Current theories on the structure of the visual system

Institute for Perception Research
PO Box 513, 5600 MB Eindhoven

07.09.1995

Rapport no. 1070

Current theories on the structure of the visual system

T.J.W.M. Janssen

Current theories on the structure of the visual system.

T.J.W.M. Janssen

September 1, 1995

Contents

List of Figures

List of Tables

Chapter 1

Introduction.

Indirect perception and the structure of the visual system.

One important conviction amongst researchers in visual perception is the believe that the visual stimulus contains insufficient information to account for the rich, three-dimensional world that is perceived when we look around us. Therefore, intermediate processing of the visual stimulus – for example using innate knowledge of the visual world, or using previously recorded experiences – is required to solve the puzzle of vision. In following this approach of *indirect perception* one is able to construct a sequence of processes likely to occur in the visual system: the *structure* of the visual system. This report attempts to give a survey of some recent, influential theories regarding the structure of the visual system.

The computational approach to visual perception.

The computational approach to visual perception, with David Marr as its most prominent advocate, is the only approach that made a serious attempt at explaining the structure of the visual system. The computational approach is greatly influenced by the scientific fields of artificial intelligence and computer vision. It can be characterized as a rather mechanistic, top-down approach, focused mainly on the processes that constitute low-level vision. According to computational theorists the visual stimulus is rich in information. However, in order to solve the under-determined problem of reconstructing a three-dimensional world out of a two-dimensional visual image, the visual system needs to use general physical knowledge of the visual world. Therefore, the computational approach stresses the importance of the information processing required to solve the ambiguities encountered by the visual system.

According to Marr, this information processing is divided into three main steps. First, starting with the original image a two-dimensional *primal sketch* is constructed which highlights the important features in the image. Second, from the primal sketch an observer-centered map specifying the distances and orientations of the visible surfaces in the image, the *$2^1/_2$-D sketch*, is formed. And third, the $2^1/_2$-D sketch is transformed into a viewpoint-independent description of the objects visible in the image, the *3-D model representation*.

The ecological approach to visual perception.

As mentioned above, the aim of this report is to give a survey of theories regarding the structure of the visual system. Nevertheless, this report also includes an overview of the ideas of James J. Gibson. Like the computational approach, Gibson's ecological approach to visual perception acknowledges the richness of the visual stimulus. But, in contrast to the former, it emphasizes the existence of a higher-order organization of the visual stimulus in everyday life to which the visual system is tuned *directly*. Thus, no elaborate signal processing by the brain is required, and perception is direct. However, the main reason for including Gibson's ideas in this report is *not* to confront the theories of direct and indirect perception.

Instead the aim is to give an extension to the relatively limited scope of the computational approach regarding the everyday environment surrounding the observer and the interaction between vision and action that occurs in real life. According to Gibson, "[...] Natural vision depends on the eyes in the head on a body supported by the ground, the brain being only the central organ of a complete visual system. [...] The single, frozen field of view provides only impoverished information about the world. The visual system did not evolve for this." [7]. Compared with the rather restricted starting point of the computational approach, i.e., the mere static image, this part of Gibson's approach intuitively seems more appropriate to describe natural vision.

Chapter 2

The visual system according to Marr.

This chapter gives a survey of the theories of David Marr [13]. His contributions to the computational approach to visual perception have been, and still are, extremely influential. Two other influential researchers that are closely related to Marr are Ellen Hildreth and Shimon Ullman. A survey of their interesting work is given in [9].

2.1 Philosophy.

2.1.1 Vision as an information-processing task.

According to Marr, the visual system must confront the loss of information that occurs when a three-dimensional scene is projected onto a two-dimensional image. To reconstruct the lost third dimension, the visual system must necessarily exploit knowledge of the nature of the visual world and the physics of imaging. Consequently, the visual system can be regarded as an information-processing device.

Any information-processing device can only be comprehended completely when it is understood at three distinct levels. First, the level of the computational theory. This is the abstract theory characterizing the performance of the device as a mapping from one kind of information into another. The abstract properties of this mapping are defined precisely, and its appropriateness and adequacy for the task are demonstrated. The second level, the level of representation and algorithm, is primarily concerned with the implementation of the computational theory, i.e., the choice of representation for the input and output, and the algorithm to be used to transform the one into the other. Finally, the third level is the level of hardware implementation: the details of how the algorithm and representation are realized *physically.*

In the eventual understanding of the information-processing device, each of the three levels of explanation will have its own importance. But although the second and third level - *algorithms* and *mechanisms* - are empirically more accessible, it is the level of the computational theory which is critically important from an information-processing perspective, the reason for this being that an algorithm is more likely to be understood by understanding the nature of the problem being solved than by examining the mechanism in which it is implemented. In Marr's opinion: "Trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers." [13]. This is the principal reason for the computational approach being - in agreement with its name - primarily concerned with the level of the computational theory, and perhaps Marr's most important contribution to the scientific field of visual perception.

2.1.2 The structure of the visual system.

According to Marr, vision is a process that produces from images of the external world a description that is *useful* to the observer. Vision can be thought of as a *sequence of representations*, starting with descriptions that can be obtained straight from an image but that are carefully designed to facilitate the subsequent recovery of gradually more objective, physical properties about an object's shape. In Marr's opinion the main step towards this goal is a description of the geometry of the visible surfaces, since the information encoded in images, for example by stereopsis, shading, texture, contours, or visual motion, is due to a shape's local surface properties. The objective of many early visual computations is to extract this information. However, this description of the visible surfaces turns out to be unsuitable for recognition tasks, the most prominent cause being that like all early visual processes, it depends critically on the vantage point. Marr's final step therefore consists of transforming the viewer-centered surface description into a representation of the three-dimensional shape and spatial arrangement of an object that does not depend upon the direction from which the object is being viewed. This final description is object-centered rather than viewer-centered.

The overall framework thus outlined divides the derivation of shape information from images into three representational stages:

1. The representation of properties of the two-dimensional image, such as intensity changes and local two-dimensional geometry. This representation will be discussed in section 2.2.

2. The representation of properties of the visible surfaces in a viewer-centered coordinate system, such as surface orientation, distance from the viewer, and discontinuities in these quantities; surface reflectance; and some coarse description of the prevailing illumination. It will be discussed in section 2.3.

3. An object-centered representation of the three-dimensional structure and of the organization of the viewed shape, together with some description of its surface properties. This representation will be discussed in 2.4.

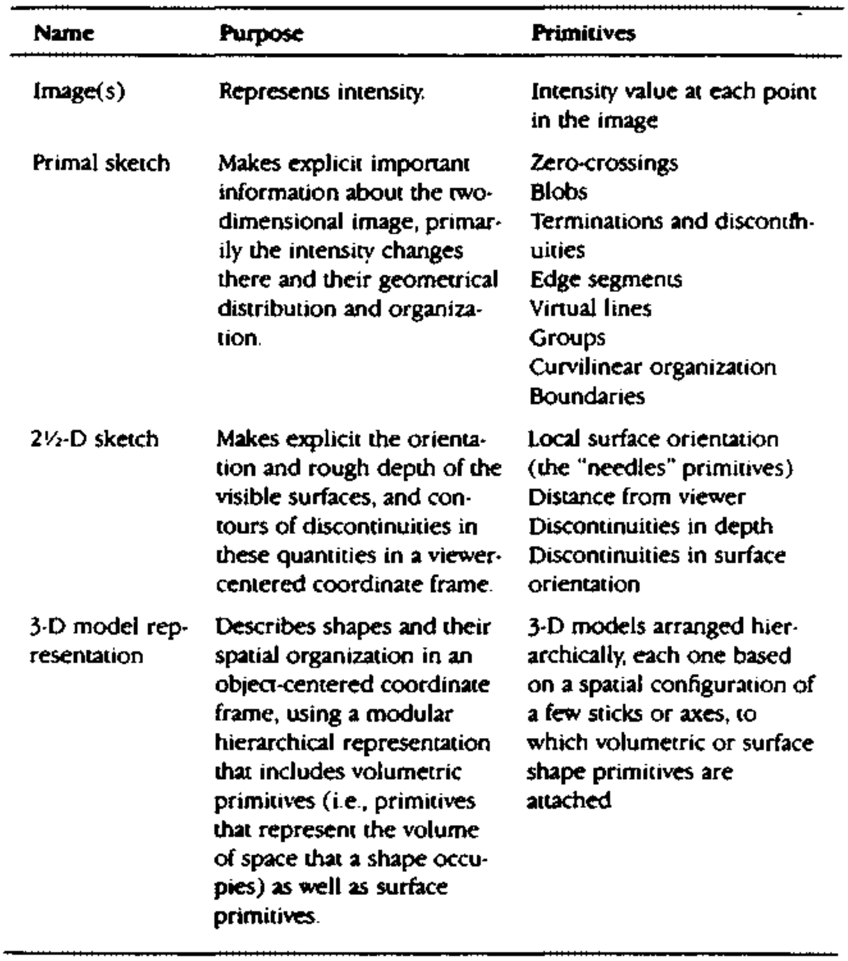
This framework is also summarized in table 2.1.

2.2 The early representations of the image.

2.2.1 Representing the image.

Four factors are mainly responsible for the intensity values in an image. They are (1) the geometry and (2) the reflectances of the visible surfaces, (3) the illumination of the scene, and (4) the viewpoint. In an image, all these factors are intertwined, some intensity changes being due to one cause, other to another, and some to a combination. The purpose of early visual processing is to identify which changes are due to what factors and hence to create representations in which the four factors are separated. This aim is accomplished in two stages. First, suitable representations are obtained of the changes and structures in the images. This involves the detection of intensity changes, the representation and analysis of local geometric structure, and the detection of illumination effects like light sources, highlights, and transparency. The result of this first stage is called the *primal sketch*. Second, a number of processes operate on the primal sketch to derive a representation – still retinocentric – of the geometry of the visible surfaces. This second representation, that of the visible surfaces, is called the *$2^1/_2$-D sketch*. Both the primal sketch and the $2^1/_2$-D sketch are constructed in a viewer-centered coordinate frame, and this is the aspect of their structures denoted by the term *sketch*.

The main purpose of the early representations is to give a description of the image suitable for detecting changes in the image's geometrical organiz-



Name	Purpose	Primitives
Image(s)	Represents intensity.	Intensity value at each point in the image
Primal sketch	Makes explicit important information about the two-dimensional image, primarily the intensity changes there and their geometrical distribution and organization.	Zero-crossings Blobs Terminations and discontinuities Edge segments Virtual lines Groups Curvilinear organization Boundaries
2½-D sketch	Makes explicit the orientation and rough depth of the visible surfaces, and contours of discontinuities in these quantities in a viewer-centered coordinate frame.	Local surface orientation (the "needles" primitives) Distance from viewer Discontinuities in depth Discontinuities in surface orientation
3-D model representation	Describes shapes and their spatial organization in an object-centered coordinate frame, using a modular hierarchical representation that includes volumetric primitives (i.e., primitives that represent the volume of space that a shape occupies) as well as surface primitives.	3-D models arranged hierarchically, each one based on a spatial configuration of a few sticks or axes, to which volumetric or surface shape primitives are attached

Table 2.1: Marr's framework for deriving shape information from images. Adapted from [13].

ation due to changes in the reflectances of the surfaces or to changes in the surfaces' orientations or distances from the viewer. Changes in orientation and perhaps also in distance are likely to give rise to a change in image intensity. If the surface is textured, then quantities like the orientation or size of tiny elements on the surface – perhaps rough length and width – and measures taken over a small area reflecting the density and spacing of these elements yield the important clues in an image. Hence, the early representations should contain some type of *token* that can be derived reliably and repeatedly from images and to which can be assigned values of attributes like orientation, brightness, size (length and width), and position (for density and spacing measurements). It is of critical importance that these tokens correspond to real physical changes on the viewed surface; they must not be artifacts of the imaging process. or else inferences made from their structure backwards to the structure of the surface will be meaningless.

The general nature of surface reflectance functions gives important clues as to how to structure the early representations. According to Marr, the underlying physical assumptions are:

- *Existence of surfaces*: The world can be regarded as being composed of smooth surfaces having reflectance functions whose spatial structure may be elaborate.

- *Hierarchical organization*: The spatial organization of a surface's reflectance function is often generated by a number of different processes. each operating at a different scale.

- *Similarity*: The items generated on a given surface by a reflectance-generating process acting at a given scale tend to be more similar to one another in their size. local contrast. colour, and spatial organization than to other items on that surface.

- *Spatial continuity*: Markings generated on a surface by a single process are often spatially organized – they are arranged in curves or lines and possibly create more complex patterns.

- *Continuity of discontinuities*: The loci of discontinuities in depth or in surface orientation are smooth almost everywhere.

- *Continuity of flow*: If direction of motion is ever discontinuous at more than one point – along a line, for example – then an object boundary is present.

The important message of these physical constraints is that although the basic elements in the image are the intensity changes, the physical world imposes on these raw intensity changes a wide variety of spatial organizations, roughly independently at different scales. This organization is reflected in the structure of images, and since it yields important clues about the structure of the visible surfaces, it needs to be captured by the early representations of the image. In Marr's opinion this can be done by a set of *place tokens* that roughly correspond to oriented *edge* or *boundary* segments or to points of *discontinuity* in their orientations, to *bars* (roughly parallel edge pairs) or to their *terminations*; or to *blobs* (roughly, doubly terminated bars). These primitives can be defined in very concrete ways - from pure discontinuities in intensity - or in rather abstract ways. This representational scheme (e.g., see figure 2.1) is called the *primal sketch.* The critical ideas behind it are the following:

1. The primal sketch consists of primitives of the same general kind at different scales but the primitives can be defined from an image in a variety of ways, from the very concrete to the very abstract.

2. These primitives are built up in stages in a constructive way, first by analyzing and representing the intensity changes and forming tokens directly from them, then by adding representations of the local geometrical structure of their arrangement, and then by operating on the latter with active selection and grouping processes to form larger-scale tokens that reflect larger-scale structures in the image, et cetera.

3. On the whole, the primitives that are obtained, the parameters associated with them, and the accuracy with which they are measured are designed to capture and to match the structure in an image so as to facilitate the recovery of information about the underlying geometry of the visible surfaces. This gives rise to a complex balance between the accuracy of the discriminations that can be made and the value of making them.

The three main stages in the processes that derive the primal sketch are (1) the detection of zero-crossings, (2) the formation of the raw primal sketch, and (3) the creation of the full primal sketch.

2.2.2 Zero-crossings and the raw primal sketch.

The first of these three stages concerns the detection of intensity changes. The two ideas underlying this detection are (1) that intensity changes occur

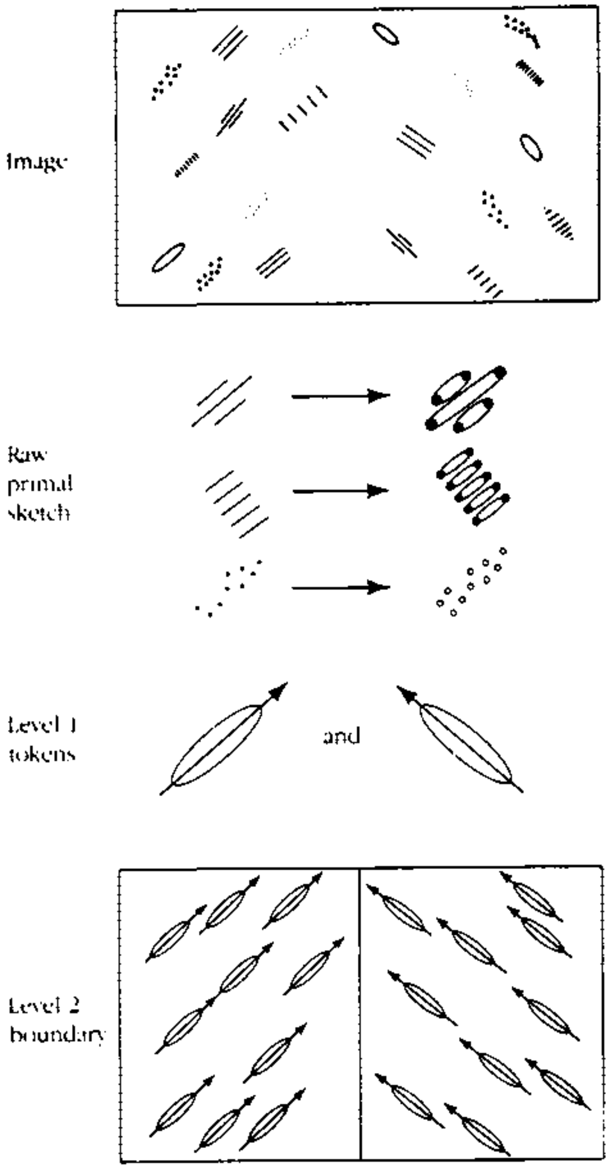


Figure 2.1: The hierarchical structure of the primal sketch. At the lowest level the intensity changes are copied into the raw primal sketch, and tokens which represent terminations are added. Next, grouping processes act upon the primal sketch, and group orientation tokens are added. At the highest level boundaries are constructed between groups with different orientations. As might be expected, the exact structure of the primal sketch depends upon the organization of the original image at the various scales. Adapted from [13].

at different scales in an image, and so their optimal detection requires the use of operators of varying sizes; and (2) that a sudden intensity change will give rise to a peak or trough in the first derivative or, equivalently, to a *zero-crossing* in the second derivative. These ideas suggest that in order to detect intensity changes efficiently, one should search for a filter that has two prominent characteristics. First and foremost, it should be a differential operator, taking either a first or a second spatial derivative of the image. Second, it should be capable of being tuned to act at any desired scale, so that large filters can be used to detect blurry shadow edges, and small ones to detect sharply focused fine detail in the image. According to Marr, the most satisfactory operator fulfilling these conditions is the filter $\nabla^2 G$, where ∇^2 is the Laplacian operator and G stands for the two-dimensional Gaussian distribution.

There are two basic ideas behind the choice of the filter $\nabla^2 G$. The first is that the Gaussian part of it blurs the image, effectively wiping out all structure at scales much smaller than the space constant σ of the Gaussian. The reason for one to choose the Gaussian for this purpose, is that the Gaussian distribution has the desirable characteristic of being smooth and localized in both the spatial and frequency domains and, in a strict sense, being the unique distribution that is simultaneously optimally localized in both domains. And the reason, in turn, for this being a desirable property of the blurring filter is that if the blurring is as smooth as possible, both spatially and in the frequency domain, it is least likely to introduce any changes that were not present in the original image. The second idea concerns the derivative part of the filter, ∇^2. The great advantage of using it is economy of computation: it is the lowest-order isotropic differential operator. Hence, in practice, the most satisfactory way of finding the intensity changes at a given scale in an image is first to filter it with the operator $\nabla^2 G$, where the space constant of G is chosen to reflect the scale at which the changes are to be detected, and then locate the zero-crossings in the filtered image.

Zero-crossings can be represented symbolically in various ways, one of them being a set of oriented primitives called *zero-crossing segments*, each describing a piece of the contour whose intensity slope and local orientation are roughly uniform. Because of their eventual physical significance, it is also important to make explicit those places at which the orientation of a zero-crossing changes *discontinuously* (according to a practical definition of discontinuity). In addition, small, closed contours are represented as blobs, each also with an associated orientation, average intensity slope, and size defined by its extent along a major and minor axis. Finally, in keeping with the overall plan, several sizes of operator will be needed to cover the range

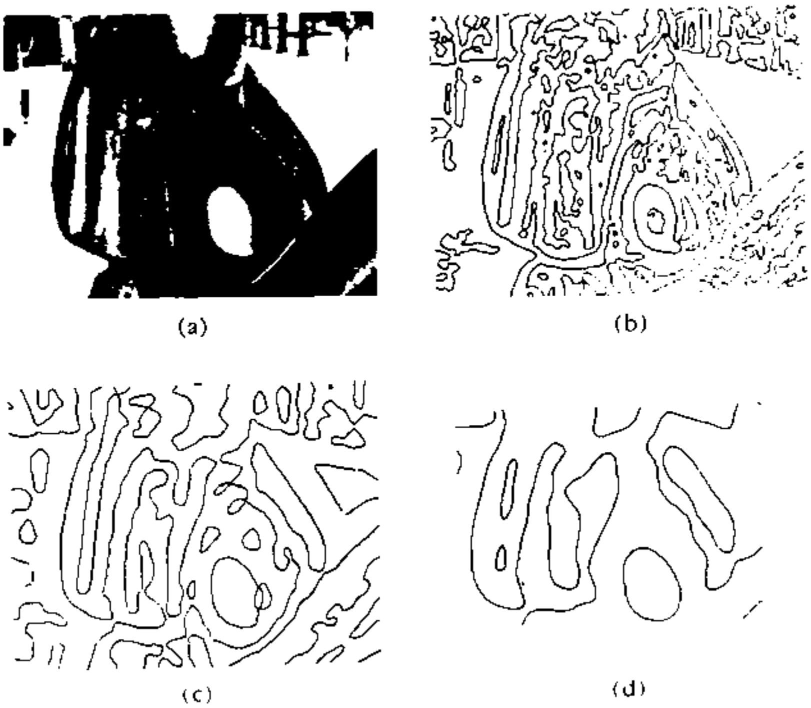


Figure 2.2: The original image (a) and the zero-crossings detected by three filters with increasing size (b – d). Adapted from [13].

of scales over which intensity changes occur. Figure 2.2 is an example of the zero-crossings obtained after filtering the original image with a set of three $\nabla^2 G$-filters with increasing sizes. As can be seen from this example, the smallest filter primarily detects minute detail. In contrast, the larger channels mainly detect overall structure.

At this point, the information that is available is the zero-crossings of the image after filtering it through $\nabla^2 G$-filters of different sizes. The next problem is how to combine the information from these different channels. The physical world constrains the geometry of the zero-crossings from the different-sized channels. This can be exploited, Marr argues, by formulating the *spatial coincidence assumption*: If a zero-crossing segment is present in a set of independent $\nabla^2 G$-channels over a contiguous range of sizes, and the segment has the same position and orientation in each channel, then the set of such zero-crossing segments indicates the presence of an intensity change in the image that is due to a single physical phenomenon - a change in reflectance, illumination, depth, or surface orientation. Thus, provided that the zero-crossings from independent channels of adjacent sizes coincide, they can be taken together. If the zero-crossings do not coincide, they probably arise from distinct surfaces or physical phenomena. It follows (1) that the minimum number of $\nabla^2 G$-channels required to establish physical reality is two and (2) that if there is a range of channel sizes, reasonably well separated in the frequency domain and covering an adequate range of the frequency spectrum, rules can be derived for combining their zero-crossings into a description whose primitives are *physically meaningful*.

The description of the image to which these ideas lead is called the *raw primal sketch*. Its primitives are edges, bars, blobs, and terminations, and these have attributes of orientation, contrast, length, width, and position. It can be thought of as a map specifying the precise positions of the edge segments, together with the specifications at each point along them of the local orientation and of the type and extent of the intensity change. Blob, bar, and discontinuity primitives can be made explicit in much the same way. The raw primal sketch is a very rich description of an image, since it contains virtually all the information in the zero-crossings from several channels. Its importance is that it is the first representation derived from an image whose primitives have a high probability of reflecting physical reality directly.

2.2.3 Grouping processes and the full primal sketch.

The purpose for which the raw primal sketch is to be used, is to infer the geometry of the underlying surfaces. The physical assumptions stated in 2.2.1, together with the natural consequences for an image of changes in depth and surface orientation, can be used to formulate a list of image properties whose detection will aid this task of decoding surface geometry:

1. Average local *intensity*, from the first physical assumption (changes in average intensity can be caused by changes in illumination, perhaps due to changes in depth, and by changes in surface orientation or surface reflectance).

2. Average *size* of items on a surface that are similar to one another, in the sense of the second and third physical assumptions (the term *size* includes the concepts of length and width).

3. Local *density* of the items defined in image property 2.

4. Local *orientation*, if such exists, of the terms defined in image property 2.

5. Local *distances* associated with the spatial arrangement of similar items (the third and fourth physical assumptions), i.e., the distance between neighbouring pairs of similar items.

6. Local *orientation* associated with the spatial arrangement of similar items (the third, fourth, and fifth physical assumptions), i.e., the orientation of the line joining neighbouring pairs of similar items.

In Marr's opinion, there are two main goals in the analysis now: (1) to construct tokens that capture the larger scale structure of the surface reflectance function and (2) to detect various types of change in the measured parameters associated with these tokens that could be of help in detecting changes in the orientation and distance from the viewer of the visible surfaces. Thus, the goals are to make tokens and to find boundaries. Both tasks require selection processes whose function is to combine roughly similar types of tokens into larger tokens or to construct boundaries between sets of tokens that differ in certain ways. In general terms, the approach is to build up descriptive primitives in an almost recursive manner. The material from which everything starts is the raw primal sketch. By doing this again and again, one builds up tokens or primitives at each scale that capture the spatial structure at that scale. Once these primitives have been constructed, they can reveal the geometry of the visible surfaces – either by means of detecting the changes in surface reflectance or by detecting changes that could be due to discontinuities in surface orientation or depth.

At a change in the surface, the change in the reflectance function is usually so great that almost any measure will detect it. Boundaries that might be caused by surface discontinuities can be detected in two ways. One is by finding sets of tokens that owe their existence to the physical discontinuity and are therefore organized geometrically along it. The second type of task is to measure locally (at different scales) the six quantities defined above and to make explicit, by means of a set of boundary or edge primitives, places where discontinuities occur in these measures. The reason for adding such boundaries to the representation of the image is that they may provide important evidence about the location of surface discontinuities. In doing so, parameters likely to have arisen because of discontinuities in the surface ought to be those that give rise to perceptual boundaries, whereas those that could probably not have their origins traced to geometrical causes should be much less likely to produce perceptual boundaries. This is Marr's *hypothesis* *of geometrical origin for perceptual texture boundaries.* The principal limitations on its usefulness come from the fact that reflectance functions seldom have a precise geometrical structure. Hence, small changes in orientation in an image that may be produced by small changes in surface orientation will not usually produce a clear signal. The same applies to changes in apparent size in an image, although density allows a more sensitive discrimination.

2.3 From images to surfaces.

2.3.1 The modular organization of the visual system,

According to Marr, the existence of a *modular organization* in the human visual system indicates that different types of information can be analyzed in relative isolation. Information about the geometry and reflectance of visible surfaces is encoded in the image in various ways and can be decoded by processes that are almost independent. Processes quite well understood are:

- stereopsis
- directional selectivity
- structure from apparent motion
- depth from optical flow
- surface orientation from surface contours
- surface orientation from surface texture
- shape from shading
- photometric stereo
- lightness and colour as an approximation to reflectance

These processes will now be discussed shortly.

Stereopsis.

The process of stereopsis can be subdivided into two parts: measuring stereo disparity, and computing distance and surface orientation from disparity. The second part follows from a rather straightforward geometrical exercise, and therefore will not be discussed any further. The first part, measuring stereo disparity, can be divided into three steps: (1) A particular location on a surface in the scene must be selected from one image; (2) that same location must be identified in the other image; and (3) the disparity between the two corresponding image points must be measured. Three matching constraints restrict the allowable ways of matching two primitive descriptions, one from each eye. First, the *compatibility constraint*: if two descriptive elements could have arisen from the same physical marking, then they can match. If they could not have, then they cannot be matched. Second, the *uniqueness constraint*: in general, each descriptive item can match only one item from the other image. And third, the *continuity constraint*: disparity varies smoothly almost everywhere. The usefulness of these constraints follows from the *fundamental assumption of stereopsis*: If a correspondence is established between physically meaningful primitives extracted from the left and right images of a scene that contains a sufficient amount of detail, and if the correspondence satisfies the three matching constraints, then that correspondence is physically correct.

Marr describes the following matching algorithm: (1) Each image is analyzed through channels of varying coarseness and matches take place between corresponding channels from the two eyes for disparity values of the order of the channel resolution; (2) coarse channels control vergence movements, thus causing fine channels to come into correspondence; (3) when a correspondence is achieved, it is held and written down in the $2^1/_2$-D sketch; and (4) there is a reverse relation between the memory and the channels, acting through the control of eye movements, that allows one to fuse any piece of surface easily once its depth map has been established in the memory. The input representation for the stereo matching process consists of the raw zero-crossings, labeled by the sign of their contrast change and their rough orientation in the image, and of terminations (local discontinuities) also labeled by contrast and perhaps very rough orientation.

Directional selectivity.

Directional selectivity is concerned with using partial information about motion - specifically, only its direction defined to within 180° - in order to discern the two-dimensional shapes or regions in the visual field based on their relative movement. The motivation for studying what direction alone can tell comes from the *aperture problem*: the only motion that can be detected directly through a small aperture placed over an edge is motion at right angles to that edge, i.e., forward or backward. The earliest stage at which direction of motion can be detected is at the level of zero-crossings segments. A zero-crossings segment is defined as a locally oriented segment of the zero values of the convolution $\nabla^2 G * I$. In an algorithm suggested by Marr the time derivative $\partial/\partial t(\nabla^2 G * I)$ is measured at the location of a zero-crossing. The direction of movement is then specified by the sign of this measure together with the sign of the contrast along the zero-crossing.

Since the motions of unconnected objects are generally unrelated, the velocity field will often be discontinuous at object boundaries. Conversely, lines of discontinuity are reliable evidence of an object boundary. Unfortunately, the complete velocity field is not directly available from measurements of small oriented elements. Because of the aperture problem, only the sign of the direction of movement is available locally. Thus, additional constraints are necessary. The sign of the local direction of motion determines neither the movement's speed nor its true direction, but is does place constraints on what the true direction can be. This constraint depends on the orientation of the local element, so if the visible surface is textured and gives rise to many local orientations, the true direction of movement may be rather tightly constrained.

Apparent motion.

Apparent motion is concerned with detecting the changes induced by motion and to use them to recover the three-dimensional structures in motion. This introduces two kinds of task. The first is to follow things around as they move in the image and to measure their positions at different times: the *correspondence problem*. The second task is to recover three-dimensional structure from these measurements: the *structure-from-motion problem*. The correspondence problem in apparent motion yields two tasks. The aim of the first task is to achieve a very detailed correspondence between accurately localizable items in the image, so that measurements of their position changes may be made to the (second order) precision needed for the structure-from-motion computations. The aim of the second task is to establish consistency of an object's identity through time. Precision is not its goal; instead its aim is to establish rough identity of an object changing its shape, configuration and even reflectance between two temporal viewpoints.

In Marr's opinion, the fundamental assumption underlying the structure-from-motion theorem is the *rigidity assumption*: Any set of elements undergoing a two-dimensional transformation that has a unique interpretation as a rigid body moving in space is caused by such a body in motion and hence should be interpreted as such. It implies that if a body is rigid, its three-dimensional structure can be found from three frames. If it is not rigid, the chances of there being an accidental rigid interpretation are vanishingly small, so in practice, the method will fail. What is needed is an algorithm that *degrades gracefully*. The algorithm should be able to deliver an account of the structure that is at first rather rough but which becomes increasingly accurate as more views and hence more information are presented. Also, if the viewed object is not quite rigid, the algorithm should be able to produce the not-quite-rigid structure, perhaps again at the price of needing more points or more views to work on.

Shape contours.

Shape contours can be divided into three categories: (1) contours that occur at discontinuities in the distance of the surface from the viewer (occluding contours), (2) contours that follow discontinuities in surface orientation, and (3) contours that lie physically on the surface, e.g., due to surface markings or to shadow lines. Occluding contours usually correspond to the silhouette of an object as seen in two-dimensional projection. They often reveal the shape of an object more accurately then they should . Three assumptions seem to be important. The first is that each line of sight from the viewer to the object should graze the object's surface at exactly one point. This assumption allows one to speak of a particular curve on the object's surface called the *contour generator*. The second assumption is that nearby points on the contour in an image arise from nearby points on the contour generator on the viewed object. The third assumption states that the contour generator is planar, i.e., the contour generator lies in a plane. These three assumptions together led Marr to formulate the following basic idea: If the surface is smooth and if the three assumptions stated above hold for all distant viewing positions in any one plane, then the viewed surface is a *generalized cone*. A generalized cone can be described as the surface created by moving a cross section along an axis; the cross section may vary smoothly in size, but its shape remains the same.

Surface orientation contours mark the loci of discontinuities in surface orientation. With regard to recovering the geometry of the surface, the most important question about such a contour is whether it corresponds to a convexity or a concavity on the surface. However, it is often difficult to distinguish convexities and concavities from purely local cues in a monocular image.

Surface contours arise for various reasons in the image of smooth surfaces, and they yield information about the three-dimensional shape of the surface. In determining the shape of the contour generator, the assumption that the contour generator is planar greatly simplifies the problem. However, it is difficult to be confident of this assumption, and the assumption that the contour-generator has the minimum possible curvature sometimes seems more useful. In the case of a number of parallel, shifted contour generators the curvature of the surface in the direction of the shift can locally be ignored. The surface can be thought of locally as a cylinder, and this restriction allows the interpretation of the global structure of the surface.

Surface texture.

In using surface texture, the first problem is how to extract from an image the uniform texture elements from which subsequent analysis must proceed. A full answer to this would include a complete understanding of the full primal sketch and of the selection by similarity, whose task it is to classify items by origin and whose importance was already encountered. As a first approximation, Marr assumes that the world's surfaces are covered with regular and sufficient markings, and that it is possible to discover them from the early representations of an image.

There are two ways in which a surface may be specified relative to the viewer: either the distance relative to local pieces of it is specified, or the surface orientation relative to the viewer is specified. Surface orientation itself is naturally split into two components: *slant* (the angle by which the surface dips away from the frontal plane) and *tilt* (the direction in which the dip takes place). Which of these quantities, distance, slant, or tilt, is actually extracted from measurements of variations in texture? According to Marr, the answer to this is as follows:

1. Tilt is probably extracted explicitly.
2. Probably distance is also extracted explicitly.
3. Slant is probably inferred by differentiating estimates of scaled distance made in accordance with point 2.
4. In particular, measurements of texture gradients, which are closely associated mathematically with slant, are probably *not* made or used, perhaps because of the inaccuracies inherent in the measurement process.

In Marr's opinion the analysis of texture lies in a somewhat unsatisfactory state. It is not at all obvious to what extent the vagaries of the natural world allow the visual system to make use of the possible mathematical relations. Once more is known about this matter it shall be understood why the human visual system handles texture information in the rather peculiar and limited way in which it appears to operate.

Shape from shading.

The human visual system incorporates some processes for inferring shape from shading. although it seems likely that the power of these processes is only slight. Shape from shading is concerned with deducing surface orientation from image intensity values. The problem is complicated because intensity values do not depend on surface orientation alone; they depend on how the surface is illuminated and on the surface reflectance function. However. for the very simple illumination condition of one distant point source and a combination of perfect matte surfaces (surfaces having a Lambertian reflectance function) and pure mirror surfaces (surfaces having a specular reflectance function) the problem is solvable. One additional assumption is necessary: the surface should be smooth *and* the surface orientation should vary smoothly. Nevertheless, it seems likely that the human visual system only uses coarse shading information. and shading information is easily overridden by other cues.

Brightness, lightness and colour.

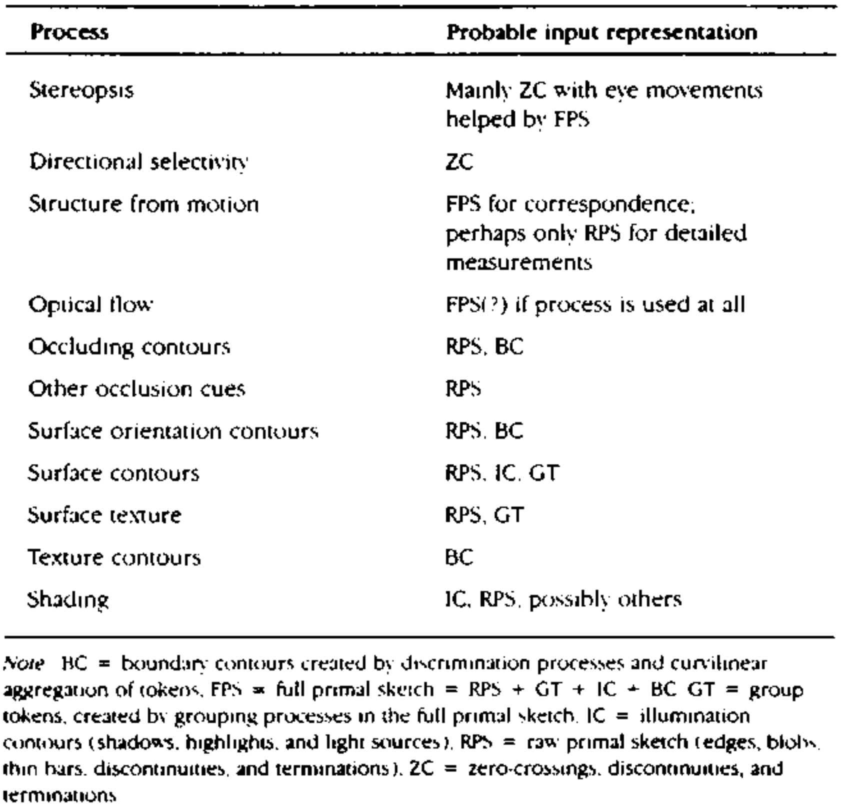
According to Marr. the only attempt at a true theory of colour vision thus far had been Land and McCann's retinex theory. The first part of this theory is concerned with lightness. The central problem of it is to separate the effects of surface reflectance from the effects of the illuminant. because what is perceived as the colour of a surface is much more closely connected with spectral characteristics of its reflectance function than with the spectral characteristics of the light falling upon the eye. What characteristics enable one to separate the effects due to changes in illumination from the effects due to changes in reflectance? Land and McCann propose that changes due to the illuminant are on the whole gradual, appearing usually as smooth illumination gradients, whereas those due to changes in reflectance tend to be sharp. If one can separate the two types of change. one is able to separate effects of changes in the illuminant from the effects of changes in reflectance in an image. According to the retinex theory, what the visual system essentially does is removing the illumination gradient by applying a thresholding operation. This computation is called the retinex computation.

In order to apply this operation to colour, Land and McCann require that it be performed independently in each of the red, green, and blue channels. What then emerges from each are signals that should depend not on the illuminant but solely on the surface reflectance. These can be combined to give a percept of colour that rests solely on properties of surface reflectance and not on the vagaries of its particular present illuminant. There still is the need of calibrating the signals. Land and McCann suggest doing this by calling the brightest point in the scene white.

In Marr's opinion, brightness perception and colour perception appear to involve at least some effects that are not predicted by Land and McCann's approach. When looking at the subject more closely, three observations can be made: (1) locally measurable illumination gradients on a flat surface can only occur if the light source is not very far away; (2) these illumination gradients are small unless the source is very near; and (3) they are approximately linear except perhaps directly under the source. These observations suggest that the following approach to the physical basis of colour vision may be profitable: In the absence of sharp changes in brightness, detectable as shadow boundaries or changes in surface orientation, all nonlinear changes in intensities may be assumed to be due to properties of the surface - either its orientation or its reflectance. In other words, in the absence of obvious illumination effects like shadows, measurable nonlinear local differences in either image intensity or spectral distributions are due to changes in lightness or colour of the surface. This assumption allows one to ignore small linear changes and provides a basis for the idea that lightness and colour may be recovered from measurements of nonlinear local changes in intensity and spectral distribution made, for example, by comparing their values at each point with their values in the surrounding neighbourhood.

To conclude.

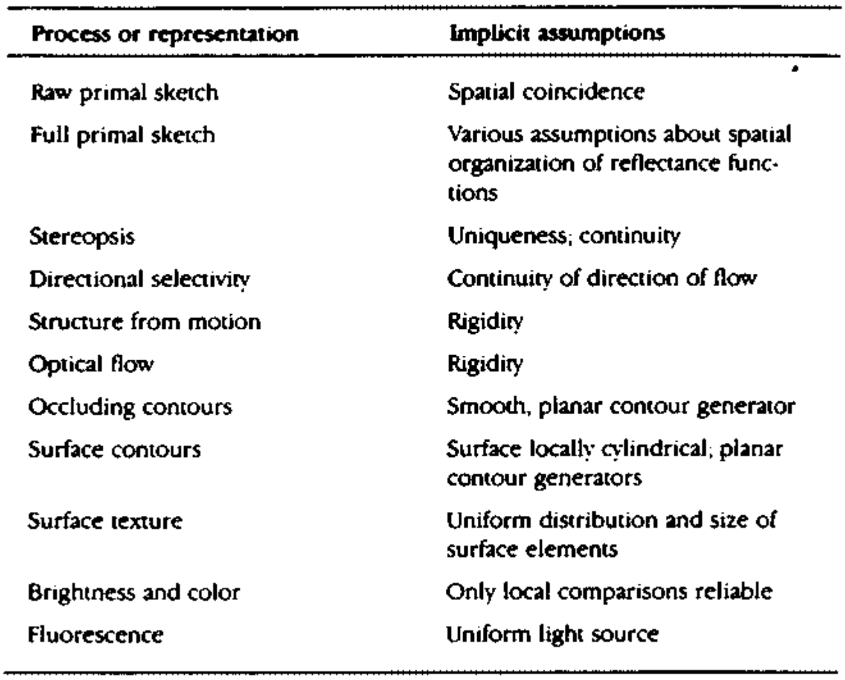
The various processes described above appear to use slightly different input representations. In addition, they all involve slightly different assumptions about the world in order to work satisfactorily. In each case the surface structure is strictly under-determined from the information in images alone. Discovering what additional information can safely be assumed about the world provides powerful enough constraints for the processes to run. In table 2.2 the input representations to the various processes are summarized. Additional assumptions implicit in these processes are summarized in table 2.3.



Process	Probable input representation
Stereopsis	Mainly ZC with eye movements helped by FPS
Directional selectivity	ZC
Structure from motion	FPS for correspondence; perhaps only RPS for detailed measurements
Optical flow	FPS(?) if process is used at all
Occluding contours	RPS, BC
Other occlusion cues	RPS
Surface orientation contours	RPS, BC
Surface contours	RPS, IC, GT
Surface texture	RPS, GT
Texture contours	BC
Shading	IC, RPS, possibly others

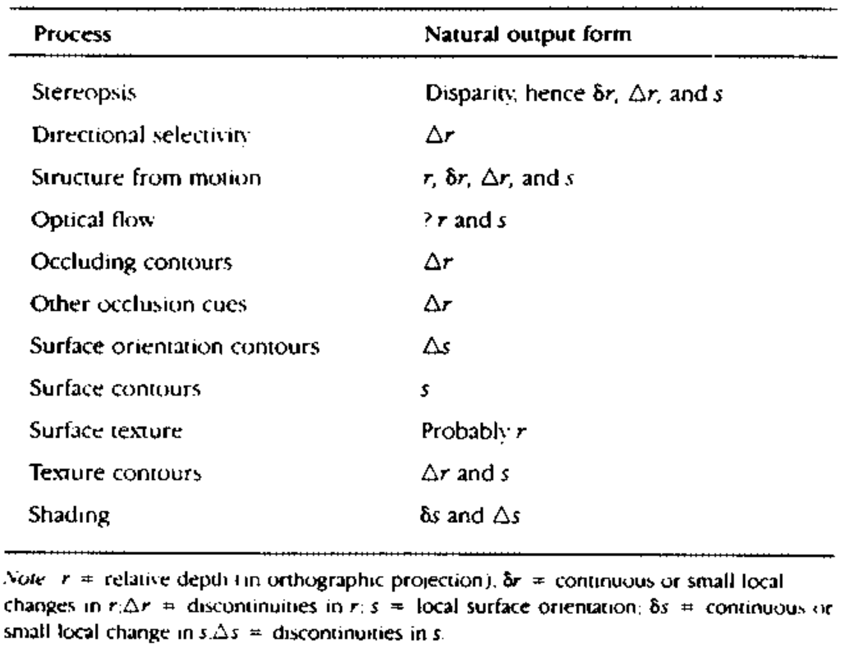
Note BC = boundary contours created by discrimination processes and curvilinear aggregation of tokens. FPS = full primal sketch = RPS + GT + IC + BC. GT = group tokens, created by grouping processes in the full primal sketch. IC = illumination contours (shadows, highlights, and light sources). RPS = raw primal sketch (edges, blobs, thin bars, discontinuities, and terminations). ZC = zero-crossings, discontinuities, and terminations

Table 2.2: Input representations of the processes that derive surface information from images. Adapted from [13].



Process or representation	Implicit assumptions
Raw primal sketch	Spatial coincidence
Full primal sketch	Various assumptions about spatial organization of reflectance functions
Stereopsis	Uniqueness; continuity
Directional selectivity	Continuity of direction of flow
Structure from motion	Rigidity
Optical flow	Rigidity
Occluding contours	Smooth, planar contour generator
Surface contours	Surface locally cylindrical; planar contour generators
Surface texture	Uniform distribution and size of surface elements
Brightness and color	Only local comparisons reliable
Fluorescence	Uniform light source

Table 2.3: Additional assumptions used by the processes that derive surface information from images. Adapted from [13].



Process	Natural output form
Stereopsis	Disparity, hence δr, Δr, and s
Directional selectivity	Δr
Structure from motion	r, δr, Δr, and s
Optical flow	? r and s
Occluding contours	Δr
Other occlusion cues	Δr
Surface orientation contours	Δs
Surface contours	s
Surface texture	Probably r
Texture contours	Δr and s
Shading	δs and Δs

Note r = relative depth (in orthographic projection); δr = continuous or small local changes in r;Δr = discontinuities in r; s = local surface orientation; δs = continuous or small local change in s.Δs = discontinuities in s.

Table 2.4: Output representations of the processes that derive surface information from images. Adapted from [13].

2.3.2 The immediate representation of visible surfaces.

The $2^1/_2$-D sketch provides a viewer-centered representation of the visible surfaces in which the results of all previously described processes can be expressed and combined. Its construction is a pivotal point, marking the last step before a surface's interpretation. The following discussion is concerned with three basic questions: First, what precisely is represented and how? Second, what precisely is the coordinate system? And third, which internal computations are carried out within the representation either to maintain its own internal consistency or to keep it consistent with what is allowed by the three-dimensional world?

What is represented and how?

Table 2.4 lists the types of information that the various early visual processes can extract from images. The interesting point is that although processes like stereopsis and motion are in principle capable of delivering depth information directly, they are in practice more likely to deliver information about local *changes* in depth, for example, by measuring local changes in disparity. Surface contours and shading provide more direct information about surface orientation. In addition, occlusion and brightness and size clues can deliver information about discontinuities in depth. The main function of the $2^1/_2$-D sketch is therefore not only to make explicit information about depth, local surface orientation, and discontinuities in these quantities but also to create and maintain a global representation of depth that is consistent with the local cues that these sources provide. It is likely that both surface orientation and depth are represented in the $2^1/_2$-D sketch, but although surface orientation can be represented quite accurately, depth is represented only roughly. Nevertheless, it might be possible that *local differences* in depth can be represented more accurately.

What is the coordinate system?

It was already observed that the coordinate system must be viewer-centered. The processes discussed so far are naturally retinocentric, so it is to be expected that the coordinate frame within which one is to express the results of each process most naturally will be retinocentric. Four observations weaken this argument. First, coordinates referring to the line of sight are not very useful to the viewer. Second, there are several different ways of representing surface orientation in a retinocentric coordinate frame, and the early visual processes may each use a slightly different one in which to express their own approximation of what the surface orientation actually is. Third, different parts of the visual field are analyzed at very different resolutions for a given direction of gaze. And fourth, because the early visual processes can run independently to a large extent, the question of consistency amongst the various types of information will arise. Inconsistency should be resolved as early as possible, because otherwise the information cannot be reduced to just one representation. In Marr's opinion, these observations lead to two conclusions. First, information from the various sources is probably checked for consistency and combined in some kind of retinocentric frame. Second, some conversion of the coordinate frame probably takes place at this point in order to express information from these processes in a standard form and probably also to allow for the angle of gaze. Such a conversion would (1) facilitate the computation of predicates like flat, convex, or concave; (2) allow easy comparison of the orientation of surfaces in different parts of the visual field; and (3) allow for eye movements.

What kind of computations are carried out to maintain consistency?

These computations are primarily concerned with discontinuities and interpolation. In an absolute sense, the resolution of the sample space does impose restrictions on what can be considered as a continuous change. If the sample values change too fast, the overall signal may exceed the bandwidth of the representation. If this occurs, then the representation is forced to attribute the change due to a discontinuity, since it is simply not rich enough to accommodate the changes that are actually occurring. The interpolation process in the human visual system, Marr argues, is conservative and reluctant to insert contours of discontinuity in depth or surface orientation unless the image itself provides reasonable evidence of their positions. However, the notion of surface continuity may give rise to various active computations in the $2^1/_2$-D sketch, including filling-in and the smooth continuation of discontinuities.

2.4 Representing shapes for recognition.

2.4.1 Form of the representation.

The viewer-centered coordinate frame, on which all representations discussed so far have been based because of their intimate connection with the imaging process, must now be abandoned. Object recognition demands a stable shape description that depends little, if at all, on the viewpoint. This, in turn, means that the pieces and articulation of a shape need to be described not relative to the viewer but relative to a frame of reference based on the shape itself. This implies that a canonical coordinate frame (a coordinate frame uniquely determined by the shape itself) must be set up within the object *before* its shape is described. Although formulating a general classification of shape representations is difficult, Marr states a number of main criteria by which they may be judged.

- *Accessibility*: can the desired description be computed from an image, and can this be done reasonably inexpensively? There are fundamental limitations to the information available in an image and the requirements of a representation have to fall within the limits of what is possible.

- *Scope and uniqueness*: what class of shapes is the representation designed for, and do the shapes in that class have canonical descriptions in the representation? If the representation is to be used for recognition, the shape description must also be unique; otherwise the difficult problem would arise of deciding whether two descriptions specify the same shape.

- *Stability and sensitivity*: to be useful for recognition, the similarity between two shapes must be reflected in their descriptions, but at the same time even subtle differences must be expressible. These opposing conditions can be satisfied only if it is possible to decouple stable information that captures the more general and less varying properties of a shape from information that is sensitive to the finer distinctions between shapes.

Using these criteria, three aspects of a representation's design are considered: (1) the representation's coordinate frame; (2) its primitives; and (3) the organization that the representation imposes on the information in its descriptions.

Coordinate systems.

For recognition tasks, viewer-centered descriptions are easier to produce but harder to use than object-centered ones, because viewer-centered descriptions depend upon the vantage point from which they are built. The alternative to relying on an exhaustive enumeration of all possible appearances is to use an object-centered coordinate system and thus to emphasize the computation of a canonical description that is independent of the vantage point. However, an object-centered description is more difficult to derive, since a unique coordinate system has to be defined for each object, and – as mentioned earlier – that coordinate system has to be identified from the image before the description is constructed.

Primitives.

There are two principal classes of shape primitives, surface-based (two-dimensional) and volumetric (three-dimensional). Volumetric primitives carry information about the spatial distribution of a shape. This type of information is more directly related to the requirements of shape recognition than information about a shape's surface structure, and this often means that much shorter and therefore more stable descriptions can still satisfy the sensitivity criterion.

Organization.

In the simplest case, no organization is imposed by the representation and all elements in a description have the same status. Alternatively, the primitive elements of a description can be organized into modules consisting, for example, of adjacent elements of roughly the same size, in order to distinguish certain groupings of the primitives from others. A modular organization is especially useful for recognition because it can make sensitivity and stability distinctions explicit if all constituents of a given module lie at roughly the same level of stability and sensitivity.

2.4.2 The 3-D model representation.

From these design requirements a limited representation, called the *3-D model representation*, can be formulated quite directly. First, in this representation a shape's object-centered coordinate system is based on axes determined by prominent geometrical characteristics of the shape, so the representation must be limited to those shapes for which this can be done. One large class of shapes satisfying this constraint is the generalized cones. In real life, a wide variety of common shapes is included in the scope of this representation, because objects are often described quite naturally in terms of one or more generalized cones. Second, the representation's primitives are also based on the natural axes of a shape. A description that uses these primitives can be thought of as a stick figure. Whilst only a limited amount of information about a shape is captured by such a description, that information is especially useful for recognition. And third, a modular decomposition of a description is achieved by basing the decomposition on the canonical axes of the shape. Each of these axes can be associated with a coarse spatial context that provides a natural grouping of the axes of the major shape components contained within that scope. Thus, the 3-D model representation specifies the following:

1. A model axis, which is the single axis defining the extent of the shape context of the model. This is a primitive of the representation, and it provides coarse information about characteristics such as size and orientation about the overall shape described.
2. Optionally, the relative spatial arrangement and sizes of the major component axes contained within the spatial context specified by the model axis. The number of component axes should be small and they should be roughly the same size.
3. The names (internal references) of 3-D models for the shape components associated with the component axes, whenever such models have been constructed. Their model axes correspond to the component axes of this 3-D model.

A distributed coordinate system, in which each 3-D model has its own coordinate system, is preferable. First, the spatial relations specified in a 3-D model description are always local to one of its models and should be given in a frame of reference determined by that model. Second, in addition to this stability and uniqueness consideration, the representation's accessibility and modularity is improved if each 3-D model maintains its own coordinate frame, because it can then be dealt with as a completely self-contained unit of shape description. An example of Marr's 3-D model representation is given in figure 2.3.

Deriving the 3-D model representation.

To construct a 3-D model, the model's coordinate system and component axes must be identified from an image, and the arrangement of the component axes in that coordinate system must be specified. Even if a shape has a canonical coordinate system and a natural decomposition into component axes, there still is the problem of deriving these features from an image. A major difficulty in the analysis of images arises when an important axis is

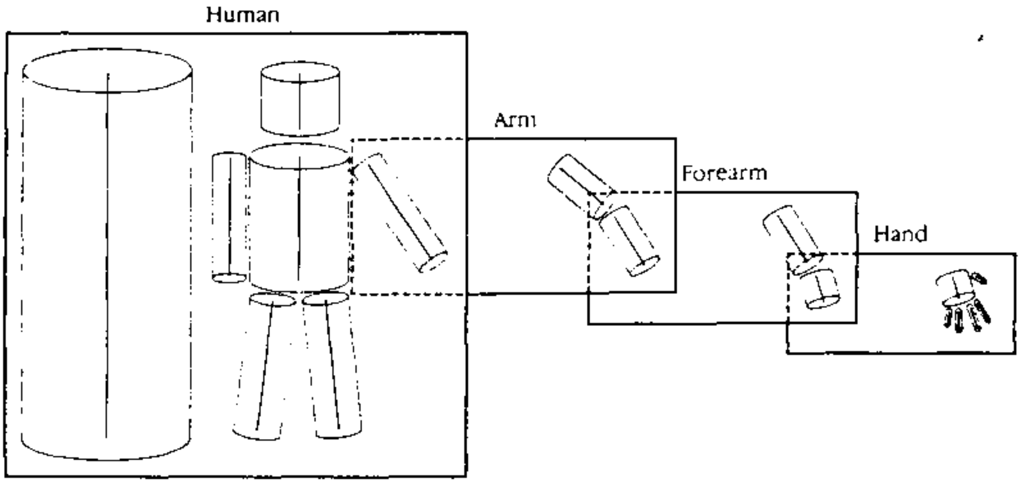


Figure 2.3: An example of the organization of shape information in Marr's 3-D model representation. Each box corresponds to a 3-D model, with its model axis on the left and the arrangement of its component axes on the right. Note that the relative arrangement of each model's component axes is shown improperly, since it should be in an object-centered system rather than the viewer-centered projection used here. Adapted from [13].

obscured because it is either foreshortened or hidden behind another part of the shape. There are three ways of dealing with such a situation. The first is to allow for recognition the use of partial descriptions based on the axes visible from the front. If this is done, the representation is slightly weakened in terms of the uniqueness criterion, but not as severely as a purely viewer-centered representation would be. Another strategy is to use a shape's visible components whenever their recognition is easy but that of the overall shape is difficult. These parts can be recognized directly and provide another route by which the shape can be recognized. Finally, a foreshortened axis can sometimes be found from an analysis of radial symmetry in the image. A local surface depth map like the $2^1/_2$-D sketch, computed by means of stereopsis, shading, or texture information, is likely to play an important role in interpreting images.

Techniques for finding axes in a two-dimensional image describe the locations of the axes in a viewer-centered coordinate system, and so a transformation is required to convert the specifications of the axes to an object-centered coordinate system. In the 3-D model representation, all axis dispositions are specified by adjunct relations, so a mechanism is required for computing an adjunct relation from the specification of two axes in a viewer-centered coordinate system. The accuracy of the computed adjunct relations is limited by the precision with which two axes are specified in the viewer-centered coordinate system. Since depth information is lost, the orientation specifications for axes derived from the retinal images are least precise in the amount the axes slant towards or away from the viewer. Axis slant parameters can often be reconstructed at least roughly by using stereopsis, shading, texture, structure-from-motion, and surface contour analysis. Constraints supplied by the recognition process can also be used to improve the precision of the slant specifications.

Using the 3-D model representation.

In Marr's opinion, recognition involves two things: a collection of stored 3-D model descriptions, the *catalogue of 3-D models*, and various indexes into the collection that allow a newly derived description to be associated with a description in the collection. Three access paths into the catalogue appear to be particularly useful. First, all 3-D models can be classified hierarchically according to the precision of the information they carry, and an index can be based on this classification called the *specifity index*. An example of this organization is given in figure 2.4. A newly derived 3-D model may be related to a model in the catalogue by starting at the top of the hierarchy and working down the levels through models whose shape specifications are consistent with the new model's until a level of specifity is reached that corresponds to the precision of the information in the new model. Second, once a 3-D model for a shape has been selected from the catalogue, its adjunct relations provide access to 3-D models for its components based on their locations, orientations, and relative sizes. This gives another access path to the models in the catalogue, called the *adjunct index*. The adjunct index provides useful defaults for the shapes of the components of a shape prior to the derivation of 3-D models for them in the image. It is also useful in situations where a catalogued model is not accessible via the specifity index because the description derived from the image is inadequate. The third access path considered, the *parent index*, is the inverse of the adjunct index. When a component of a shape is recognized, it can provide information about what the whole shape is likely to be. This index would play an important role in the situation where an important axis of a shape is obscured or foreshortened. According to Marr, the adjunct and parent indexes play a role secondary to that of the specifity index. Their purpose is primarily to provide contextual constraints that support the derivation process, for example, by indicating where the principal axes is likely to be when such information cannot be obtained from the image. Also, other indexes in the catalogue might be possible, for example based on colour or texture characteristics (e.g., think about the stripes of a zebra), or even non-visual clues (the sound an animal makes).

The overall recognition process may be summarized as follows. First a model from the catalogue is selected based on the distribution of components along the length of the principal axes. This model then provides relative orientation constraints that help to determine the absolute orientations (relative to the viewer) of the component axes in the image, and this information can be used to compute the relative lengths of the component axes. This new information can in turn be used to disambiguate shapes at the next level of the specifity index.

2.5 Concluding remarks.

Until the present day, Marr's theories continue to be extremely influential, and although some modifications concerning details of his algorithms have been suggested, his proposition of the basic structure of the visual system remains undisputed within the computational approach. However, research since Marr has concentrated primarily on early visual processing, thus leav-

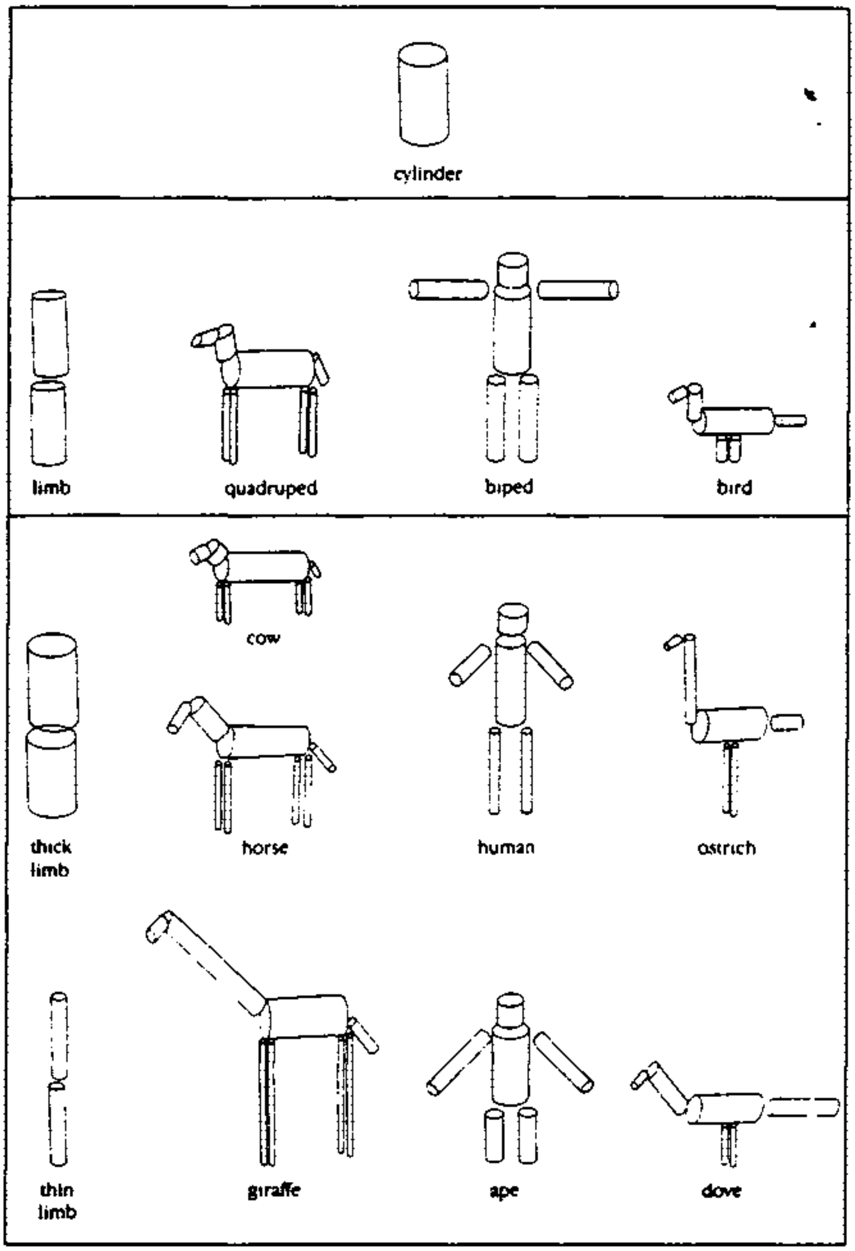


Figure 2.4: The catalogue of 3-D models, showing an organization according to specifity. Adapted from [13].

ing higher-level visual processing somewhat unexplored. Moreover, research effort has been directed more towards the level of the computational theory, at the penalty of spending less attention to the level of representation and algorithm and the level of hardware implementation.

The computational approach to visual perception results from an interdisciplinary effort of the scientific fields of visual perception, artificial intelligence and, in particular, computer vision. Indeed, the influence of computer vision on the computational approach is very strong, resulting in a strictly constrained, deterministic view on visual perception. Processing in the visual system is regarded as being first and foremost bottom-up, and top-down influences like goals and expectations of the observer are neglected almost completely. In addition, the image from which all the information processing starts is preeminently static. Innate biological behaviour, specifically exploratory movements such as directing one's gaze towards a potentially interesting part of the visual field, approaching an object of interest for a closer investigation, or moving one's head to and fro to get a better insight into the three-dimensional structure of the scene one is looking at, does not get the attention it ought to have. This is one important handicap of the scientific inheritage of the computational approach, and the primal cause to include the ideas of James J. Gibson in this report.

Chapter 3

Alternatives to Marr.

In this chapter two computational theories other than Marr's are considered. First, the ideas of H.G. Barrow and J.M. Tenenbaum [1] are discussed. Although, in general, their theory is very closely related to Marr's, it is directed even more towards the field of computer vision. Barrow and Tenenbaum propose that instead of combining the results of the processes decoding surface geometry immediately into a $2^1/_2$-D sketch, these results are written into a set of *intrinsic images*. This writing process succeeds in parallel, and corrections for maintaining consistency between the various intrinsic images are being made instantly.

Second, the *utilitarian approach*, primarily advocated by V.S. Ramachandran [16], is discussed shortly. According to Ramachandran, the visual system does not use the highly advanced signal-processing techniques of the type proposed by most computational theorists. Instead, it uses what is called a *bag of tricks*: a set of rules-of-thumb that evolved during millions of years simply because they proved to be useful. The simultaneous use of multiple parallel shortcuts, Ramachandran argues, allows a more rapid processing of images and a greater tolerance for noise than what would be possible with a single sophisticated algorithm.

Besides these two theories, other interesting literature worth reading includes M. Eimer's [3] philosophical discussion about the conditions that have to be satisfied in order for the human visual system to be characterized as an information-processing system, and about the way in which this information-processing might *itself* be explained. In [10] R. Jackendoff extends the computational approach towards a broader field, thus placing consciousness and memory at the point at which Marr ended. In [14] M.W. Matlin and H.J. Foley give a broad, introductory survey of theories regarding the visual system. In their discussion they include. besides the computational approach. Gibson's ecological approach, the Gestalt approach. and empiricism. Therefore, [14] gives a good extension to this report. Theories of visual perception related more directly to computer vision can be found in M.A. Fischler and O. Firschein [4]. and examples of typical computational theories regarding numerous aspects of visual perception can be found in M.S. Landy and J.A. Movshon [11]. Finally. K. Nakayama [15]. S.J. Thorpe [17] and R.J. Watt [18. 19] give an impression of some of the ideas that have been developed since Marr.

3.1 Barrow and Tenenbaum's intrinsic images.

3.1.1 The concept of intrinsic images.

Barrow and Tenenbaum's proposition of the structure of the visual system [1] is very closely related to Marr's. The most important deviation concerns the combination of the output from the processes that decode surface geometry into the $2^1/_2$-D sketch. Instead of combining this output *directly* into a $2^1/_2$-D sketch, Barrow and Tenenbaum propose using a set of *intrinsic images* instead. Thus. intrinsic characteristics of surfaces are represented as a set of images in parallel. each image containing values and explicit indication of discontinuities in value or gradient for a particular characteristic of the surface element visible at the corresponding location in the original image. The primary intrinsic images are those of surface reflectance, surface orientation. and incident illumination. Also. range, transparency, and specularity might be useful as intrinsic images. Note that the intrinsic images of distance and orientation together correspond to Marr's $2^1/_2$-D sketch. An example set of intrinsic images is given in figure 3.1.

3.1.2 Applicability of the processes deriving surface characteristics.

The processes that derive surface characteristics from image features (e.g., see 2.3.1) each provide absolute or relative values for a particular scene characteristic. However, the applicability of each process is restricted by its dependence upon certain assumptions about the nature of the visual world and the scene that is observed. In general, therefore, multiple processes will be needed to obtain a complete description, and the problem emerges of determining which processes are applicable in each region of the image, and how to integrate the output of these processes into a consistent global

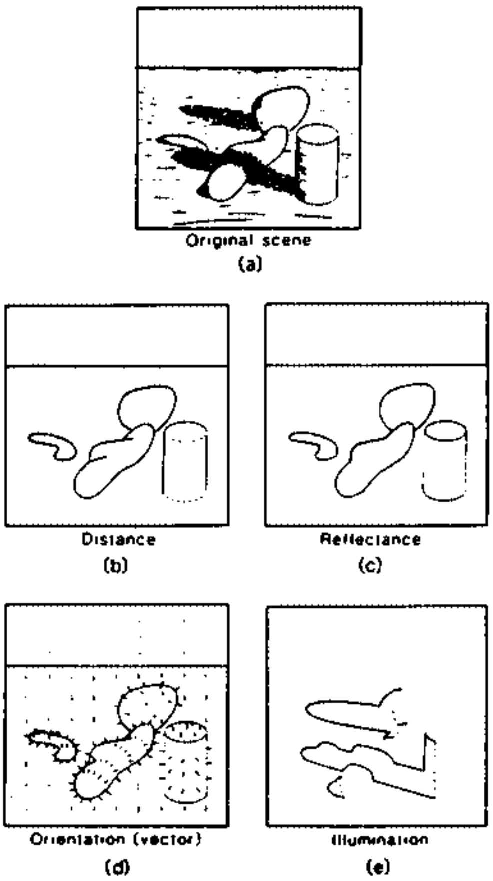
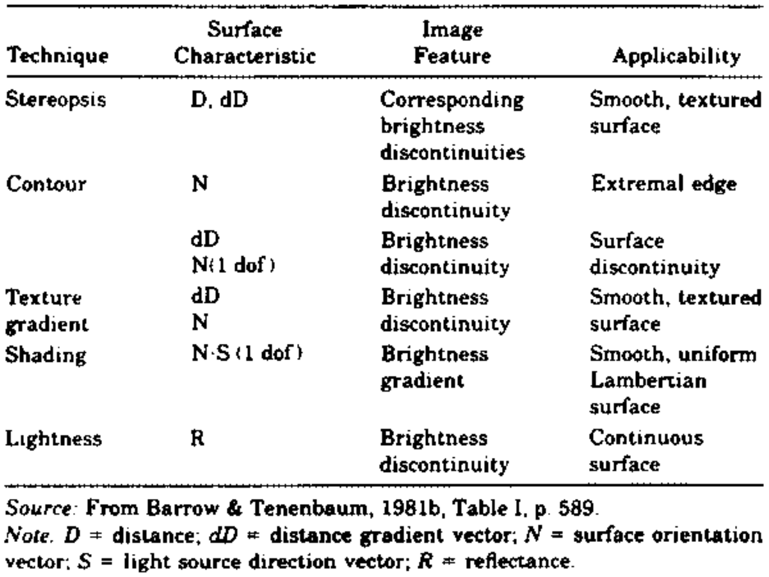


Figure 3.1: A set of intrinsic images. Solid lines represent discontinuities in the depicted scene characteristic, and dashed lines represent discontinuities in its derivative. Adapted from [1].



Technique	Surface Characteristic	Image Feature	Applicability
Stereopsis	D, dD	Corresponding brightness discontinuities	Smooth, textured surface
Contour	N	Brightness discontinuity	Extremal edge
	dD N(1 dof)	Brightness discontinuity	Surface discontinuity
Texture gradient	dD N	Brightness discontinuity	Smooth, textured surface
Shading	N·S (1 dof)	Brightness gradient	Smooth, uniform Lambertian surface
Lightness	R	Brightness discontinuity	Continuous surface

Source: From Barrow & Tenenbaum, 1981b, Table I, p. 589.
Note. D = distance; dD = distance gradient vector; N = surface orientation vector; S = light source direction vector; R = reflectance.

Table 3.1: The processes that derive surface characteristics from image features, the characteristics they derive, what image features they exploit, and their applicability. Adapted from [1].

interpretation. To solve this problem, it is necessary to determine which physical characteristics are, in fact, responsible for an observed intensity variation and which are discontinuous across intensity edges. The applicability of the various processes, as proposed by Barrow and Tenenbaum, is listed in table 3.1.

3.1.3 An implementation of the theory of intrinsic images.

In Barrow and Tenenbaum's opinion, the simultaneous recovery of the primary intrinsic characteristics from a brightness image succeeds in four steps:

1. Find the brightness discontinuities in the input image.
2. Determine the physical nature of the discontinuity.
3. Assign boundary values for intrinsic characteristics along the edges, based on the physical interpretation.
4. Propagate from these boundary values into the interiors of smoothly shaded regions, using continuity assumptions.

The proposed computation depends on local processes only, and can be performed rapidly in parallel. A possible implementation is given in figure 3.2. Essentially, it consists of the original image on top of a stack of intrinsic images. Processing is initialized by detecting intensity edges in the original image, interpreting them, and then creating the appropriate edges in the intrinsic images. Parallel local operations (shown as circles) modify the values in each intrinsic image to make them consistent with intra-image continuity and limit constraints. Simultaneously, a second set of processes (shown as vertical lines) operates to make the values at each point consistent with the corresponding intensity value. A third set of processes (shown as X's) operate to insert and delete edge elements, which inhibit continuity constraints locally. The constraint and edge modification processes operate continuously and interact to recover accurate intrinsic scene characteristics and to perfect the initial edge interpretation.

3.1.4 To conclude.

The implementation described above shows the possibility of simultaneously recovering orientation, reflectance, and illumination from a single monochrome image without recourse to primary depth cues such as stereopsis, motion parallax, or texture gradient. Nevertheless, such additional cues can be added to aid initialization, and would be particularly useful in shadowed areas or in areas of complex illumination and reflectance functions.

The recovery of intrinsic characteristics provides a set of intrinsic images that describe the scene on a point-by-point basis in a viewer-centered coordinate frame. Higher levels of processing, such as object recognition, require that this information be reorganized to provide a more concise symbolic representation that captures global properties in a viewpoint-independent coordinate frame. Barrow and Tenenbaum's proposition for this stage of vision is similar to Marr's 3-D model representation.

3.2 Ramachandran's utilitarian approach.

3.2.1 Machine vision versus biological vision.

According to Ramachandran, visual perception may be defined as a biological process whose goal it is to rapidly compute a three-dimensional representation of the world that the organism can use for navigation and for object manipulation. The visual image is inherently ambiguous so perception is essentially a matter of resolving ambiguities by using knowledge of

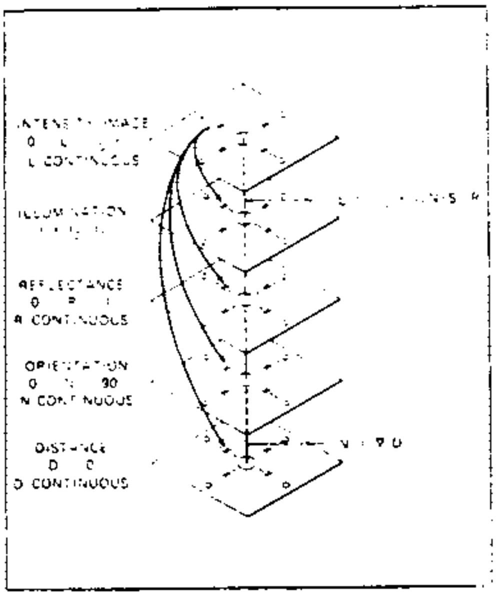


Figure 3.2: An implementation of a parallel computation for recovering intrinsic image characteristics. Adapted from [1].

the external world. Therefore, the visual system can be regarded as an information-processing system.

The first step towards understanding any complex information-processing system is to clearly identify the problems it was designed to solve. In Ramachandran's opinion, the computational approach to vision has been extremely useful in this regard because it allows a much more rigorous formulation of perceptual problems than what would be possible with psychophysics of physiology alone. However, simulation of a biological system does not necessarily reveal how the system actually works, since biological vision differs in several important respects from machine vision:

1. Considerations of optimality have obvious importance in engineering but they have only a limited role in biology. The goal of biological visual systems is to rapidly compute approximate solutions to perceptual problems. The solutions are always adequate for the situation at hand, but rarely optimal.

2. The constraints imposed by the environment (natural constraints) reduce the computational burden on the visual system but they do not impose a unique solution to perceptual problems. There are often far

too many ways of solving a problem theoretically; the only way to distinguish between them is by means of psychophysics and neuroanatomy.

3. The central dogma of computational vision has been that the strategies used by any complex information-processing system can be understood independent of hardware implementation. Contrary to this, Ramachandran argues that biological vision is strongly constrained by the actual neural machinery that mediates it. There may be certain things that neurons simply cannot accomplish and this automatically eliminates a wide range of theoretically plausible solutions.

4. In science one typically wants to understand the whole lineage of causes that governs a given phenomenon - not merely the remote ancestral cause. In this sense, specifying natural constraints, although useful, is an incomplete account of visual processes since it doesn't explain exactly how a given constraint is actually exploited.

5. Biological systems were not designed from scratch - they often had to be built from preexisting implementations. Nature is inherently opportunistic and will often adopt ad hoc solutions that may actually seem very inelegant to an engineer.

6. For any given perceptual problem biological systems often seem to use multiple parallel mechanisms. The perception of three-dimensional shapes, for example, is made possible by stereopsis, occlusion, shading, relative motion, et cetera. Why use multiple mechanisms when a single one will suffice on computational grounds? There are at least two reasons. First, by using multiple strategies for any one problem, the system can get away with each of them being relatively crude and, therefore, easy to implement. Second, the simultaneous use of multiple parallel shortcuts allows more rapid processing of images and a greater tolerance for noise than what would be possible with a single sophisticated algorithm. It is this remarkable tolerance for noisy (sometimes even camouflaged) images that characterizes biological vision and discerns it from machine vision.

3.2.2 The utilitarian theory of perception.

According to this theory, perception does not involve intelligent reasoning, nor does it involve resonance with the world or does it require creating elaborate internal representations or solving equations. Perception is essentially a *bag of tricks*. Through millions of years of trial and error the visual system has evolved numerous shortcuts, rules-of-thumb and heuristics which were adopted not for their aesthetic appeal or mathematical elegance but simply because they worked; the visual system uses a bewildering array of special-purpose tricks and adaptive heuristics to solve its problems. The mechanisms used are in part the result of an interaction between the organism's adaptive needs and certain natural constraints, but also in part due to the organism's evolutionary history. Although it is possible to provide post hoc rationalizations for these mechanisms, it seems highly unlikely that any of them could have been deduced from first principles. Therefore, in understanding vision a bottom-up research strategy is just as important as a strictly top-down approach.

One general remark that can be made about perception concerns its biological goal - that it has to be extremely rapid and highly tolerant to noisy inputs. Through millions of years of trial and error the organism has learned that the best way to achieve this is to simultaneously deploy multiple parallel shortcuts or heuristics for each perceptual problem. Thus, Ramachandran argues, the task of vision researchers is to unravel the internal logic of these mechanisms - preferably at all three levels prescribed by Marr - and to discover how they interact with each other to generate an observer's perceptual experience of the world.

Chapter 4

The ecological approach.

Although this report is primarily concerned with the *structure* of the visual system, in this chapter a survey is given of one major influential theory of *direct perception*: James J. Gibson's ecological approach to visual perception [5, 6, 7]. As already mentioned in the introduction, the purpose of including Gibson's theories in this report is *not* to confront the theories of direct and indirect perception. Instead, Gibson's ideas are included to give an extension to the somewhat limited scope of the computational approach with regard to the everyday environment surrounding the observer and the interaction between vision and action.

In contrast to Marr's work in perception spanning only a few years, Gibson was active in the field of visual perception for nearly four decades. During World War II his prime research concerned visually guided behaviour, specifically with respect to taking off and landing aircraft. His research led him to regard perception first and foremost as a biologically adaptive activity. Therefore, theories on perception should do justice to the everyday perceptual accomplishments that contribute to the survival of the species. It was this conviction that finally brought Gibson to formulate his ecological approach to visual perception.

Besides Gibson's own work [5, 6, 7], two references for further reading are particularly worth mentioning. First, an interesting survey of the history of Gibson's ideas can be found in an essay by W.M. Mace [12]. Its title, "James J. Gibson's strategy for perceiving: Ask not what's inside your head, but what your head's inside of.", is an excellent one-line summary of Gibson's philosophy. And second, a discussion of Gibson's ecological approach from the point of view of biological vision can be found in V. Bruce and P.R. Green [2].

4.1 Gibson's early work.

4.1.1 The Ground Theory of visual perception.

During World War II, Gibson studied the perception of motion in space during flight, specifically during the takeoff and landing of aircraft. This research convinced him that information from the ground and the sky played an important role in motion perception, and that this information lay in the optic flow of textures which arises as a result of motion relative to the ground and the clouds. This finding led him to formulate his Ground Theory of visual perception. According to the Ground Theory, visual space should be conceived not as an object or an array of objects in air but as a *continuous surface or an array of adjoining surfaces*; the spatial character of the visual world is given not by the objects in it but by the background of objects. Stated more specifically, the central assumptions of Gibson's Ground Theory are:

1. The fundamental condition for seeing a visual world is an array of physical surfaces reflecting light and projected on the retina.

2. In any environment, these surfaces are of two extreme types, frontal (i.e., transverse to the line of sight) and longitudinal (parallel with the line of sight).

3. The perception of depth and distance is reducible to the problem of the perception of longitudinal surfaces.

4. The general condition for the perception of a surface is the type of stimulation which yields texture.

5. The general condition for the perception of an edge, and hence for the perception of a bounded surface in the visual field, is the type of stimulation consisting of an abrupt transition.

6. The perception of an object in depth is reducible to the problem of the changing slant of a curved surface or the differing slant of a bent surface.

7. The general condition for the perception of a longitudinal or slanted surface is a kind of stimulation called a gradient. The gradient of texture, gradients dependent on outlines, the gradient of retinal disparity, gradients of shading, the gradient of deformation when the observer moves, and possibly others, all attribute to the impression of distance on a surface.

4.1.2 The richness of the visual stimulus.

The Ground Theory was first published in [5]. In this work, Gibson emphasizes the *richness* of the visual stimulus: the visual image is adequately rich in information to account for the depth and distance of the visual world without the necessity of supposing a special mental process to supplement it. According to Gibson, the basis of the perception of space is the projection of its objects and elements as an image, and the consequent gradual change of size and density in the image as the objects and elements recede from the observer. Thus, the gradient of density in a projection of a physical surface bears a fixed relation to the slant and facing of the physical surface projected. Likewise, the gradient of binocular disparity bears a fixed relation to the distance of the surface projected. And the illumination of a given section of surface is a function of the orientation of the surface towards or away from the source of light. Variations in texture and size, in binocular disparity, and in shading should therefore be in exact geometrical correspondence with the dimensions of the physical world and, most importantly, they should yield corresponding variations in perceptual experience. Gibson concludes these findings in summarizing eight varieties of *perspective* related to the perception of distance over a surface or an array of surfaces, and five varieties of *sensory shift* related to the perception of depth at a contour:

1. Texture-perspective: a gradual increase in texture density with increasing distance.

2. Size-perspective: a gradual decrease in the apparent size of shapes with increasing distance.

3. Linear perspective: a gradual decrease in the spacing between either outlines or inlines of rectilinear contours with increasing distance.

4. Binocular perspective: the gradual decrease in horizontal skew of one retinal image relative to the other with increasing distance.

5. Motion perspective: the gradual decrease in the rate of displacement of texture elements or contours with increasing distance.

6. Aerial perspective: a gradual increase in haziness, blueness, and desaturation of colours with increasing distance.

7. The perspective of blur: the gradual increase of blur with increasing distance to the plane of focus.

8. Relative upward location in the visual field: the gradual decrease in apparent distance from the horizon with increasing distance.

9. Shift of texture density or linear spacing: a sudden change in texture density or spacing between outlines or inlines resulting from a sudden change in distance.

10. Shift in the amount of binocular disparity: a sudden change in horizontal skew of one retinal image relative to the other resulting from a sudden change in distance.

11. Shift in the rate of displacement: a sudden change in the rate of displacement resulting from a sudden change in distance.

12. Completeness of outline: completeness of a visual contour tends to be associated with the near side of a common contour and incompleteness with the far side.

13. Transitions between light and shade: these seem to be capable of giving a surface the quality of shape in depth.

Texture-perspective, size-perspective and linear perspective are the perspectives of position. and binocular and motion perspective together constitute the perspectives of parallax. whereas aerial perspective. the perspective of blur and the relative upward location in the visual field are independent of an observer's motion or position. Completeness of outline is related to the superposition of objects and might be a result of *visual experience*. Finally. the relationship between transitions of shading and the corresponding depth-shape is complicated and at present not well understood.

Gibson's central statement in [5] is that the objective world does not require for its explanation a process of construction. translation. or even organization. The visual world can be analyzed into impressions which are object-like. and these impressions are traceable to stimulation. The fundamental impressions obtained by introspection, Gibson argues. are contour. surface. slant. corner. motion. distance, and depth, in addition to colour, all of which correspond to the variables of a distribution of focused light. In Gibson's opinion, these impressions do not require any putting together since the togetherness exists on the retina. His conclusion is "[...] that order exists in stimulation as well as in experience; that order is just as much physical as mental." [5].

4.1.3 Stimulus invariants and perceptual systems.

A further development of these ideas appears in [6]. In this work, Gibson comes to regard the senses no longer as the transient channels of sensations but as dedicated *perceptual systems* that during the animal's evolution have so developed that they got precisely tuned to pick up the vital information available in the everyday environment. This information is specified in the *invariants* of visual stimulation. Invariants can be thought of as higher-order properties of patterns of stimulation which remain constant during changes associated with the observer, the environment, or both. Imagine, for example, moving in a natural environment: the changes in the.patterns of texture one observes will seem chaotic and complex. Nevertheless there is a higher-order property underlying these changes: the optical flow (see figure 4.1). Thus, optical flow is the invariant of an observer's motion. Another example, an invariant possibly underlying size constancy, is given in figure 4.2. In his late work, Gibson comes to distinguish types of stimulus invariants, such as invariants of optical structure under changing illumination, invariants of optical structure under change of the point of observation, and invariants of the process of looking around.

As a consequence of regarding the senses as perceptual systems, the brain is relieved of the necessity of constructing the information contained within the visual stimulus by *any* process - be it innate rational powers (according to theoretical nativism), memory (according to empiricism), or form-fields (according to Gestalt theory). The brain can be treated as the highest of several centers of the nervous system, governing the perceptual systems. So, instead of postulating that the brain constructs information from the input of a sensory nerve, Gibson proposes that the centers of the nervous system, including the brain, *resonate* to information. The reason for this resonance to occur being that active perceptual systems, as contrasted with passive receptors, are precisely tuned to pick up the invariants from information available in the everyday environment.

4.2 Gibson's late work.

The preceding two sections were included in this report to give an introduction to Gibson's ecological approach, since the central ideas of this approach are closely related to those in Gibson's earlier work. However, in Gibson's late work [7] emphasis shifted more towards the relationship between the perceiving organism and the environment surrounding it, hence its name *ecological* approach.

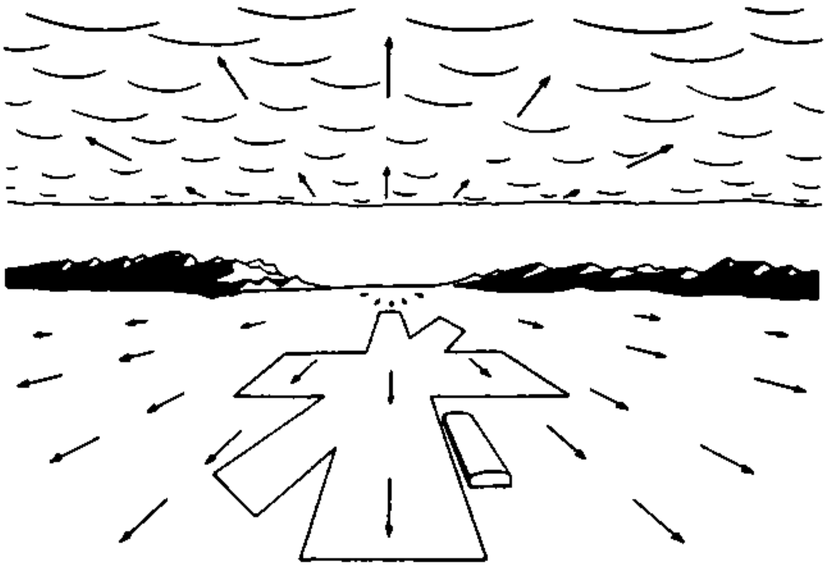


Figure 4.1: Optic flow during the landing of an aircraft. The point of alight is indicated by the center of expansion. The conclusion to be drawn from this example is that, without further action, this aircraft is bound to pass over the runway. Adapted from [8].

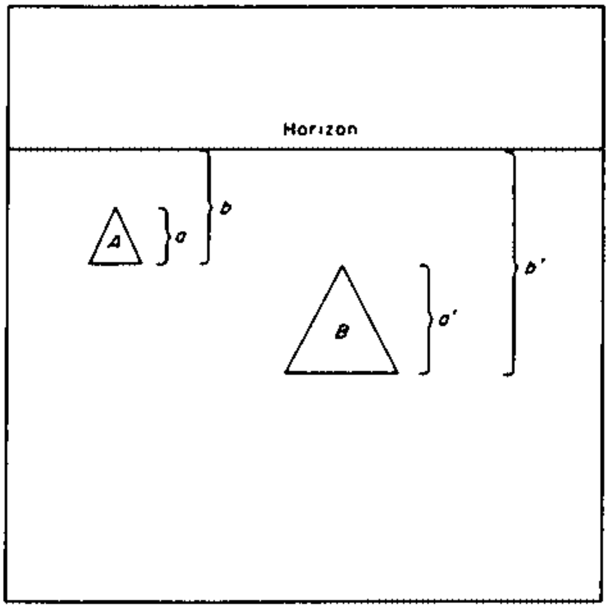


Figure 4.2: A candidate invariant underlying size constancy. If $a : b = a' : b'$ then A and B are the same size. Adapted from [8].

4.2.1 The necessity of an ecological approach.

To understand what an animal's perceptual systems are able to accomplish, the environment in which they evolved must be considered, since it is precisely the same environment which shaped these systems. Consider, for example, an object moving away from an observer, its retinal image – in accordance with classical optics – steadily getting smaller. Under ordinary circumstances observers will *never* report seeing a shrinking object; instead, observers will always correctly report seeing an object moving away. Classical optics alone cannot account for this feat.

However, receding objects do not just produce shrinking retinal images. Most objects have textured surfaces and this texture gets finer as the object moves away. Objects obscure a portion of the textured ground against which they are seen. and this portion gets smaller as the object recedes. The further away an object is, the closer it will be to the horizon, so a receding object will seem to move towards the horizon. Hence, what is needed is a new kind of optics which addresses to these effects: an *ecological optics*. capable of explaining everyday vision, which takes into account the environment surrounding the observer.

4.2.2 Natural vision.

Traditionally, vision is told to depend on the eye, which is connected to the brain. In Gibson's opinion however, natural vision depends on the eyes in the head on a body supported by the ground, the brain being only the central organ of a complete visual system. When no constraints are put on the visual system, people look around, walk up to something interesting and move around it so to see it from all sides. Looking around and getting around do not fit into the standard idea of what visual perception is. But, according to Gibson, the single, frozen field of view provides only impoverished information about the world; the visual system did not evolve for this. Moreover, visual awareness is panoramic and does in fact persist during long acts of locomotion.

4.2.3 The fundamentals of the ecological approach.

In this subsection a short survey is given of the fundamentals of the ecological approach; any discussion going more into the details of this approach would almost certainly move beyond the scope of this report.

According to Gibson, the ecological approach starts with considering the visual stimulus of an observer who walks from one point of observation to another, who moves around an object of interest, and who can approach such an object for a closer inspection - thereby extracting the invariants that underlie the changing perspective structure and seeing the connections between hidden and unhidden surfaces - and with considering panoramic visual awareness. Only then is the awareness of a single scene considered, the surfaces seen with the head fixed and the visual image frozen. The classical puzzles that arise with this kind of vision are resolved by recognizing that the invariants are weaker and the ambiguities stronger when the point of observation is motionless.

Finally, the kind of visual awareness obtained in laboratory experiments - with the head fixed and the retina either briefly exposed or made to stay fixed - is considered, in Gibson's opinion "[...] a peculiar result of trying to make the eye work as if it were a camera at the end of a nerve cable.". Even at this photographic level the visual system continues to operate, but the constraints imposed on it are so severe that very little information can be picked up. Should the perception of the environment truly be based on this kind of glimpses, then it *has* to be a process of construction. If the data are insufficient, the observer must go beyond the data. Gibson gives no clear answer as to how this should be done. Nevertheless, Gibson argues, explanations of perception based on sensory inputs must fail because they can all be reduced to this: in order to perceive the world, one must already have ideas about it. And knowledge of the world is explained by assuming that knowledge of the world exists. In Gibson's opinion, whether the ideas are acquired or innate makes no difference; the fundamental failure of these explanations lies in their circular reasoning.

But if, following Gibson's reasoning, perception of the environment is not based on a sequence of snapshots but on invariant-extraction from a continuous flux, one does not need to have ideas about the environment in order to perceive it. The information for the perception of an object is not its image; the information in light to specify something does not have to resemble it, or copy it, or even be an exact projection. Gibson puts it even stronger: "*Nothing* is copied in the light to the eye of an observer, not the shape of a thing, not the surface of it, not its substance, not its colour, and certainly not its motion. But all these things are *specified* in the light.".

Chapter 5

Concluding remarks.

Apart from their fundamental difference in the one being a theory of direct and the other a theory of indirect perception, the ecological and the computational approach to visual perception do have much in common. Both start explaining visual perception by investigating the properties of the everyday visual environment. As a result, both acknowledge the richness of the visual stimulus. Both state that it is constrained by the environment and that, albeit through processes as different as construction and resonance, the visual system somehow uses these constraints.

Definitely, differences exist concerning the way in which the visual system is supposed to use these constraints, the computational approach being directed more towards general physical knowledge of the visual world, and the ecological approach more towards the invariants available in the everyday environment. Nevertheless, these differences may perhaps be subtle, since Gibson's invariants can at least partly be explained in terms of the general physical knowledge advanced by the computational approach and, inversely, an extension of this knowledge could lead to the formulation of some kind of invariants in everyday visual stimulation.

Both approaches have their weaknesses. The ecological approach is preoccupied in regarding the richness of the visual stimulus as the single possible explanation of the accomplishments of visual perception. Therefore, in the case of a degraded visual stimulus – for example when looking at an image – the ecological approach can no longer fully account for the rich perception of the depicted scene. The computational approach, however, limits itself too much to the static image, thus overlooking the enormous richness of the visual stimulus in everyday life, and condemning itself to the development of a laborious set of dedicated information-processing tasks in order to retrieve as much information as possible from the visual stimulus. A less limited scope, for example by examining the visual stimulus as occurring in everyday life and by including factors like exploratory behaviour, could mean an important enhancement to the already accumulated knowledge of the visual world, thereby lessening the urgent need for highly sophisticated algorithms and at the same time developing a less mechanistic, more natural vision of the way in which we perceive the world around us.

Appendix A

Further reading.

This report of course cannot give a complete account of all existing theories of visual perception. Amongst the theories not mentioned in this report are extremely influential approaches like the Empiricist approach and the Gestalt approach. There is, however, one reference that I would *really* like to recommend to anyone interested in a more thorough discussion of the approaches to visual perception that during history have been pursued. It is I.E. Gordon's book "Theories of visual perception" [8], available in IPO's library, code JF238.

As can be concluded from the portion of this report dedicated to his theory, D. Marr's "Vision: a computational investigation into the human representation and processing of visual information" [13] is another reference I truly recommend, since this report barely presented the skeleton of his ideas. "Vision" is available in IPO's library, code JF144.

And finally, I would like to recommend Gibson's works "The perception of the visual world" [5], "The senses considered as perceptual systems" [6] and "The ecological approach to visual perception" [7], since this report could not possibly give a complete survey of Gibson's ideas. For example, the *affordance*, Gibson's most radical and most controversial concept, has not been attended to. Whilst [5] and [6] are available in IPO's library, codes JF26 and PD54 respectively, [7] is only available in the library of the faculty Technology Management of the EUT, code PKN79GIB.

Bibliography

[1] H.G. Barrow and J.M. Tenenbaum. Computational approaches to vision. In K.R. Boff, L. Kaufman, and J.P. Thomas, editors, *Handbook of perception and human performance*, pages 38.1 - 38.70. Wiley-Interscience, Chichester, 1986.

[2] V. Bruce and P.R. Green. *Visual perception: physiology, psychology and ecology*. Erlbaum, London, first edition, 1985.

[3] M. Eimer. Representational content and computation in the human visual system. *Psychological Research*, 52:238 - 242, November 1990.

[4] M.A. Fischler and O. Firschein, editors. *Readings in computer vision: issues, problems, principles, and paradigms*. Kaufman, Los Altos, 1987.

[5] J.J. Gibson. *The perception of the visual world*. Allen and Unwin, London, 1950.

[6] J.J. Gibson. *The senses considered as perceptual systems*. Houghton Mifflin, London, 1966.

[7] J.J. Gibson. *The ecological approach to visual perception*. Houghton Mifflin, London, 1979.

[8] I.E. Gordon. *Theories of visual perception*. Wiley, New York, 1994.

[9] E.C. Hildreth and S. Ullman. The computational study of vision. In M.I. Posner, editor, *Foundations of cognitive science*, pages 581 - 630. MIT Press, London, 1989.

[10] R. Jackendoff. *Consciousness and the computational mind*. MIT Press, Cambridge, MA, 1987.

[11] M.S. Landy and J.A. Movshon, editors. *Computational models of visual processing*. MIT Press, Cambridge, MA, 1991.

[12] W.M. Mace. James j. gibson's strategy for perceiving: ask not what's inside your head, but what your head's inside of. In R. Shaw and J. Bransford, editors, *Perceiving, acting and knowing: towards an ecological psychology*, pages 43 – 66. Erlbaum, Hove, 1977.

[13] D. Marr. *Vision: a computational investigation into the human representation and processing of visual information.* Freeman, San Fransisco, 1982.

[14] M.W. Matlin and H.J. Foley. *Sensation and perception.* Allyn and Bacon, Boston, third edition, 1991.

[15] K. Nakayama. The iconic bottleneck and the tenuous link between early visual processing and perception. In C. Blakemore, editor, *Vision: coding and efficiency*, pages 411 – 422. Cambridge University Press, Cambridge, 1990.

[16] V.S. Ramachandran. Visual perception in people and machines. In A. Blake and T. Troscianko, editors, *AI and the eye: 11th European conference on visual perception, Bristol, 1988*, pages 21 – 77, Chichester, 1990. Wiley.

[17] S.J. Thorpe. Image processing by the human visual system. Technical Report 4, Eurographics, Montreux, September 1990.

[18] R.J. Watt. *Visual processing: computational, psychophysical and cognitive research.* Erlbaum, Hove, 1989.

[19] R.J. Watt. *Understanding vision.* Academic Press, London, 1991.