finite element method in Ritz-Galerkin theory," *Studies in Appl. Math.* 48 (1969), 265-273.

Frederickson, P.O. "Quasi-interpolation, extrapolation, and approximation on the plane," *Conf. Numerical Maths.*, Winnipeg, 1971, 159-167.

Goel, J.J. "Construction of basic functions for numerical utilization of Ritz's method," *Numer. Math.* 12 (1968), 435-447.

Gregory, R.L. "Cognitive contours," *Nature* 238, 5358 (1972), 51-52.

Gregory, R.L. and Harris, J.P. "Illusory contours and stereo depth," *Perception and Psychopysics* 15, 3 (1974), 411-416.

Grimson, W.E.L. A computer implementation of a theory of human stereo vision, MIT Artificial Intelligence Laboratory, Memo 565, 1980.

Grimson, W.E.L. and Marr, D. "A computer implementation of a theory of human stereo vision," *Proceedings: Image Understanding Workshop*, Palo Alto, Cal., 1979, 41-47.

Guenther, R.B. and Roetman, E.L. "Some observations on interpolation in higher dimensions," *Math. Comp.* 24 (1970), 517-522.

Hall, C.A. "Bicubic interpolation over triangles," *J. Math. Mech.* 19 (1969), 1-11.

Haussman, W. "On multivariate spline systems," *J. Approximation Theory* 11 (1974), 285-305.

Haussman, W. and Munch, H.J. "On construction of multivariate spline systems," *in Approximation Theory, G.G. Lorentz, (ed.), Academic Press, New York* (1973), 373-378.

Hayes, J.G. and Halliday, J. "The least squares fitting of cubic spline surfaces to general data sets," *J. Inst. Maths. Applics.* 14 (1974), 89-103.

Hildreth, E.C. Implementation of a theory of edge detection, M. Sc. Thesis, Department of Dept. Of Computer Science and Electrical Engineering, Massachusetts Institute of Technology, 1980.

Hirai, Y. and Fukushima, K. "An inference upon the neural network finding binocular correspondence," *Trans. IECE* J59-d (1976), 133-140.

Horn, B.K.P. Focusing, MIT Artificial Intelligence Laboratory, Memo 160, 1968.

Horn, B.K.P. Shape from shading: a method for obtaining the shape of a smooth spaque object from one view, MIT Project MAC Technical Report, MAC TR-79, 1970.

Horn, B.K.P. "Obtaining shape from shading information," *The Psychology of Computer Vision, P.H. Winston (ed), McGraw-Hill* (1975), 115-155.

REFERENCES

Horn, B.K.P. "Understanding image intensities," *Artificial Intelligence* 8 (1977), 201-231.

Horn, B.K.P. and Bachman, B.L. Using synthetic images to register real images with surface models, MIT Artificial Intelligence Laboratory, Memo 437, 1977.

Horn, B.K.P. and Sjoberg, R.W. Calculating the reflectance map, MIT Artificial Intelligence Laboratory, Memo 498, 1978.

Hosaka, M. "Theory of curve and surface synthesis and their smooth fitting," *Information Processing in Japan* 9 (1969), 60-68.

Howard, J.H. "A test for the judgement of distance," *Am. J. Ophthal.* 2 (1919), 656-675.

Huffman, D.A. "Impossible objects as nonsense sentences," *Machine Intelligence 6, B. Meltzer and D. Michie (eds), Edinburgh University Press* (1971), 295-393.

Ikeuchi, K. Numerical shape from shading and occluding contours in a single view, MIT Artificial Intelligence Laboratory, Memo 566, 1979.

Ikeuchi, K. Shape from regular patterns (an example of constraint propagation in vision), MIT Artificial Intelligence Laboratory, Memo 567, 1980.

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

Julesz, B. "Towards the automation of binocular depth perception (AUTOMAP-1)," *Bell System Tech. J.* (1963).

Julesz, B. *Foundations of Cyclopean Perception*, The University of Chicago Press, Chicago, 1971.

Julesz, B. and Chang, J.J. "Interaction between pools of binocular disparity detectors tuned to different disparities," *Biol. Cybernetics* 22 (1976), 107-120.

Julesz, B. and Miller, J.E. "Independent spatial-frequence-tuned channels in binocular fusion and rivalry," *Perception* 4 (1975), 125-143.

Kanizsa, G. "Subjective contours," *Scientific American* 234, 4 (1976), 48-52.

Kaufman, L. "On the nature of binocular disparity," *Am. J. Psychol.* 77 (1964), 393-40.

Kender, J.R. "Shape from texture: a brief overview and a new aggregation transform," *Proceedings of the ARPA Image Understanding Workshop*, 1978.

Kender, J.R. Shape from texture, Ph.D. Thesis, Carnegie-Mellon Univ., 1979.

Knight, T.F., Moon, D.A., Holloway, J. and Steele, G.L. CADR, MIT AI Lab, Memo 528, 1979.

Koelling, M.E.V. and Whitten, E.H.T. "Fortran IV program for spline surface interpolation and

contour map production," *Geocomprograms* 9 (1973), 1-12.

Kuhn, H.W. and Tucker, A.W. "Nonlinear Programming," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability ed. J. Neyman. Berkeley and Los Angeles: University of California Press*, Berkeley, Cal., 1951, 481-492.

Kunz, K.S. *Numerical Analysis*, McGraw-Hill, New York, 1957.

Laurent, P.J. *Approximation et Optimisation*, Hermann, Paris, 1972.

Lawson, C. Generation of a triangular grid with applications to contour plotting, JPL, T.M. 299, 1972.

Leadbetter, M.R. "On the distributions of times between events in a stationary stream of events.," *J. R. Statist. Soc. B* 31 (1969), 295-302.

Leipnik, R. "The extended entropy uncertainty principle," *Inform. and Control* 3 (1960), 18-25.

Longuet-Higgins, M.S. "The distribution of intervals between zeros of a stationary random function.," *Phil. Trans. R. Soc. Lond. A* 254 (1962), 557-599.

Luenberger, D.G. *Introduction to Linear and Nonlinear Programming*, Addison-Wesley Publishing Co., Reading, Mass., 1973.

Lyche, T., and Schumaker, L. "Local spline approximation methods," *J. Approximation Theory* 15 (1975), 294-325.

Mackworth, A.K. "Interpreting pictures of polyhedral scenes," *Artificial Intelligence* 4, 2, 121-137.

Mansfield, L.E. "On the optimal approximation of linear functionals in spaces of bivariate functions," *SIAM J. Numer. Anal.* 8 (1971), 115-126.

Mansfield, L.E. "On the variational characterization and convergence of bivariate splines," *Numer. Math.* 20 (1972a), 99-114.

Mansfield, L.E. "Optimal approximation and error bounds in spaces of bivariate functions," *J. Approximation Theory* 5 (1972b), 77-96.

Mansfield, L.E. "On the variational approach to defining splines on L-shaped regions," *J. Approximation Theory* 12 (1974), 99-112.

Mansfield, L.E. "Interpolation to boundary data in triangles with application to compatible finite elements," *Approximation Theory II, ed. G.G. Lorentz, C.K. Chui, L.L. Schumaker, Academic Press* (1976).

Marr, D. "A theory of cerebellar cortex," *J. Physiol.* 202 (1969), 437-470.

Marr, D. A note on the computation of binocular disparity in a symbolic, low-level visual processor, MIT AI Lab, Memo 327, 1974.

REFERENCES

Marr, D. Artificial Intelligence – a personal view, MIT AI Lab, Memo 355, 1976a.

Marr, D. "Early processing of visual information," *Philosophical Transactions of the Royal Society of London* 275, 942 (1976b), 483-534.

Marr, D. "Representing visual information," *AAS 143rd Annual Meeting, Symposium on Some Mathematical Questions in Biology, February 1977. Published in Lectures in the Life Sciences* 10 (1978), 101-180.

Marr, D. *VISION: A computational investigation in the human representation and processing of visual information*, W.J. Freeman, San Francisco, 1980.

Marr, D. and Hildreth, E.C. "Theory of edge detection," *Proc. R. Soc. Lond. B* 207 (1980), 187-217.

Marr, D. and Nishihara, H.K. "Representation and recognition of the spatial organization of three-dimensional shapes," *Proc. R. Soc. Lond. B.* (1978).

Marr, D. and Poggio, T. "From understanding computation to understanding neural circuitry," *Neuroscience Research Program Bulletin* 15, 3 (1977a), 470-488.

Marr, D. and Poggio, T. "Cooperative computation of stereo disparity," *Science* 195 (1977b), 283-287.

Marr, D. and Poggio, T. "A theory of human stereo vision," *Proc. R. Soc. Lond. B* 204 (1979), 301-328.

Marr, D., Poggio, T. and Hildreth, E. "The smallest channel in early human vision," *JOSA* (1979), in press.

Mayhew, J. E. W. and Frisby, J. P. "Rivalrous texture stereograms," *Nature* 264 (1976), 53-56.

Mayhew, J. E. W. and Frisby, J. P. "Stereopsis masking in humans is not orientationally tuned," *Perception* 7 (1978), 431-436.

Mayhew, J. E. W. and Frisby, J. P. "Convergent disparity discriminations in narrow-band-filtered random-dot stereograms," *Vision Res.* 19 (1979a), 63-71.

Mersereau, R.M. and Oppenheim, A.V. "Digital reconstruction of multi-dimensional signals from their projections," *Proc. IEEE* 62 (1974), 1319-1338.

Mitchell, A.R. "Introduction to the mathematics of finite elements," *in The Mathematics of Finite Elements, Academic Press, London* (1973), 37-58.

Mitchell, A.R. and Phillips, G.M. "Construction of basis functions in the finite element method," *Nord. Tidskr. Inform. BIT* 12 (1972), 81-89.

Munteanu, M.J. "Generalized smoothing spline functions for operators," *SIAM J. Numer. Anal.* 10 (1973), 28-34.

Munteanu, M.J. "On the construction of multidimensional splines," *in Spline Functions and Approximation Theory, A. Meir and A. Sharma, (eds.), ISNM 21, Birkhauser* (1973), 235-265.

Munteanu, M.J. and Schumaker, L.L. "Some multidimensional spline approximation methods," *J. Approximation Theory* 10 (1974), 23-40.

Nelson, J.I "Globality and stereoscopic fusion in binocular vision," *J. Theor. Biol.* 49 (1975), 1-88.

Nicoliades, R.A. "On the class of finite elements generated by Lagrange interpolation," *SIAM J. Numer. Anal.* 9 (1972), 435-445.

Nicoliades, R.A. "On the class of finite elements generated by Lagrange interpolation, II," *SIAM J. Numer. Anal.* 10 (1973), 182-189.

Nielson, G.M. "Bivariate spline functions and the approximation of linear functionals," *Numer. Math.* 21 (1973), 138-160.

Nielson, G.M. Surface approximation and data smoothing using generalized spline functions, Univ. of Utah, 1970.

O'Brien, B. "Vision and resolution in the central retina," *J. Opt. Soc. Am.* 41 (1951), 882-894.

Pavlidis, T. *Structural Pattern Recognition*, Springer-Verlag, Berlin, 1977.

Pelto, C., Elkins, T. and Boyd, H. "Automatic contouring of irregularly spaced data," *Geophysics* 33 (1968), 424-430.

Pivovarova, N.B., and Puknacheva, T.B. "Smoothing experimental data with local splines," *Sem. Num. Methods of Applied Mathematics*, Novosibirsk, 1975.

Poeppelmeir, C. A Booelan sum interpolation scheme to random data for computer aided geometric design, Univ. Of Utah, 1975.

Poggio, G.F. and Fischer, B. "Binocular interaction and depth sensitivity of striate and prestriate cortical neurons of the behaving rhesus monkey," *J. Neurophysiol.* (1978).

Prenter, P.M. *Splines and Variational Methods*, Wiley-Interscience, New York, NY., 1975.

Rabinowitz, P. "Applications of linear programming to numerical analysis," *SIAM Rev.* 10 (1968), 121-159.

Rabinowitz, P. "Mathematical programming and approximation," *Approximation Theory, A. Talbot (ed), Academic Press, London* (1970), 271-231.

Rashbass, C. and Westheimer, G. "Disjunctive eye movements," *J. Physiol. Lond.* 159 (1961), 339-360.

Rice, S.O. "Mathematical analysis of random noise.," *Bell Syst. Tech. J.* 24 (1945), 46-156.

REFERENCES

Richards, W. "Stereopsis and stereoblindness," *Exp. Brain Res.* 10 (1970), 380-388.

Richards, W. "Anomalous steroscopic depth perception," *J. Opt. Soc. Amer.* 61 (1971), 410-414.

Richards, W. "Stereopsis with and without monocular cues," *Vision Res.* (1977).

Richards, W.A. and Regan, D. "A stereo field map with implications for disparity processing," *Investigative Ophthalmology* 12 (1973), 904-909.

Riggs, L. A. and Niehl, E. W. "Eye movements recorded during convergence and divergence," *J. Opt. Soc. Am.* 50 (1960), 913-920.

Ritter, K. "Two-dimensional splines and their extremal properties," *Z. Agnew. Math. Mech.* 49 (1969), 597-608.

Ritter, K. "Two-dimensional spline functions and best approximations of linear functionals," *J. Approximation Theory* 3 (1970), 352-368.

Rosen, J.B. "Minimum error bounds for multidimensional spline approximation," *J. Comput. System Sci.* 5 (1971), 430-452.

Rudin, W. *Functional Analysis*, McGraw-Hill Book Company, New York, 1973.

Sard, A. "Approximation based on nonscalar observations," *J. Approximation Theory* 8 (1973), 315-334.

Sard, A. "Instances of generalized splines," *in Splinefunctionen, K. Bohmer, G. Meinardus, W. Schempp, (eds.), Bibliographisches Institut, Mannheim* (1974), 215-241.

Saye, A. and Frisby, J. P. "The role of monocularly conspicuous features in facilitating stereopsis from random-dot stereograms," *Perception* 4 (1975), 159-171.

Schaback, R. "Konstruktion und algebraische eigenschaften von M-spline interpolierenden," *Numer. Math.* 21 (1973), 166-180.

Schaback, R. "Kollakation mit mehrdimensionalen spline-funktionen," *Numerische Behandlung nicht-linearer Integrodifferential und Differentialgleichungen, R. Ansorge and W. Tornig, eds., Lecture Notes 395., Springer-Verlag, Heidelberg* (1974), 291-300.

Schultz, M.H. "L-infinity-multivariate approximation theory," *SIAM J. Numer. Anal.* 6 (1969a), 161-183.

Schultz, M.H. "Multivariate L-spline interpolation," *J. Approximation Theory* 2 (1969b), 127-135.

Schumaker, L. L. "Fitting surfaces to scattered data," *Approximation Theory II, ed. G.G. Lorentz, C.K. Chui, L.L. Schumaker, Academic Press* (1976), 203-268.

Schumann, F. "Einige beobachtungen uber die zusammenfassung von gesichtseindrucken zu ein-

heiten," *Psychologische Studien* 1 (1904), 1-32.

Schwartz, E.L. "Afferent geometry in the primate visual cortex and the generation of neuronal trigger features," *Biol. Cyb.* 28 (1977), 1-14.

Shepard, D. "A two-dimensional interpolation function for irregularly spaced data," *Proc. 1968 ACM Nat. Conf.*, 1968, 517-524.

Silver, W. Determining shape and reflectance using multiple images, M.Sc. Thesis, Massachusetts Institute of Technology, 1980.

Spath, H. "Algorithmus 10-zweidimensionale glatte interpolation," *Computing* 4 (1969), 178-182.

Spath, H. "Two-dimensional exponential splines," *Computing* 7 (1971), 364-369.

Sperling, G. "Binocular vision: a physical and a neural theory," *Am. J. Psychol.* 83 (1970), 461-534.

Steffensen, J.F. *Interpolation*, Williams and Wilkins, Baltimore, Md., 1927.

Stevens, K.A. Surface perception from local analysis of texture and contour, MIT AI Laboratory, TR 512, 1979.

Strang, G. and Fix, J. *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, N.J., 1973.

Sugie, N. and Suwa, M. "A scheme for binocular depth perception suggested by neurophysiological evidence," *Biol. Cybernetics* 26 (1977), 1-15.

Thacher, H.C.,Jr. "Derivation of interpolation formulas in several independent variables," *Annal. of the New York Acad. of Sciences* 86 (1960), 758-775.

Thacher, H.C.,Jr. and Milne, W.E. "Interpolation in several variables," *SIAM J. Appl. Math.* 8 (1960), 33-42.

Theilheimer, F. and Starkweather, W. "the fairing of ship lines on a high-speed computer," *Math. Comp.* 15 (1961), 338-355.

Thomann, J. Determination et construction de fonctions spline a deux variables definies sur un domaine rectangulaire ou circulaire, Univ. of Lille, 1970a.

Thomann, J. "Obtention de la fonction spline d'interpolation a deux variables sur un domaine rectangulaire ou circulaire," *Proc. Algol en analyse Numerique II*, Centre National de la Recherche Scientifique, Paris, 1970b, 83-94.

Tyler, C.W. "Spatial limitations of human stereoscopic vision," *SPIE* 120 (1977).

Ullman, S. "Filling in the gaps: The shape of subjective contours and a model for their generation," *Biol. Cyb.* 25 (1976), 1-6.

REFERENCES

Ullman, S. *The Interpretation of Visual Motion* , MIT Press, Cambridge, Mass., 1979a.

Ullman, S. "Relaxation and constrained optimization by local processes," *Computer Graphics and Image Processing* 10 (1979b), 115-125.

Whaples, G.W. "A note on degree N independence," *SIAM J. Appl. Math.* 6 (1958), 300-301.

Whiteman, J.R. (ed) *The Mathematics of Finite Elements* , Academic Press, London, 1973.

Whitten, E.H.T. "Orthogonal polynomial trend surfaces for irregularly spaced data," *The Mathematics of Finite Elements* 2 (1970), 141-152.

Whitten, E.H.T. "The use of multi-dimensional cubic spline functions for regression and smoothing," *Austral. Comput. J.* 3 (1971), 81-88.

Whitten, E.H.T. "More on irregularly spaced data and orthogonal polynomial trend surfaces," *Int. Assoc. Math. Geol. J.* 4 (1972), 83.

Whitten, E.H.T. and Koelling, M.E.V. Computation of bicubic-spline surfaces for irregularly-spaced data, Northwestern Univ. Geology Dept., T.R. 3, 1974.

Wilson, H.R. and Bergen J.R. "A four mechanism model for threshold spatial vision," *Vision Res.* 19 (1979), 19-32.

Wilson, H.R. and Giese, S.C. "threshold visibility of frequency grating patterns," *Vision Res.* 17 (1977), 1177-1190.

Woodburne, L.S. "The effect of a constant visual angle upon the binocular discrimination of depth differences," *Am. J. Psychol.* 46 (1934), 273-286.

Woodham, R.J. Reflectance map techniques for analyzing surface defects in metal castings, MIT AI Lab, TR 457, 1978.

Zavialov, Yu. S. "Interpolating L-splines in several variables," *Mat. Zametki* 14 (1973), 11-20.

Zavialov, Yu. S. "Smoothing L-splines in several variables," *Mat. Zametki* 15 (1974a), 371-379.

Zavialov, Yu. S. "L-spline functions of several variables," *Soviet Math. Dokl.* 15 (1974b), 338-341.

Zenisek, A. "Interpolation polynomials on the triangle," *Numer. Math.* 15 (1970), 283-296.

Zienkiewicz, O.C. "The finite element method: from intuition to generality," *Appl. Mech. Rev.* 23 (1970), 249-256.

Zlamal, M. "A finite element procedure of the second order of accuracy," *Numer. Math.* 14 (1970), 394-402.

Zlamal, M. "On the finite element method," *Numer. Math.* 12 (1968), 394-409.

Zlamal, M. "Curved elements in the finite element method, I," *SIAM J. Numer. Anal.* 10 (1973), 229-240.

Zlamal, M. "Curved elements in the finite element method, II," *SIAM J. Numer. Anal.* 11 (1974), 347-362.