CHAPTER 1

THEORETICAL PRELIMINARIES

Although our world has three spatial dimensions, the projection of light rays onto the retina presents our visual system with an image of the world that is inherently two-dimensional. We must use such images to physically interact with this three-dimensional world, even in situations new to us, or with objects unknown to us. That we accomplish this task easily implies that one of the functions of the human visual system is to reconstruct a three-dimensional representation of the world from its two-dimensional projection onto our eyes. The study described in this book constitutes a computational theory of this process: *creating representations of surface shape from their images.* An example of a process which the human visual system uses to construct this three-dimensional representation is stereopsis, which refers to the use of the two viewpoints, provided by the left and right eyes, to establish the depth of surfaces around us. Illustrated in Figure 1.1 is a stereogram; the top two images represent two different viewpoints of a surface. In this case, we cannot see the surface in either image alone but when presented separately to the left and right eyes, (for example, under a stereoscope), a vivid impression of surfaces lying at different depths is perceived. The bottom figure illustrates a three-dimensional reconstruction of the perceived surfaces.

The principal question to be investigated is: *how does this three-dimensional reconstruction take place?* And, in particular, *how does this three-dimensional reconstruction take place in the human visual system?* There are many levels at which we could attempt to answer this question. Traditional methods have included neurophysiological approaches, which have sought to identify the neural structures that perform the reconstruction, and psychophysical approaches, which have sought to identify the perceptual processes involved in the reconstruction. In contrast, this book presents a new approach to the question of "how the processing takes place". This approach is the *computational paradigm* of visual processing [Marr, 1976a, 1976b, 1981; Marr and Poggio, 1977a], which we will outline in greater detail in the rest of this chapter. One of its basic tenets is that the human visual system can be regarded as an *information processor*, performing computations on internal symbolic representations of visual information. As a consequence, we can distinguish between abstract computations and the meaning of the symbols embodied in those representations on the one hand, and the physical manifestation of those symbols on the other. In other words, we can make a distinction between aspects of the visual process which are specific to biological

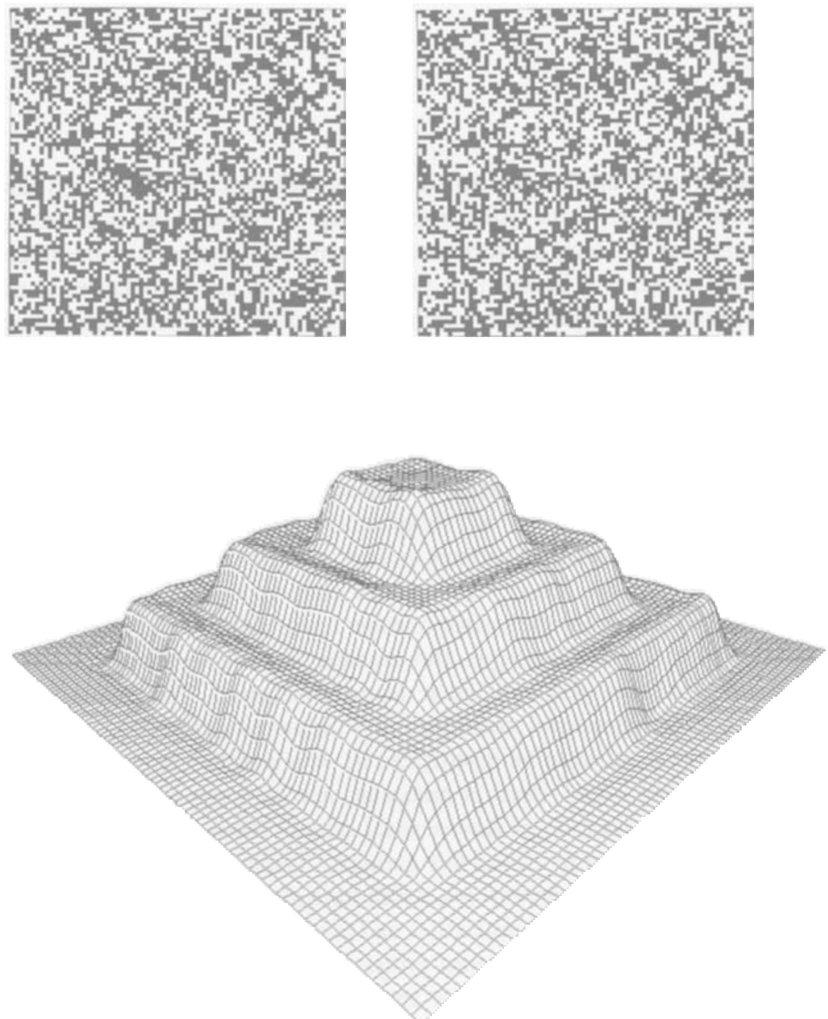

Figure 1.1
Random Dot Stereogram and Its Interpolation. Each image in the top pair is a collection of black and white dots, constructed in such a way that they represent two viewpoints of a set of surfaces in space. When the images are viewed stereoscopically, a series of squares is perceived as separated in depth from the rest of the pattern, although each monocular image contains no cue to this effect. The lower figure represents a perspective view of the surface reconstructed by analyzing the top pair of images with the visual processing algorithms that we will develop in this book.

hardware, and aspects of the process which are specific to the problem being solved, independent of the particular implementation of that process.

Because of this distinction, we can concentrate on the computational process that is occurring independent of the means by which that process is incorporated into the human brain. Our goal is to understand the reconstruction of three-dimensional surfaces from two-dimensional images at the level of computational theories and algorithms. While more precise definitions of computational theory and algorithm will be given in Section 1.1, we can informally consider our goal in the following manner. We begin by investigating the symbolic representations and the transformations between representations that are involved in a computational theory of the construction of surface shape. Next, we consider specific algorithms for performing these transformations, where by algorithm, we mean an ordered set of simple instructions to be performed. In general, there could be many algorithms for solving a particular computational problem. We will attempt to concentrate on the algorithm used by the human visual system. Neurophysiology and psychophysics play an important role in our investigation, by providing information about the architecture of the algorithms which are used to perform the computation, the form of the symbolic representations, and constraints on the transformations that convert one representation into another. It is my hope that computational studies such as this will help to focus research attention in the study of the human early visual system, and will provide a bridge between psychophysics, neurophysiology, mathematics and other areas that can contribute to an understanding of the visual system.

1.1 The Stages of Visual Processing

Information about the three-dimensional geometry of the visible surfaces of a scene is encoded in an image in a variety of ways. Hence, there are several sources of information in the retinal images that can be used for this three-dimensional reconstruction. One of the crucial insights of vision research is that this information can be decoded by independent processes. This will allow us to concentrate on specific modules of the visual system, such as stereopsis, without requiring an understanding of the entire system.

A schematic of the basic stages of analysis involved in early visual processing is presented in Figure 1.2. This architecture has evolved through many computational studies of vision [Marr, 1978, 1981; Marr and Poggio, 1977; Marr and Nishihara, 1978; Barrow and Tenenbaum, 1978]. It can be divided into three stages. The first stage transforms the images into a representation of the locations in an image at which there is a change in some physical property of the corresponding surface in the scene. This representation has been labelled the Primal Sketch, and constitutes the primary source of information for all later

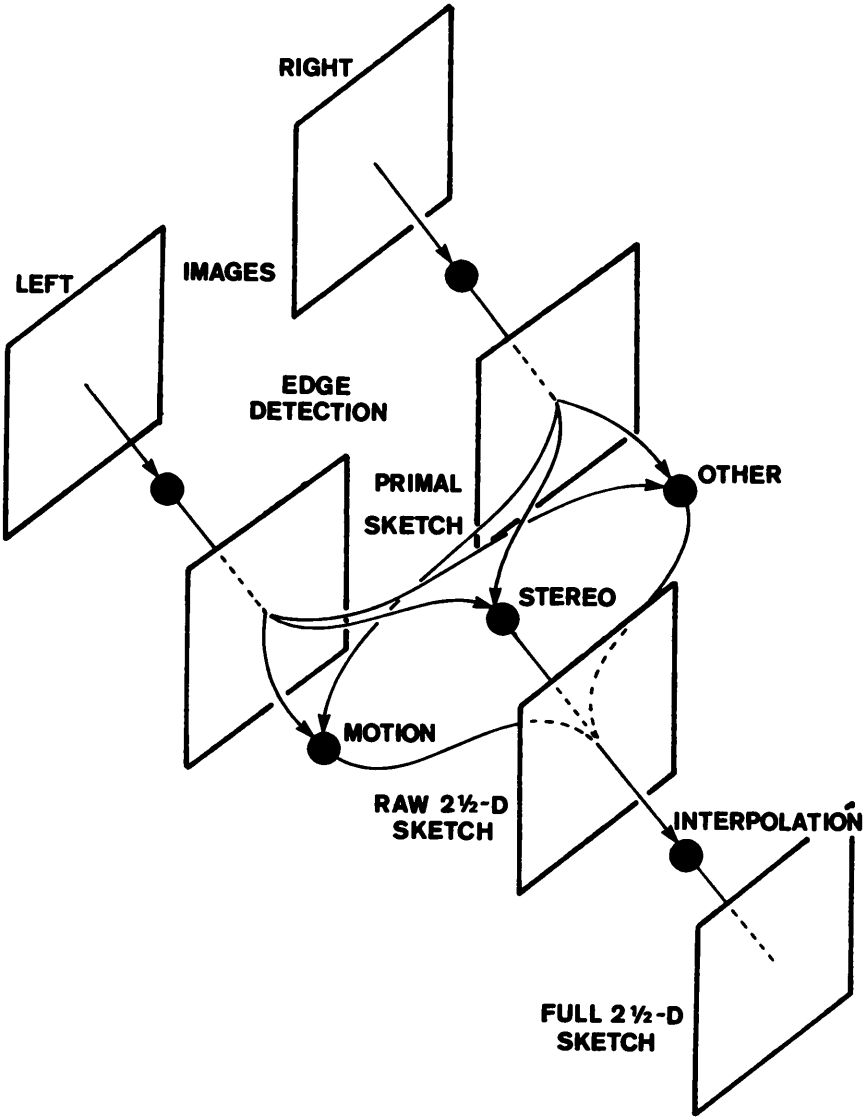

Figure 1.2
Human Early Visual Processing. This diagram schematically illustrates the different representations in the early visual system and the transformations between them. The images obtained by each eye are transformed into Primal Sketch descriptions, consisting of those image positions at which a change of intensity takes place. The Primal Sketch descriptions are then processed by stereo, motion, texture, and other modules of the system. Each computes explicit surface information at the demarked locations of the Primal Sketch. The information from these modules is combined into a single representation. This representation is interpolated, in order to compute explicit surface information at all points in the scene, yielding a representation of the shapes of the visible surfaces, called the $2\frac{1}{2}$D Sketch.

stages of processing. From the Primal Sketch, we would like to compute information about the surface shape at points in the image. A number of modules of the visual system, which to a first approximation are considered independent, perform this computation. Two of the main ones are stereo vision (using Primal Sketch descriptions from the two eyes, which are obtained at different points in space) and motion correspondence (using Primal Sketch descriptions from images which are obtained at different points in time). All of these modules feed a representation, which has been labelled the raw $2\frac{1}{2}$-D Sketch, and which consists of explicit surface information at those locations demarked in the Primal Sketch. Finally, in order to compute explicit surface information at all points in the scene, the raw $2\frac{1}{2}$-D Sketch is interpolated to yield a representation of the shapes of the visible surfaces, which has been labelled the full $2\frac{1}{2}$-D Sketch.

In this book, we will concentrate on aspects of each of the three stages: computing the basic form of the Primal Sketch [Marr and Hildreth, 1980; Hildreth, 1980]; computing surface information based on the Primal Sketches of the left and right eye, using a theory of stereo vision [Marr and Poggio, 1979; Grimson, 1980b, 1981a]; and interpolating the raw $2\frac{1}{2}$-D Sketch to obtain a complete surface description [Grimson, 1981b]. There are other sources of visual information that are important, of course. For example, one can obtain shape from *shading information* [Horn, 1970, 1975; Woodham, 1978; Ikeuchi, 1979; Silver, 1980]; *the motion of objects over time* [Miles, 1931; Johansson, 1964, 1975; Wallach and O'Connell, 1953; Wallach, 1959; Braunstein, 1976; Ullman, 1979a; Longuet-Higgins and Pradzny, 1981]; *surface contours* [Stevens, 1979; Barrow and Tenenbaum, 1981]; *texture* [Helmholtz, 1925; Gibson, 1950a, 1950b, 1966; Purdy, 1960; Bajcsy, 1972; Haber and Hershenson, 1973; Rosinski, 1974; Bajcsy and Lieberman, 1976; Stevens, 1979; Kender, 1978, 1979; Ikeuchi, 1980; Witkin, 1980]; *focusing* [Horn, 1968]; *occluding contours* [Marr, 1977]; *or stereo vision* [Kamal al-Din Abul-Hasan al-Farisi, 1433; Wheatstone, 1838, 1852; Helmholtz, 1925; Julesz, 1971; Quam, 1971; Hannah, 1974; Dev, 1975; Marr and Poggio, 1979; Binford, 1979; Barnard and Thompson, 1980; Mayhew and Frisby, 1981; Baker and Binford, 1981]. Although all of these processes transform representations of the images into representations of the surface shapes, the concentration in this book will be on stereo vision. Thus, while the method we will describe for transforming images into surfaces is not the most general possible (since it neglects sources of surface information other than stereo, for example, motion parallax, shading information, and so forth), the techniques discussed for obtaining the Primal Sketch, and the methods developed for interpolating the raw $2\frac{1}{2}$-D Sketch are generally applicable.

1.2 The Computational Approach

Although our most immediate goal is to investigate the process of creating three-dimensional representations of surface shape, we also have a more global goal of illustrating a computational approach to the study of vision. In the following section, we describe in more detail the characteristics of a computational theory of visual processing, as developed by Marr [1976a, 1976b, 1981, also Marr and Poggio, 1977a].

The motivation for this study of visual processes arose from a desire to understand and model the human visual system. As a result, the theories developed are designed to be both consistent with known evidence about that system and feasible for implementation in a biological system. Although the human system forms the basis for our study, a distinction can be made between components of the computation that are specific to the demands of implementation in a biological system and components that must be performed by any visual processor. Marr's computational approach stresses the importance of distinguishing between these two components when considering computational theories of visual processing. The human visual system can thus serve as a tool for understanding the general processing involved in computing surface descriptions, without enmiring us with details of a specific neural model for such processing in the human system. If an accurate model of the human visual system can be formed, it may also provide a method for solving the visual problem in general situations. In this book, we shall be primarily concerned with the more general questions, that is, with those elements of visual processing which apply to any visual processor.

For the computational paradigm of the visual system as an information processing system to be effective, the representations on which it performs its computations must be *useful descriptions of the visible environment.* Any information that can be obtained from the previous description of the image and that is useful for the construction of the next representation is made explicit at each stage. Thus, every proposed representation is judged by two criteria: the computability of this description, and its suitability for higher level processing (see Marr and Nishihara [1978] for a study of the application of these criteria to judging shape representations).

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 location in the image, and
3. The 3D Model, which makes explicit the three-dimensional shape of objects in the scene, in object-centered coordinates.

In this book, we will investigate the first two representations, formed by creating a description of image locations at which the corresponding scene location undergoes a change in a physical property, and then processing the Primal Sketch descriptions to create a description of the surface geometry.

1.2.1 Levels of Description

Critical to the computational approach is the distinction between several levels of description of a process. Since one is dealing with the manipulation of symbolic information, one 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) independently of the *mechanisms* that actually perform the computation. Although a physical system, and the computation it supports, are related (by the very fact that one is computing the other), they cannot be equated. To illustrate this point, I borrow the example of an electronic calculator from Ullman [1979a, page 2]:

> Some of the events in the electronic calculator have their meaning in the domain of arithmetic. Other events and components, e.g. those inside the power supply do not have such a meaning. The theory of the electronic calculator and the theory of the computation it performs are consequently distinct and non-isomorphic. But the distinction between mechanism and computation runs deeper than this non-isomorphism: the logic governing the computation is not entirely expressible in terms of the physical system. For example, the fact that a standard pocket calculator presents only the first eight digits of the square root of the number 2.0 is a property of the particular device in question. The fact that this number cannot be represented by *any* finite decimal belongs to a different realm, i.e. to the theory of arithmetic. Furthermore, the theory of the mechanism and the theory of the computation deal with different entities. The theory of the calculator deals with electronic circuits, currents, and voltages. The theory of what is being computed, on the other hand, deals with arithmetic objects.

In applying the computational approach, we will study the processing of the visual system at three different levels: the computational theory, an algorithm to solve the theory, and the underlying implementation of the computation [Marr and Poggio, 1977a].

At the level of the computational theory, we must determine 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. But if we can identify additional constraints on the process, imposed naturally as a consequence of the way the world is made, we can restrict the result sufficiently to allow a correct solution to be found. For example, Ullman's [1979a] *rigidity assumption* in the interpretation of three-dimensional structure from motion, Marr and Hildreth's [1980] *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 instances of physical constraints that restrict the problem at hand. This enunciation of additional valid or plausible constraints is a crucial step in the formulation of the computational theory.

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 a process may utilize a constraint. The constraint may be treated as an assumption that is taken always to be true with or without verification (optical illusions often illustrate situations where these assumptions are not valid). An example of this is the case of linear variation [Marr and Hildreth, 1980]. In contrast, some other process might explicitly "look for" the satisfaction of the constraint; if it is consistent with visual input, the constraint is assumed to be true. An example of this is the rigidity constraint [Ullman, 1979a]. Alternatively, the constraint may be explicitly embedded in the process, such as the continuity and uniqueness assumptions in stereo [Marr and Poggio, 1979].

At the level of the computational theory, we must consider representations as well as constraints. All the processes of early vision take properties of the image as their input and compute properties of surfaces, either relating to their geometry or their reflectance, as their output. In the stereo process, it is important to determine the representation of the input to the process, the means of transforming this information into a representation of surfaces, and 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 important for answering these questions.

A fundamental 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 clear to what extent stereoscopic processing is independent of the monocular analysis of each image. 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 but one type of information have been removed. The objective is to determine whether one can make use of just that one type of information. 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 dots, identical except that a centrally located square-shaped region is shifted horizontally in one image relative to the other. Other than this disparity, the images contain no information about visible surfaces. Yet, when the pair is viewed stereoscopically and fused (the two images are brought into correspondence), 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 that the stereo process may be studied in relative isolation from other visual processes.

The idea that a large computation should be divided 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, 1981] calls the *principle of modular design.* This principle forms a cornerstone to the approach. Its importance lies in its heuristic value, that is, without modularity, a small change in one place in a process could have consequences in many other places. This means that the process as a whole becomes extremely cumbersome, and difficult to debug and analyze. While the main role of the principle of modular design is to enable us to derive computational algorithms for particular visual problems, it is worth noting that many of the components of the human visual system exhibit, to a first approximation, some aspects of modularity. The example of the random dot stereogram (Figure 1.1) in stereo vision, and the two cylinders demonstration in the structure-from-motion computation [Ullman, 1979a] both serve to illustrate that the human early visual system can, as a first approximation, be considered a modular one. Of course, we do not necessarily imply that the human visual system is strictly modular, since clearly computations performed by one visual component can influence the computations performed by another. We shall use the principle of modular design as a basic guide for first creating computational algorithms that are modular, and will then consider extensions of the system to account for interaction among the modules.

Having developed 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. We are 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 review for the computational theory. Any implementation of a theory uncovers otherwise unnoticed difficulties with the task and demonstrates the adequacy of the theory. Furthermore, any assumptions made by the theory are tested not just by having an implementation, but by running the implementation on examples. Chapter 4 of this book will detail how a particular implementation of a theory of human stereo vision helped to refine the theory, both through the act of implementation and through the use of the implementation on trial data.

Marr [1976b, 1981] outlines two other criteria that can also be used to guide the design of algorithms, and that ought to be satisfied by any serious candidate for an early process in the human visual system. The first, the *principle of graceful degradation,* says that whenever possible, degraded or impoverished data should not prevent the delivery of at least some of the answer. The second, the *principle* *of least commitment*, says that nothing should ever be done that may later have to be undone.

It is important to note that there may be several possible algorithms for embedding a particular computational theory. In many cases, one can distinguish between the acceptability of different algorithms. For models of the human visual system, I shall adopt a set of algorithmic criteria, outlined in Chapter 7, that support biological feasibility [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, only a few theories have been developed to the point where specific neural implementations can be proposed [for example, Marr, 1969], and none have been confirmed experimentally in every detail. Thus, it is not yet possible to formulate such rules.

Although I have outlined the three levels of description in order from computational theory through algorithm to implementation, this should not be taken as an implication that the process of solving a computational problem also follows this order. Rather, as in any scientific endeavor, the different levels interact in a wide variety of ways, each one serving to provide useful feedback for the other levels. For example, when considering algorithms for solving a particular problem, it is useful to keep in mind the types of architecture which the human system has available for implementing its algorithms. Similarly, while the computational theory is important in providing constraints on an algorithm, the development of the algorithm itself can serve to illuminate constraints on the computational theory which might otherwise have been overlooked. The importance of the levels of description is to identify which questions are particularly relevant to the different aspects of the entire computational problem.

Finally, I wish to note that for a particular computational theory, there may be many algorithms, even with the overall structure that we will derive here, and many more implementations of the algorithms. The algorithms that are developed in this book are one possible set of methods for constructing three-dimensional representations from pairs of images, and there may be many ways of modifying or improving these algorithms. The computational theory is expected to have somewhat more permanency, however. The intent of the computational theory, especially the constraints derived as a part of that theory, is to capture those aspects of the problem that are inherent to any solution of that problem. We now turn to a brief outline of the derivation of a computational theory of three-dimensional surface representations, and specific algorithms for performing the surface reconstruction.

1.3 Overview

The remainder of this book is devoted to a discussion of computational theories of the processing of the human early visual system. Figure 1.2 illustrated schematically the stages of processing used by the human system to transform retinal images into a representation of surface shape.

Initially, the images obtained by each eye are transformed into *Primal Sketch* descriptions, which make explicit those places in an image at which some physical property of the underlying surface changes in a noticeable manner [Marr, 1976b, Marr and Hildreth, 1980, Hildreth, 1980]. In Chapter 2, we examine this process within the context of stereo vision.

We first illustrate the basic problem of stereo, which is to locate points in the images of the right and left eye that correspond to the same location on a surface. We show that if such locations can be found, the difference in the positions of the two retinal locations can be used to compute the distance to the surface. We argue that to perform this correspondence computation, we need to describe image attributes that can be unambiguously identified with specific surface locations. In general, positions on a surface at which a physical property, such as surface material, surface texture, or surface shape, changes radically will satisfy our requirement and such surface locations can be identified in the image by a sudden change in image irradiance. We then derive a method for extracting a description of these image locations, specifically by isolating the *zero-crossings* of the convolution of the image irradiances with a filter whose form is a Laplacian of a Gaussian (Section 2.3). We also show that this processing is consistent with current psychophysical information about the human early visual system.

Given these basic descriptions of the images, we must address the problem of determining the correspondence between descriptors in each image. We state two simple rules, based on the physical structure of objects, to apply to the computation. These are: each descriptor from one image should match at most one descriptor from the other image, and the difference in retinal position of matching descriptors should change smoothly over the image. To apply these rules, we must face the *false targets* problem: since the basic descriptors are relatively simple, there may be several possible descriptors in one image which could correspond to a particular descriptor in the other image. The false targets problem is directly proportional to the range and resolution of depth information over which a match is sought. We show that we may solve the false targets problem without sacrificing either, by matching descriptions obtained at several levels of resolution, and using the rough depth information obtained at a coarse resolution to guide the matching at a fine resolution, by changing the orientation of the eyes. The algorithm we use, derived in Section 2.5, is that proposed by Marr and Poggio [1979]. In the remainder of Chapter 2, we show the relationship between the Marr-Poggio algorithm and currently available psychophysical and neurophysiological information about the human stereo system.

Having developed the stereo problem and an algorithm for solving it, we turn in Chapter 3 to an implementation of the algorithm, first discussed in Grimson [1981a]. Each step of the algorithm is specified in detail. In Chapter 4, we examine the performance of the Marr-Poggio stereo algorithm using the implementation described in Chapter 3. Since the algorithm was developed as a model of the human stereo system, we evaluate its performance by comparing the results of the algorithm to human perception for a wide range of random dot stereograms, (see also [Grimson, 1980a, 1980b, 1981a]). We further demonstrate the capabilities of the algorithm by considering its performance on a series of natural images. Sections 4.3 and 4.4 discuss various aspects of the algorithm and its implementation. Finally, we develop the computation of depth from disparity.

While our first few chapters focus on the process of stereo vision, there are other early visual modules, such as structure-from-motion [Ullman, 1979a], which also compute descriptions of surface shape from Primal Sketch descriptions. Both of these algorithms compute specific surface information only at certain isolated points in the images — in this case, the zero-crossings of the convolved images. (Note that the important point here is not whether zero-crossings or some other descriptor are used as the basic representation on which the computation is performed, but rather that explicit surface information is available only at such points.) In Chapter 5, we argue that such a surface description is not sufficient. We show both computationally and psychophysically that the representation of surface shape should be complete, in the sense of containing a specific depth value everywhere in the representation, rather than just at a set of scattered points. The problem we consider in Chapters 5 through 9 is how to create a complete surface representation, given the results of the stereo algorithm (or the structure-from-motion algorithm).

In principle, there are infinitely many surfaces which could fit any given set of boundary conditions, as provided by the stereo algorithm. In Sections 5.3 and 5.4 we show that most of these surfaces are not consistent with the known information. In particular, a surface whose orientation undergoes a series of radical changes should generally give rise to image irradiances which also undergo a series of radical changes. Such changes would then give rise to zero-crossings in the convolved image. If there are no corresponding zero-crossings in the Primal Sketch, such a surface is inconsistent with the information in the image. We use this argument to derive the *surface consistency constraint* [Grimson, 1981b] which states that

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

In Section 5.5, we review the factors which combine to form the image irradiances and derive Horn's image irradiance equation [Horn, 1970, 1975, 1977]. In Section 5.6, we use this equation to make the surface consistency constraint precise, by the *surface consistency theorem* that relates the probability of a zero-crossing to the variation in surface orientation [Grimson, 1981b].

The surface consistency constraint states that to find a complete surface to fit through the known points, we should choose a surface which is most consistent with that known information. The surface consistency theorem of Section 5.6 indicates that one way of measuring surface consistency is to measure the amount of variation in surface orientation over a region of the surface. In Chapter 6, we consider the problem of measuring this variation. In general, the problem is to determine, given two possible surfaces, which one is more consistent with the zero-crossings. A traditional method for comparing two surfaces is to assign to each surface a real number. Then, in order to compare two surfaces, one need only compare the corresponding real numbers. To do this, we need to define a functional, mapping the space of possible surfaces into the space of real numbers, $\Theta: X \mapsto \mathfrak{R}$. This functional should be such that the more consistent the surface, the smaller the real number assigned to it. In this case, the most consistent surface is that which is minimal under the functional. The development of Chapter 5 leads to a functional that measures variation in surface orientation.

We also require that the problem of finding a most consistent surface be well-defined, that is, that there be a unique most consistent surface. This is not just a mathematical nicety, but follows from the notion that if we create a local, parallel, iterative algorithm to compute the most consistent surface, we need to guarantee that the computation will converge to a unique answer. In Chapter 6 we derive a simple set of mathematical conditions on the functional which guarantee a unique family of solutions. We also show that there are many possible functionals which measure surface consistency and satisfy these conditions. To determine the best functional to use, we investigate the differences in the solution surfaces corresponding to each of these functionals. We also consider the conditions under which the family of solutions will consist of only a single minimal surface. Based on these facts, we argue in Section 6.5 that the best functional to use is the *quadratic variation*. We claim that the *visual surface interpolation problem* is solved by finding the unique surface which minimizes the quadratic variation, while passing through the known points provided by the stereo algorithm (or the structure-from-motion algorithm).

In Chapter 7, we consider what types of algorithms are best applied to solving the interpolation problem. Based on a series of algorithmic constraints, we suggest that the techniques of mathematical programming are appropriate to our problem. We review the *conjugate gradient method*, appropriate to the problem of approximating the surface, and the *gradient projection method*, appropriate to the problem of interpolating the surface.

In Chapter 8, we create explicit algorithms, based on both of these methods, for solving the visual surface interpolation problem. We demonstrate the suitability of the surface interpolation theory by testing the algorithms on a set of synthetic examples. Finally, we complete our original task, by processing a series of stereo images with the Marr-Poggio algorithm, and then creating a complete surface description by applying the interpolation algorithm. Thus, we complete the task of computing surfaces from images.

To complete our discussion, in Chapter 9, we analyze the performance of the algorithm and sketch additional modifications which should improve the performance of the system. We consider the possible benefits of improving the initial input to the stereo algorithm, the interactions between the stereo module and other components of the early visual system, and the problem of detecting discontinuities and their role in the interpolation process. Finally, we indicate some of the implications of this theory for neurophysiology and psychophysics.

Thus, we will develop a theory of stereo vision, which transforms the images into Primal Sketch representations and then transforms these representations into the Raw $2\frac{1}{2}$-D Sketch, and a theory of visual surface interpolation, which transforms the Raw $2\frac{1}{2}$-D Sketch into the Full $2\frac{1}{2}$-D Sketch. Throughout this study, our focus will be on the development of computational theories and algorithms.