(1)

STANFORD ARTIFICIAL INTELLIGENCE PROJECT
MEMO AIM-134

COMPUTER SCIENCE DEPARTMENT
REPORT No. CS 182

ACCOMMODATION IN COMPUTER VISION

by

Jay Martin Tenenbaum

October 1970

COMPUTER SCIENCE DEPARTMENT

STANFORD UNIVERSITY

ACCOMMODATION IN COMPUTER VISION

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Jay Martin Tenenbaum

September 1970

STANFORD ARTIFICIAL INTELLIGENCE PROJECT
MEMO AIM-134

OCTOBER 1970

COMPUTER SCIENCE DEPARTMENT REPORT
NO. CS 182

ACCOMMODATION IN COMPUTER VISION

BY

Jay Martin Tenenbaum

ABSTRACT: We describe an evolving computer vision system in which the parameters of the camera are controlled by the computer. It is distinguished from conventional picture processing systems by the fact that sensor accommodation is automatic and treated as an integral part of the recognition process.

A machine, like a person, comes in contact with far more visual information than it can process. Furthermore, no physical sensor can simultaneously provide information about the full range of the environment. Consequently, both man and machine must accommodate their sensors to emphasize selected characteristics of the environment.

Accommodation improves the reliability and efficiency of machine perception by matching the information provided by the sensor with that required by specific perceptual functions. The advantages of accommodation are demonstrated in the context of five key functions in computer vision: acquisition, contour following, verifying the presence of an expected edge, range-finding, and color recognition.

We have modeled the interaction of camera parameters with scene characteristics to determine the composition of an image. Using a priori knowledge of the environment, the camera is tuned to satisfy the information requirements of a particular task.

Task performance depends implicitly on the appropriateness of available information. If a function fails to perform as expected, and if this failure is attributable to a specific image deficiency, then the relevant accommodation parameters can be refined.

This schema for automating sensor accommodating can be applied in a variety of perceptual domains.

The research reported here was supported in part by the Advanced Research Projects Agency of the Office of the Secretary of Defense (SD 183).

ACKNOWLEDGMENTS

One of the more pleasant aspects of finishing a dissertation is this opportunity to acknowledge the many people whose contributions made it possible.

I would like first to thank my advisor, Professor Jerome Feldman. His wide ranging interests, breadth of knowledge, and intriguing conjectures contributed in many ways to the progress of this research. I am especially appreciative of his style of guidance; he encourages students to widen and then explore their own interests.

I am deeply indebted to Dr. C.E. Duncan. His commitment to graduate education and his confidence in my work helped me surmount the inevitable frustrations that befall a graduate student.

All computer scientists have benefitted from the foresight of Professor John McCarthy, director of the Artificial Intelligence Project. I, in particular, thank Dr. McCarthy for my first introduction to computer science, ten years ago at MIT, and for the stimulating environment of the research program he subsequently established at Stanford.

I am appreciative of the interest shown by the members of my reading committee. Professor Thomas Cover has provided wise counsel throughout the course of my graduate work. He was especially influential in helping me define my research interests. Professor Von Eschleman suggested

interesting applications of this work to the problem of information reduction associated with remote sensing from outer space.

I am grateful to Dr. Tom Binford, who carefully scrutinized this text. His suggestions contributed significantly to the cogency and clarity of the ideas presented in this thesis.

The evolution of any work is inherently shaped by daily interaction with one's colleagues. My association with the computer research programs at both Lockheed and Stanford provided a rare opportunity for cross-fertilization of ideas. I would like to acknowledge the particular contributions of Dr. Martin Fischler, Oscar Firschein, Roy Merrill, and Ken Last at Lockheed, and of Dr's. Alan Kay, Irwin Sobel and Michael Kelly at Stanford.

Dr. Fischler provided my initial introduction to the fascinating problems of machine vision. His extensive experience in this area has been invaluable. Many of the implications of this work for vision system strategies first arose in conversations with Dr. Kay. At Stanford, my work was most closely related to that of Dr. Sobel. He was consequently enlisted on many occasions as a sounding (and "ironing") board for emerging ideas. He served this role well. More importantly, both he and Alan Kay have become my friends. Dr. Kelly provided useful insights regarding the need for accommodation in a more general vision system.

My work was dependent on the programming and system support of many people, notably: Les Earnest, Dr. Gil Falk, Aharon Gil, Jerry Gleason, Dr. Manfred Heuckel, Dick Helliwell, Ed McGuire, Ted Panofsky, Karl Pingle, Lynn Quam, Bob Sproull, Dan Swinehart, and Joe Zinghelm.

Norm Briggs, Les Earnest and Don Schaaf helped in an administrative capacity to expedite completion of this work.

The appearance of this manuscript testifies to the competence of many people who contributed to the publication effort. I am grateful to Margret Collins, Virginia Whipple, and Phyllis Winkler. I will remember most, however, what the reader cannot see--- the dependability and moral support extended by these individuals in a time of storm and stress.

To my wife Bonnie, I dedicate this dissertation. She has participated in this endeavor since its inception, working with dedication to ensure its success. Her contributions extend far beyond the vital supportive role of a loving wife. She has been intimately involved in the supplementary research and preparation of this thesis, complementing my own efforts on an almost round the clock schedule during the past four months. The anxieties, frustrations, and satisfactions which we have shared in this project has added new dimensions to a very deep relationship. Thank you, Bonnie.

The research reported here was supported in part by the Advanced Research Projects Agency of the Office of the Department of Defense and in part by the Lockheed Missiles & Space Company Independent Research Program.

TABLE OF CONTENTS

LIST OF CHARTS

LIST OF FIGURES

CHAPTER I: INTRODUCTION

This thesis considers the problem of adjusting the characteristics of a visual sensor to obtain the most appropriate image for a specific perceptual task. The goal is to improve the reliability and (to a lesser extent) the efficiency of machine perception by matching the information provided by the sensor with that required by the task. A specific context for this work was provided by the vision requirements of the hand-eye project at the Stanford Artificial Intelligence Project.

I.1 THE HAND-EYE PROBLEM

The goal of hand-eye research (Feldman et al [1969]) is to use a computer to coordinate a visual sensor and a mechanical manipulator. We hope to demonstrate interesting perceptual- motor behavior in a realistically complex environment.

Figure 1.1 shows a typical hand-eye problem domain. An assortment of planar-faced children's toy blocks of various shapes, sizes, colors, and textures, are scattered on an arbitrary table surface. The lighting and surroundings are typical of a computer room environment.

The system can presently execute elementary tasks, for example, "PICK UP THE ARCH AND PLACE IT ON TOP OF THE

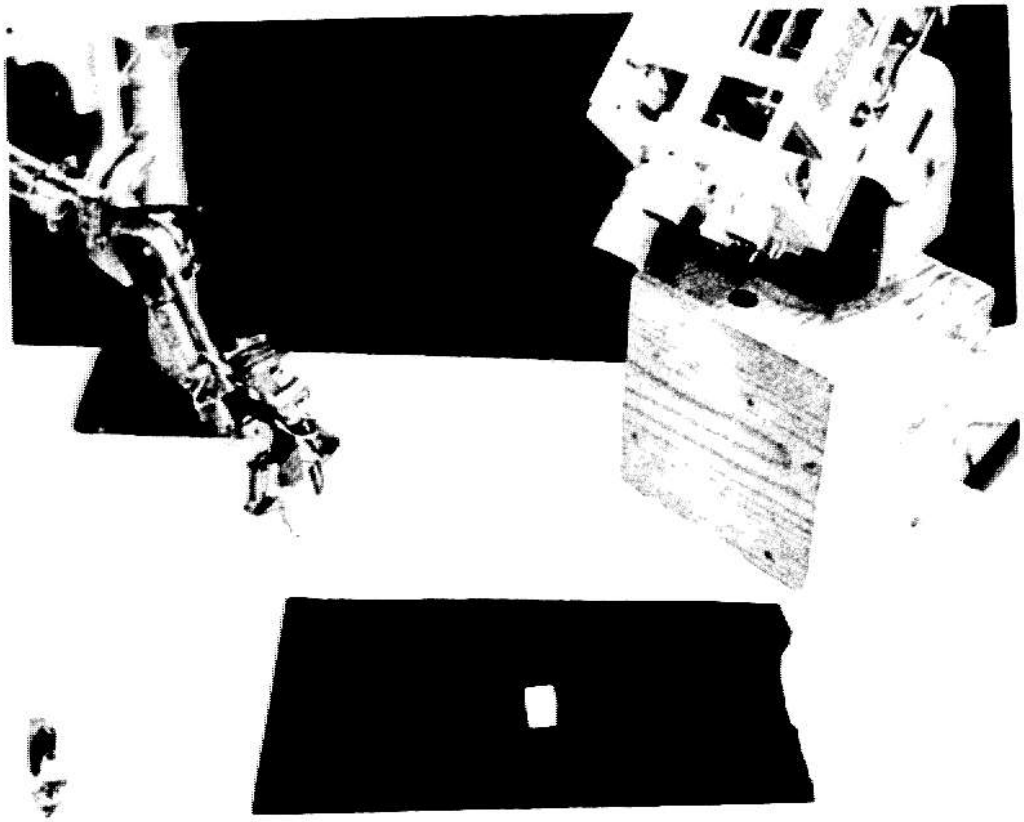


Fig. 1.1 Hand-Eye Work Space

CUBE". A short term goal is to compound these simple actions to realize behavior on the order of difficulty required to move a stack of blocks: ie. observe it, disassemble it, and, then, reassemble it elsewhere on the tabletop.

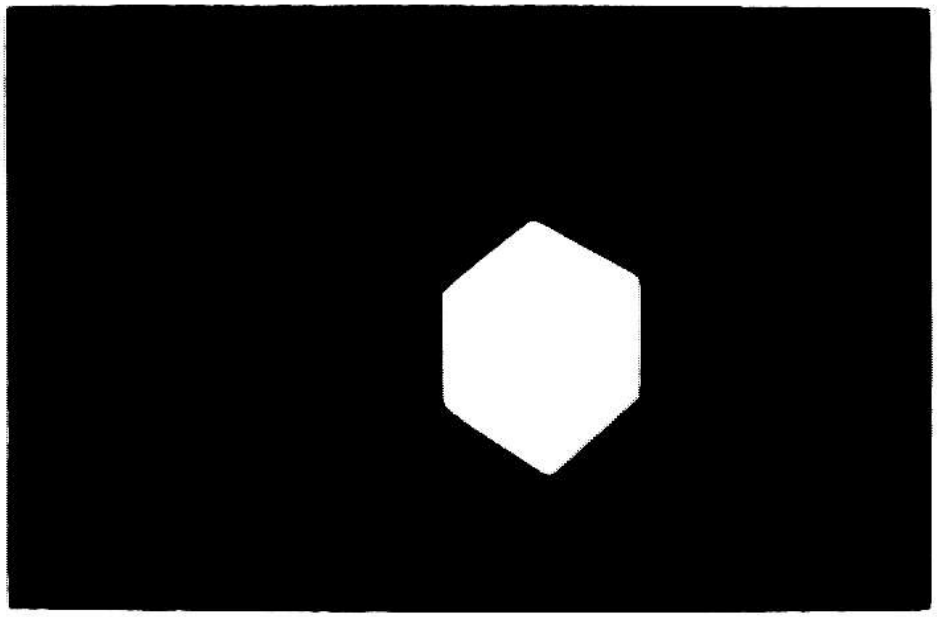
A basic stacking task requires the perceptual ability to locate specified objects (arch, cube,etc.) within the overall environment and to establish their orientation so that the hand can manipulate them.

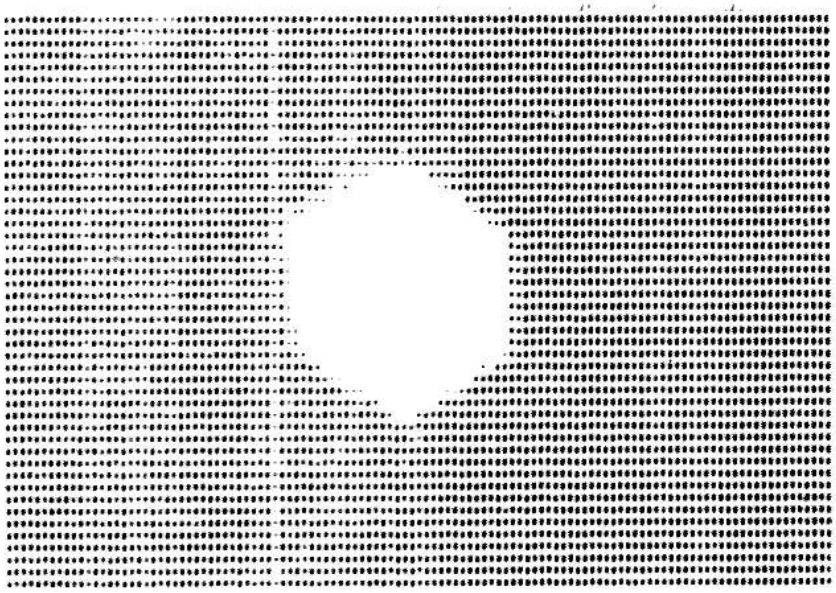
1.2 THE PROBLEMS OF COMPUTER VISION

1.2.1 A BASIC PERCEPTUAL PROBLEM

The computer views the world through an ordinary television camera. Figure 1.2a shows the image on the television monitor of the table scene depicted in Figure 1.1. Like the human eye, information is provided to the computer as an array of intensity samples, representing a point by point projective mapping of the environment. Figure 1.2b shows what the computer sees. (This figure was obtained by dumping the contents of the computer's memory on a line printer, using over-printed characters to simulate a grey scale.)

A full image contains 333x250, 4 bit intensity samples. The computer's problem is to reduce this mass of





information to a concise description of the scene. This desciption should include the location and orientation of objects (cube, arch, etc.) required to accomplish a hand-eye task.

The computer interprets the sensory data by recognizing correspondences between patterns in the data and patterns characteristic of toy blocks. The desired patterns must either be built into the program or somehow acquired on previous runs. Unfortunately, the computer is severely handicapped as a pattern extractor. It is forced by its lack of parallelism to view the world, as through a pinhole, a single intensity sample at a time. (The reader can get a feeling for the problems involved by viewing Figure 1.2b one point at a time through a hole punched in a paper mask.)

To overcome this handicap, the computer must be very clever in how it acquires and organizes the data it considers. An important simplification comes with the realization that the sensory data need initially supply only the coarse information required to locate interesting (that is, non-empty) regions of the tabletop. Sufficient additional detail can then be extracted in this localized area to describe what is contained.

The level of detail that must be provided by the receptors is largely a function of how much is already known. When the properties of the different blocks are

well-known to the program, the sensory data need only provide enough detail to distinguish among the possible alternatives. This detail presumably is substantially less than the information that would be required to completely describe the object.

The planar-faced solids used in hand-eye tasks are distinguished primarily on the basis of shape. For such objects the intensity edges that are formed at surface boundaries constitute a suitably complete description. Finer levels of detail, such as texture, will contribute little to the recognition process, unless, of course, a task specifies a particular object by its texture. (Texture may also help when the contrast is too low to find clean intensity edges.) When the set of possible objects is limited to a few simple geometric forms, a subset of the edges of any member (eg. the exterior contour) will strongly constrain (and often uniquely determine) the identity of that member (Falk [1970]).

1.2.2 A SIMPLE SOLUTION

The perceptual system shown in Figure 1.3 is an elementary embodiment of these ideas. It is similar in concept to the initial hand-eye vision package assembled by Wichman [1967]. Applying this system to the image in Figure 1.2b, the program would begin by sampling the picture in a

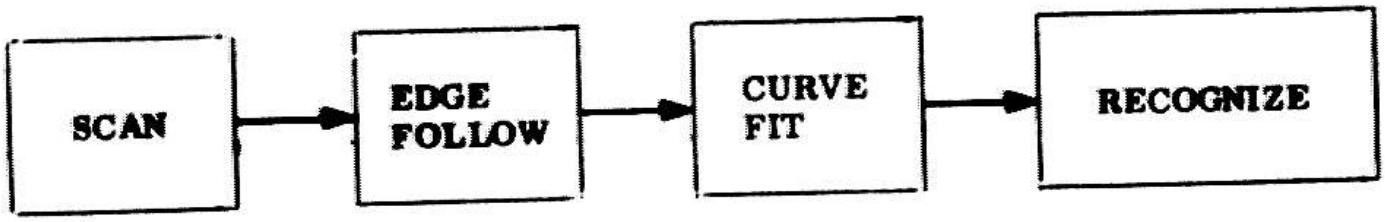


Fig. 1.3 Hierarchical Recognition Paradigm

coarse raster looking for the boundary of an object (as indicated by a discontinuity in intensity). Starting at the lower left-hand corner of the image, the program would look at perhaps 20 points on each horizontal line until it encountered a dark to light transition. After localizing the discontinuity between the coarse sampling points, a gradient operator is used to track the sequence of intensity transitions which define the object's contour (see Figure 1.4). Finally, straight lines are fit through these boundary points to obtain the exterior edges and vertices (see Figure 1.5) needed to match the machine's internal model of a cube.

1.2.3 LIMITATIONS OF SIMPLE SOLUTION

The apparent success of this strict hierarchical approach to perception required many simplifying assumptions which often escape the casual observer:

1. The system worked only with homogeneous, highly contrasting objects and backgrounds viewed under strong, uniform illumination.
2. The system knew only about cubes.
3. The system assumed all objects were resting on a horizontal table surface.

Assumption 1 made it trivial to extract reliable exterior

Fig. 1.4 Exte[illegible] [illegible]sity Gradients

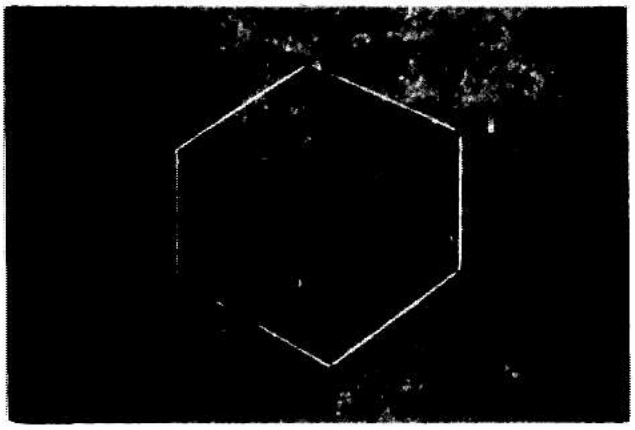


Fig. [illegible] [illegible]ints

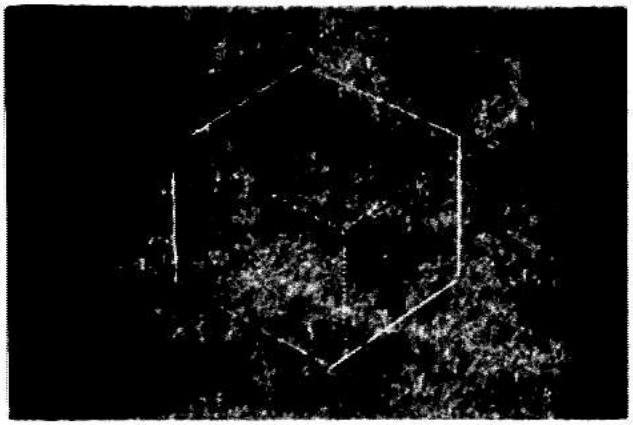


Fig. [illegible] [illegible]nts

contours. Assumption 2 allowed a unique (sic) identification on the basis of just this contour. Assumptions 2 and 3 together allowed the interior edges (shown dashed in Figure 1.6) and, thus, orientation to be inferred from the contour. This is very helpful since the interior edges are usually of low contrast and consequently much harder to see then the exterior boundary.

Needless to say, this combination of factors will not often be found in more interesting hand-eye tasks. In this thesis we thus consider perceptual strategies appropriate in the absence of any or all of the special conditions noted above.

1.2.4 PRACTICAL DIFFICULTIES

Figure 1.7 shows the computer's view of a more elaborate hand-eye work space. It contains a realistic sampling of the difficulties that make three-dimensional scene analysis such a challenge. Consider the information required to place the wedge on top of the arch. Once again, simple line drawings of these objects provide sufficient detail to determine location and orientation. However, interior edges will be needed to interpret the complex outline that results from a cluster of objects in close proximity.

Unlike the simple world of Figure 1.2b, the wedge

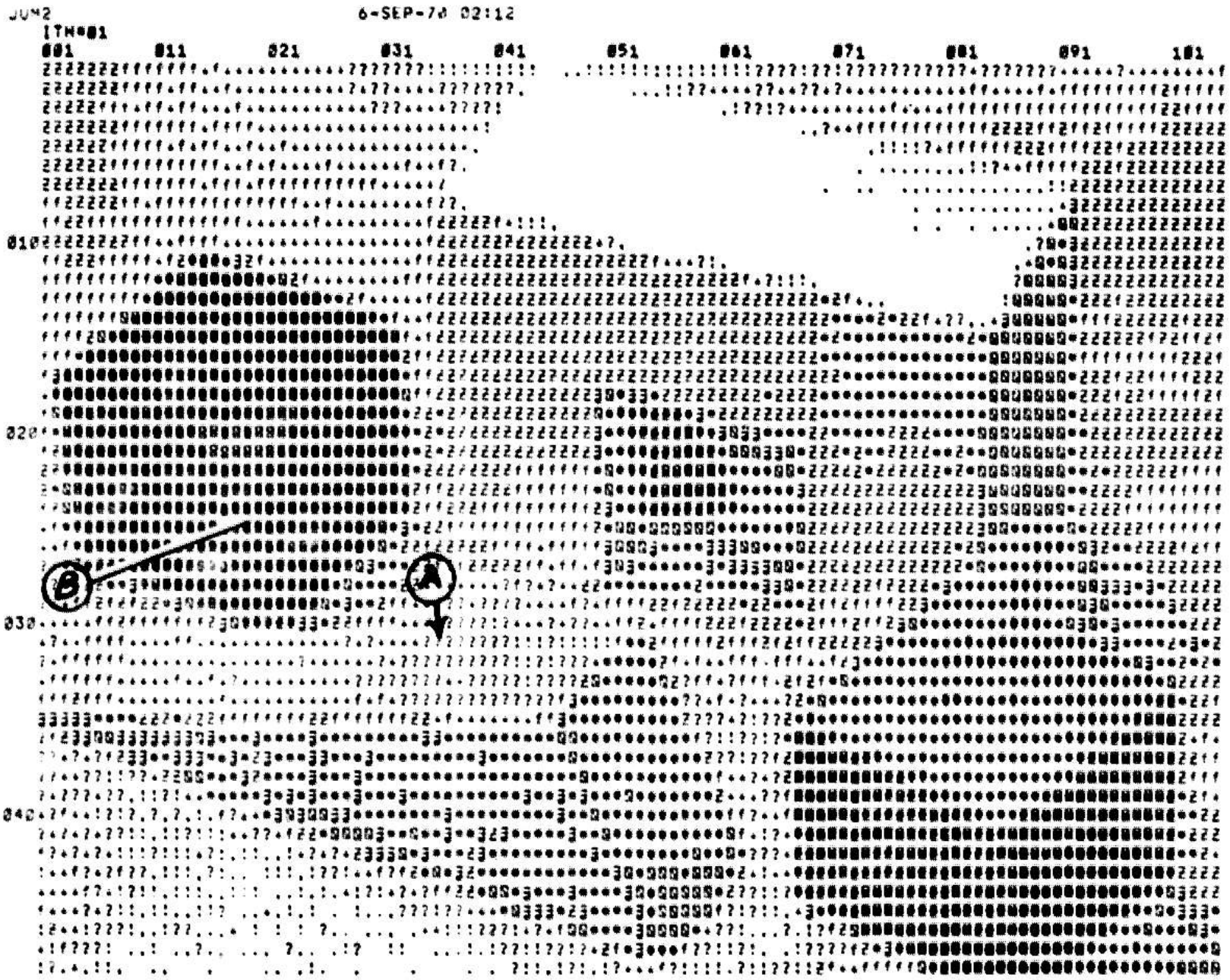


Fig. 1.7 Non-Trivial Hand-Eye Work Space

and arch are embedded in a complex background of extraneous textures and shading gradations, and, most noticeably, other objects. The overwhelming majority of the detail contained in Figure 1.7 is thus either redundant or irrelevant in the context of this specific task.

Isolating the desired objects without exhaustively analyzing the scene is a formidable search problem (Recall the pinhole!). To make such a search practical, it is essential that the machine's perceptual system utilize every available clue as to where and what to look for. Color, texture, and depth cues, for example, can augment the intensity discontinuities that sufficed in simpler settings. Attributes like color can be particularly helpful in testing and eliminating unwanted candidates during the acquisition process, thus avoiding costly edge extraction.

Search clues may be provided explicitly in a task description, ("PICK UP THE RED WEDGE IN THE LOWER LEFT-HAND CORNER"). They may also be found implicitly in the machine's internal world model (that is, the attribute red could be a given intrinsic property of wedges).

The presence of irrelevant textures and random noise, within the objects as well as the background, can considerably complicate the edge following process once a suitable candidate is found. These disturbances are especially successful at camouflaging low contrast edges, (such as the ill-defined rear border of the wedge labeled A)

from detection by local, gradient type operators. Larger operators (Heuckel [1969]) that process many intensity samples simultaneously can sometimes recover these weak edges, though at a substantial overhead in processing time. The low contrast of this edge is, however, symptomatic of a more fundamental problem.

I.2.5 SENSOR LIMITATIONS

The television system, like any communications channel, can transmit a finite quantity of information about a scene in the time frame represented by a single image. The vidicon provides 83000 intensity samples every 1/60 second. The capacity of our computer's high speed data channel (also used by the swapping disk) is 24 million bits/second. Thus, without local buffering each sample can be encoded as a 4 bit number. It is then possible to distinguish just 16 levels of intensity between some absolute limits. In our system these limits are adjustable.

It is not surprising that when these limits are set wide apart, there will be intensity discontinuities which the available quantization density will fail to resolve. It is not possible to cover the full range of intensities found in Figure 1,7 and to resolve edge A in a single image. In fact, even the widest quantization window will often fail to span the full brightness range of a scene.

The vertical interior edge of the left hand cube (labeled B) is absent, because the brightness of the entire cube was compressed into the lowest digitization level.

In any scene encompassing a sufficiently wide range of intensities, it is inevitable that some detail will be lost due to clipping by the quantizer. If the digitization range had been concentrated at a lower absolute level of intensity, Edge B would then likely be in evidence. However, other features, such as edge A, would be lost because of clipping at the high end. The limitations of the sensory channel necessitate a compromise allocation of the available intensity resolution.

This trade-off is typical of conflicts that exist in each dimension of the camera's response. There is always more visual information available than can be handled. For instance, consider the contradictory requirements of spatial resolution and field of view. The image magnification, established by the focal length of the lens and the object distance, determines whether the 83,000 intensity samples will provide coarse coverage of a wide field of view or high resolution in a small area. The combination of spatial resolution and field of view available in any image is strictly constrained by the number of available raster samples.

I.3 NEED FOR ACCOMMODATION

The important conclusion to be drawn from these examples is that no single image will contain adequate information for every perceptual goal. Taking a clue from nature, the most appropriate way to resolve these trade-offs is to tune the perceptual system, concentrating the available resolution on that portion of the sensory data that is currently of most interest. The effects of this process, which we shall call "ACCOMMODATION", can be illustrated by comparing the information emphasized in Figures 1.7 and 1.8. In Figure 1.8 the intensity quantization range is intentionally compressed to enhance the contrast across the boundary between the wedge and the background. Notice the striking improvement in the quality of edge A, produced by the combination of local contrast enhancement and suppression of irrelevant textures. This edge can now be extracted by a simple gradient operator with more reliability than could be expected from more costly processing applied to the inferior raw information in Figure 1.7.

Figures 1.7 and 1.8 graphically demonstrate that what will be seen depends strongly on the parameters of the visual channel, that is, how one looks. This result has obvious implications for the design of a context-sensitive

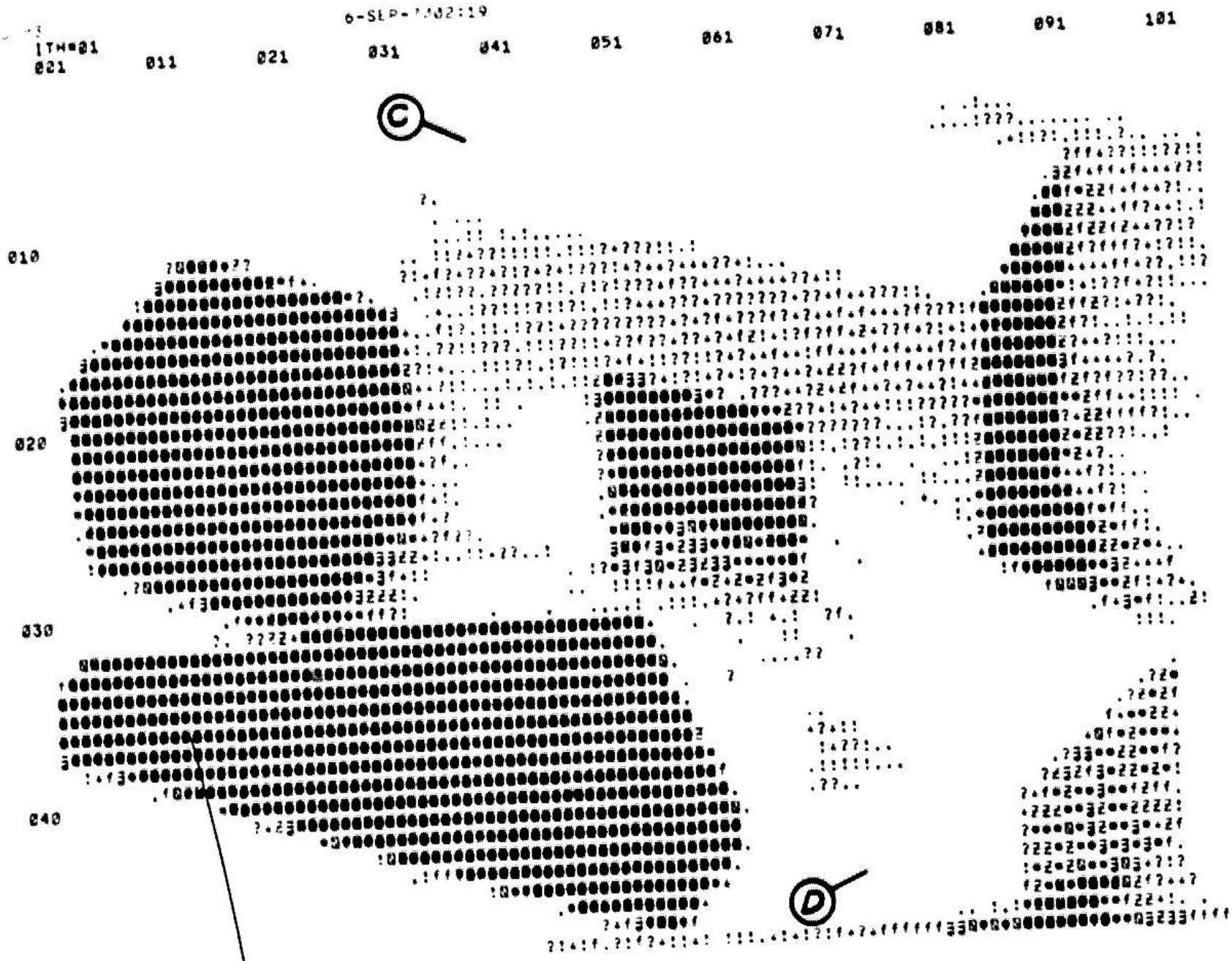


Fig. 1.8 Accommodation Optimized for Contour of Wedge

perceptual system whose objective is to seek specific information, needed to accomplish a task. This thesis studies some of the ways in which sensor accommodation can contribute to the computer's efficiency and success in seeing what it is looking for.

1.3.1 PURPOSE OF ACCOMMODATION

The purpose of accommodation is to obtain from the sensors the most appropriate image for the current perceptual goal. What is meant by an appropriate image depends, of course, on the specific goal. In general, the image should provide the required information in the simplest possible form to minimize the required scene analysis. This requirement usually means that the desired features are present with high contrast and, equally important, that all unnecessary detail is suppressed. Figure 1.7 was inferior to Figure 1.8 as a source of information about the exterior boundary of the wedge because of the combination of low contrast and irrelevant textural noise along edge A. Presumably, the loss of detail in other parts of Figure 1.8, such as the disappearance of edges at points C and D, is of no consequence when interest is focused on the wedge.

1.3.2 RATIONALE FOR ACCOMMODATION

The low contrast of edge A in Figure 1.7 is a consequence of the fact that intensity resolution has been sacrificed to obtain dynamic range. One could thus question the motivation of using accommodation to overcome what might be considered as a limitation of our present hardware. Why not devote the effort to the development of an improved camera (and data channel) with better resolution over a wide dynamic range? To answer this question, we must clarify the dual role of accommodation.

One use of accommodation is to extend and improve the capabilities of the sensor. For example, an adjustable iris helps the human eye to make fine intensity discriminations over an extraordinary range of brightness (spanning ten orders of magnitude). An accommodating lens allows the eye to observe a wide field of view or to concentrate on fine textural detail.

Clearly, a more developed sensor will depend less on accommodation to obtain this kind of flexibility. (A sensor with more independent intensity samples over a given size raster will provide more spatial detail over a given field of view. There is thus less necessity to use a longer lens to observe texture.)

However, the human eye is unique among visual sensors in its capabilities for maintaining high levels of

performance over an exceedingly large range of scene characteristics. Man has yet to create a visual sensor which can duplicate the flexibility and performance standards established by nature. Thus, the need to accommodate artificial sensors is correspondingly more acute.

Still, the dynamic range required to observe the constrained hand-eye environment is not so great that a suitably broad sensor could not be designed. A more basic problem is the inability of the computer to process the volume of data already available. The solution to this predicament is not to seek more data but rather to be more selective.

The second role of accommodation is then to concentrate the limited sensor capacity on information that emphasizes selected visual characteristics. Since the sensors must presumably exclude some detail, it is preferable that the lost detail be irrelevant in the current task context. Accommodation thus capitalizes on the "limitations" of the sensors in order to obtain more appropriate images.

Some evidence of peripheral masking is also found in humans (Munn [1961]). The most significant of these receptor adjustments are readily observable. The direction of the head and eyes is the principal determiner of what will be seen. The position of the fovea selects a portion of the

visual field for consideration at high resolution. The accommodation of the lens affects the detail with which the contents of this field are seen, as a function of depth. These sensor adjustments account for much of the relative clarity with which you perceive this text compared with the details of other objects in the room. Behavioral acts such as bringing the paper close to your eyes (so that it fills the field of view), or putting it under a strong light (to enhance the contrast) are also helpful.

The information theoretic capacity of the human optic array far exceeds the observed capability of the human brain to process it in real time (Triesman [1964]). At some point prior to the limited capacity decision channel, man must thus select the most significant information about the environment from that provided by his sensors. Broadbent [1958] devised a general "information flow" model of this process (Figure 1,9).

It is clearly advantageous to defer selection to the highest level of processing for which adequate channel capacity exists. In man, a substantial number of nerve fibers can be traced all the way from the optic nerve to the cerebral cortex. The role of "receptor set" in human selective attention is thus likely to be limited.

The processing capacity of the central nervous system in lower animals is more modest than in humans. It is common for such animals to avoid information overload by

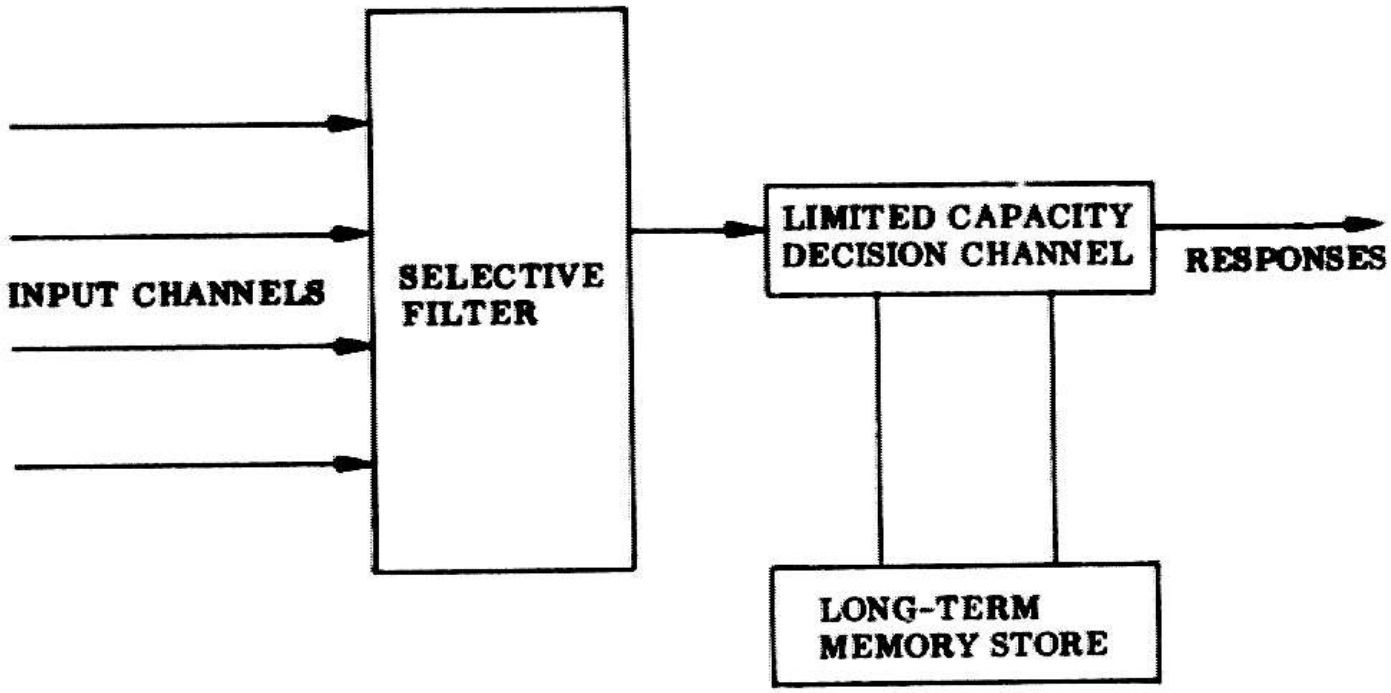


Fig. 1.9 "Filter" and "Information Flow" Model for Selective Attention (Broadbent, [1958]).

relying on specialized sensors, permanently adapted to respond to those environmental characteristics most crucial to its well-being,

Lettvin [1959] in his work on the perceptual system of a frog, succeeded in isolating several neurons capable of complex, though highly specialized visual discrimination within the animal's retina. He found cells dedicated to blue-green discrimination as well as cells that fired only in the presence of small, rapidly moving, convex objects. How convenient, considering the importance of grass and water and insects to a frog's survival! An important advantage of utilizing specialized sensors is that the selection of relevant data is accomplished at the earliest stage of processing where because of sheer volume, the potential for data reduction is greatest.

Information selection at the sensor level is perhaps most essential for machine perception. The limited bandwidth of the data channel provides an immediate constraint which is surmountable, but at excessive cost.

A more basic factor is that current digital computers lack the parallel structure needed to extract visual gestalts directly. All features must be laboriously assembled out of individual intensity samples. If an undesired object is included in the image, the computer will be forced to exert considerable effort to discriminate it from the desired object.

The efficiencies that are gained by using the sensors to filter out irrelevance at the lowest level (data entry) are crucial to the successful implementation of selective attention in machines. It is, therefore, surprising that up until now, no one but nature herself has explicitly included provision for goal-directed sensor control in the design philosophy of a functioning vision system.

I.3.3 SOME OTHER EXAMPLES OF ACCOMMODATION

Let us examine two additional instances in which accommodation can be used to facilitate basic perceptual functions. These examples are intended to be introductory in nature and will be reconsidered in greater detail in later chapters.

I.3.3.1 ACCOMMODATION FOR SELECTIVE ATTENTION

We have mentioned the use of accommodation for selective attention. Most hand-eye tasks designate explicit visual features (for example, "PICK UP ALL RED CUBES."). In the context of this task, only bright red, hexagonal shaped regions of the scene are of interest. In a complex environment it would prove excessively costly to identify the color or shape of every coherent region,

The problem is thus to isolate the required set of objects from the background environment without analyzing everything in the scene.

We have shown that in any environment what is seen depends on how the camera parameters are tuned. It is thus appropriate to adjust these parameters so that the characteristics of the desired objects are enhanced at the expense of irrelevant features. Consider the task of locating the red cube in the specific environment shown in Figure 1.10a. Here, the desired object appears against a black background, in close proximity with two non-red cubes. (The weak interior edges in this picture are again due to a wide quantization window.)

These known characteristics can be used to set the limited range of the camera to emphasize the red object. Figure 1.10b shows how the scene in Figure 1.10a would appear to the computer if the image were obtained through a red color filter with the digitization range set to clip out below average intensities. The effect of this accommodation is first to enhance the contrast of the red object with the background relative to the other colors. The darker intensities are then compressed by the quantizer, forcing those portions of the image containing the undesired objects into the same equivalence class as the background.

In Figure 1.10b the problem of locating a red object has been reduced to that of a simple search for the largest

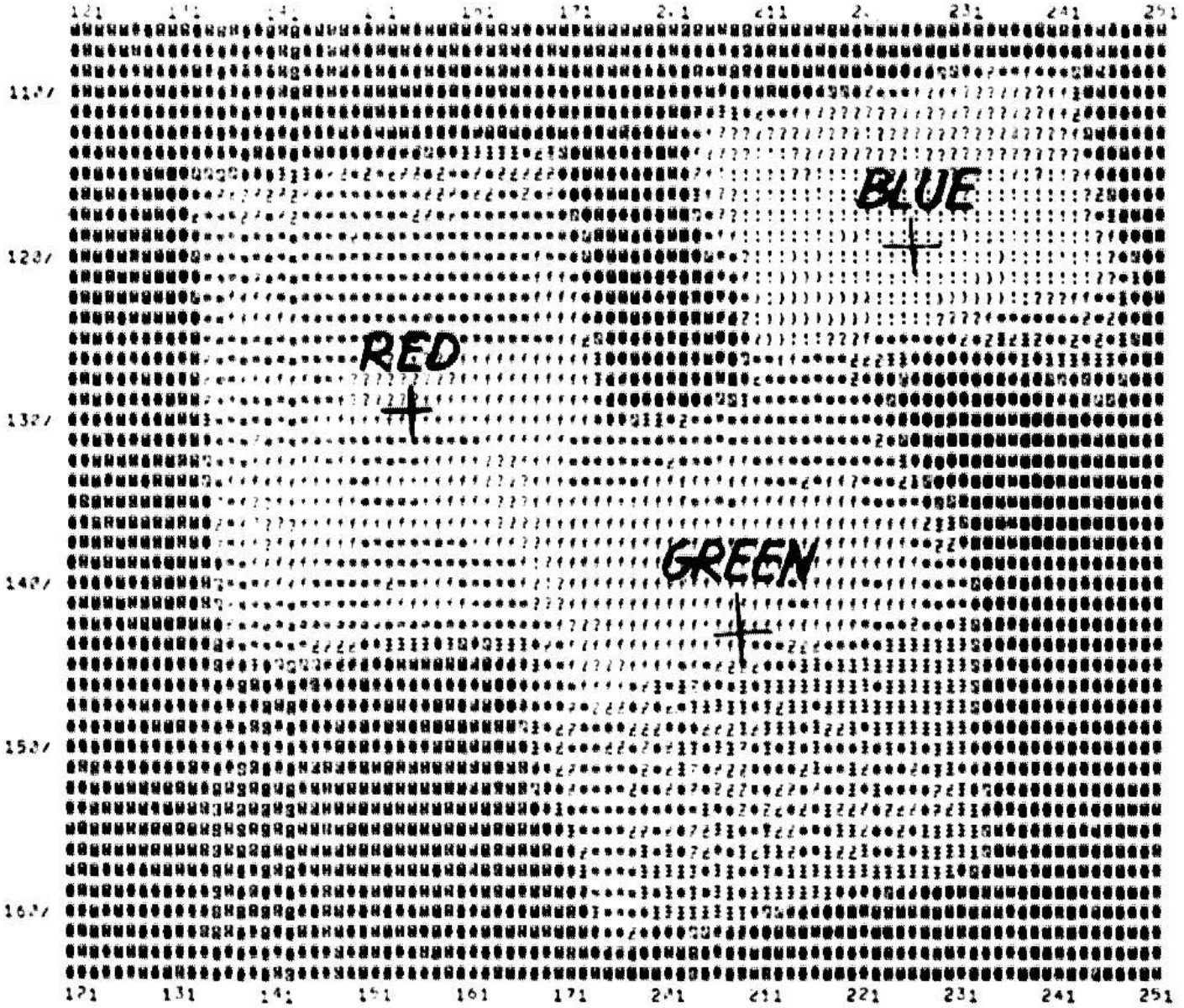


Fig. 1.10a Selective Attention: View with Clear Filter

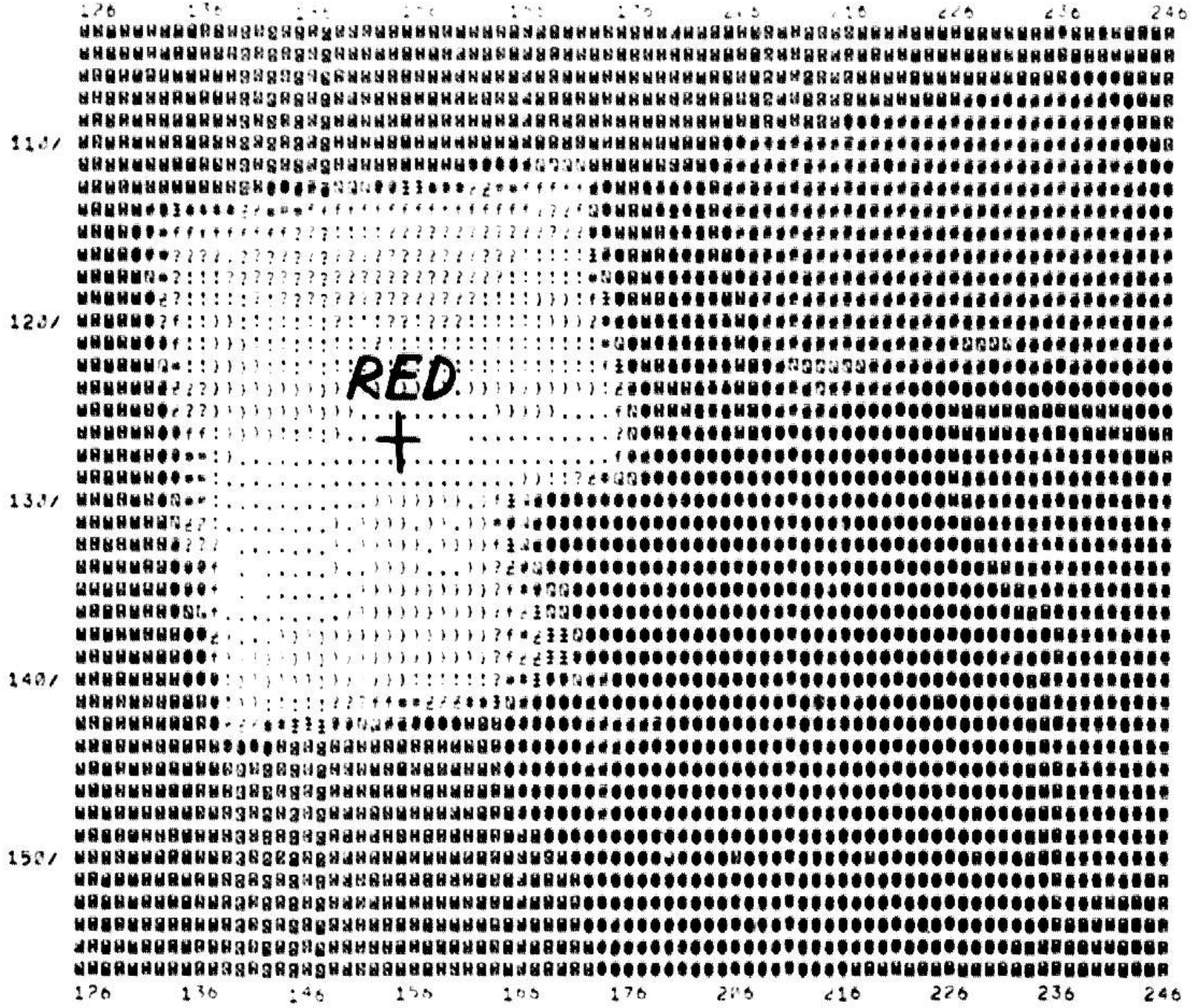


Fig. 1.10b Selective Attention: View with Red Filter

black to white intensity transition. The color and shape of the associated region can then be tested to confirm that they fulfilled the search objectives. Although it cannot be guaranteed that this region will correspond to the desired object, the red filter has considerably increased the a priori likelihood that this will be so. This accommodation reduces the expected cost of locating the red cube. It eliminates tests that would otherwise be needed to reject false prospects had they been acquired first.

Figures 1.10c and 1.10d obtained through green and blue color filters are respectively the most appropriate images in which to seek green and blue objects. In general, the choice of accommodation for selective attention depends on the acquisition requirements and on what is already known about the environment from which these characteristics must be isolated. Consider, for example, the influence of the predominant background hue (in this case black) on the choice of color filter. If the background had been white instead of black, a green filter would have been needed to emphasize the presence of a red object (as a dark region). The digitization range would then be selected to clip above average intensities. If the shade of the background had not yet been determined, one could not specify a preferable color filter. In this case, the most appropriate image would be Figure 1.10a, because it contains the widest range of intensities.

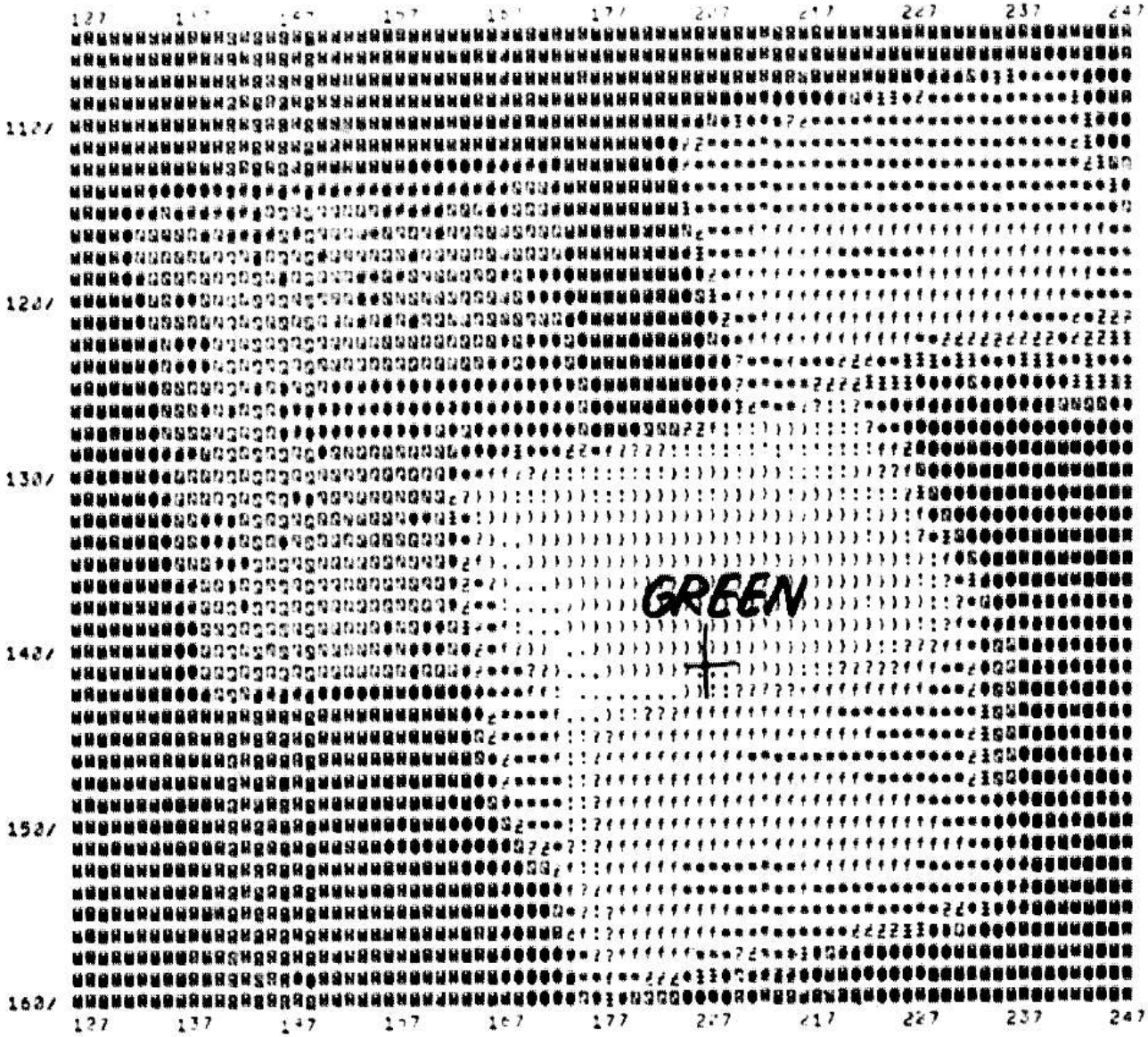


Fig. 1.10c Selective Attention: View with Green Filter

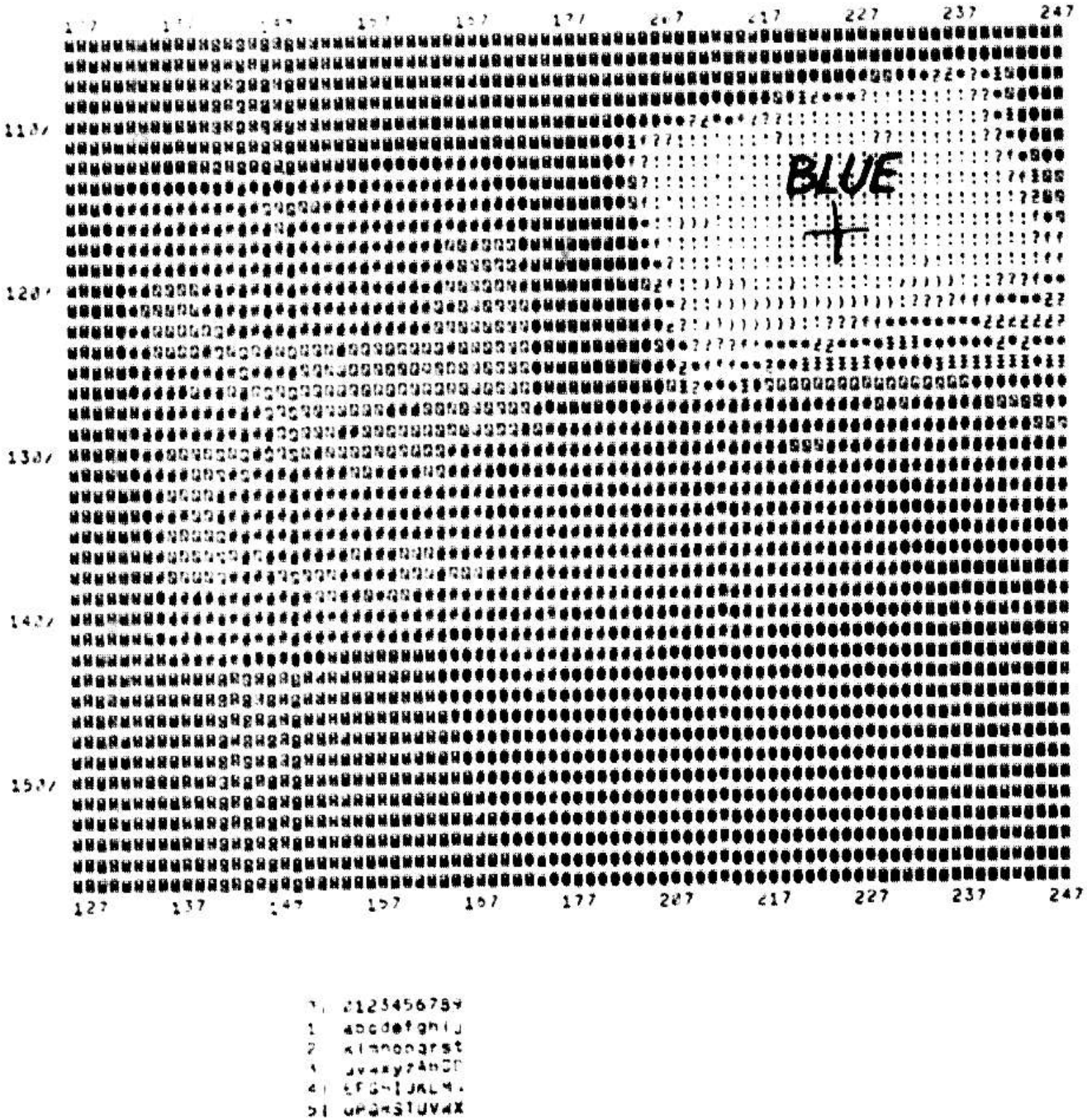


Fig. 1.10d Selective Attention: View with Blue Filter

The selection of information by an appropriate accommodation can provide the principal means of information reduction in a context-sensitive scene analysis. This power can be attributed to two factors:

1. Accommodation provides selectivity at the level of raw image data. Here the need is most critical because of the volume of potential information in a scene. Furthermore, the most significant efficiencies can be gained by filtering out irrelevant details before any processing is actually expended.

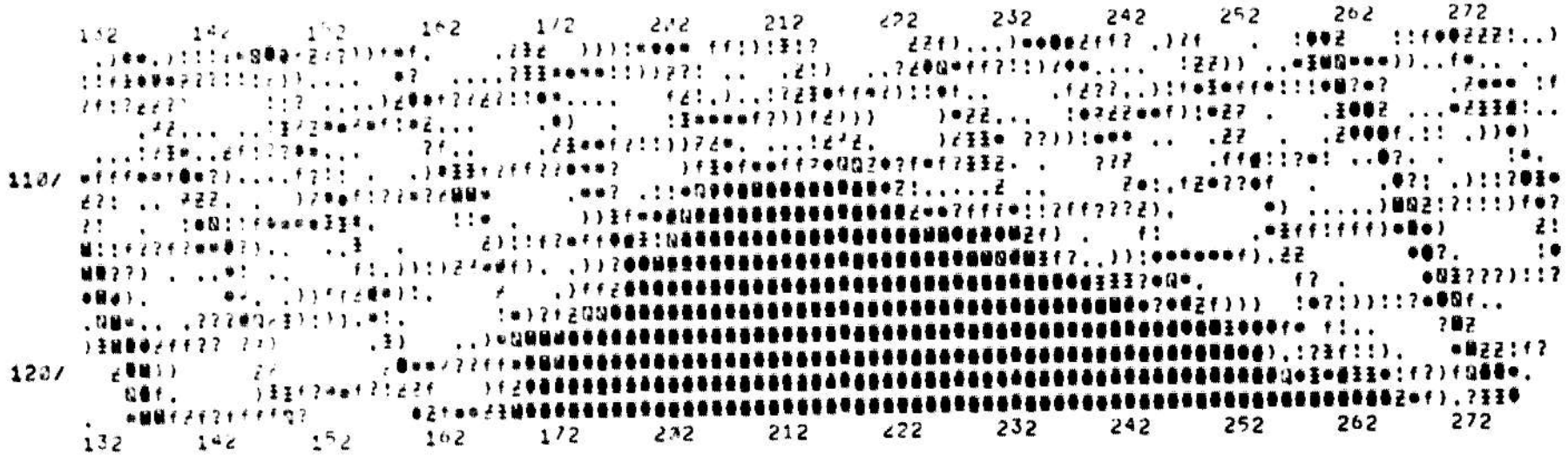
This selectivity should not be confused with simple thresholding, as might be applied with software to intensities obtained with a wide range sensor. There is no basis by which to distinguish intensities corresponding to the three colors in the unselective image (Figure 1.10a). Camera focus is another useful accommodation for selective attention. With a suitably narrow depth of field, the contrast of objects outside of a selected range can be made so low that they are effectively out of the image. This type of discrimination cannot be simply simulated in software.

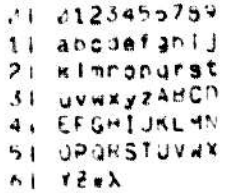
2. Accommodation filters in parallel over the entire image. All features sharing some common characteristic simultaneously vanish from view. This feature is crucial since the computer is fundamentally a serial processor.

I.3.3.2 ACCOMMODATION FOR SELECTING AN APPROPRIATE LEVEL OF DETAIL

In the introductory example on the enhancement of edge contrast, we alluded to the importance of eliminating unnecessary textural detail. In the context of edge following, texture can be defined as all detail of significantly smaller spatial dimensions than the structure whose boundary we wish to extract. Such texture is especially prominent in the situation sketched in Figure 1.11b. Here, a cube is sitting on a table covered by a coordinate grid. (The grid provides a common reference frame with which to calibrate the coordinate systems of the camera and the arm.) Figure 1.11a shows the data available to the computer in a full range image of the area in the vicinity of the top rear corner of the cube. The grid lines are an attractive distraction to a myopic edge follower trying to trace the boundary of the cube on the basis of local gradients.

There is no clear separation between the cube and the grid lines on the basis of intensity. Thus, the unwanted detail cannot be removed merely by clipping it into the background with an appropriate quantiziation window. The only basis for enhancing the contrast between the object and background, relative to that for the grid lines, is to





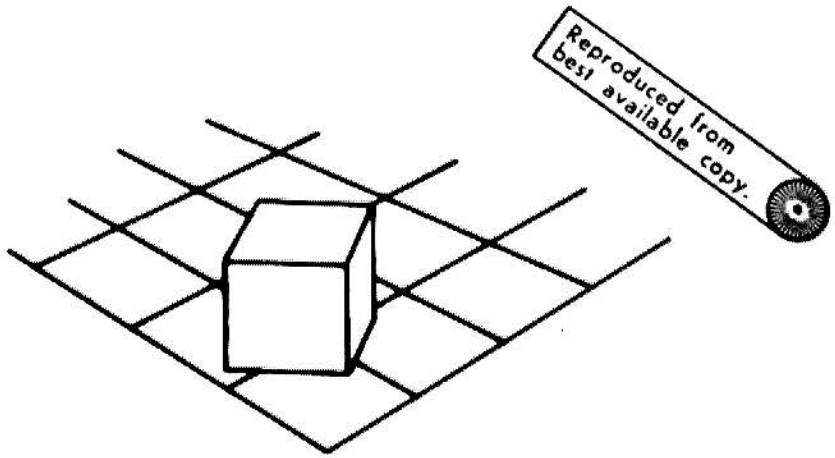


Fig. 1.11a Corner of Cube on Grid Background

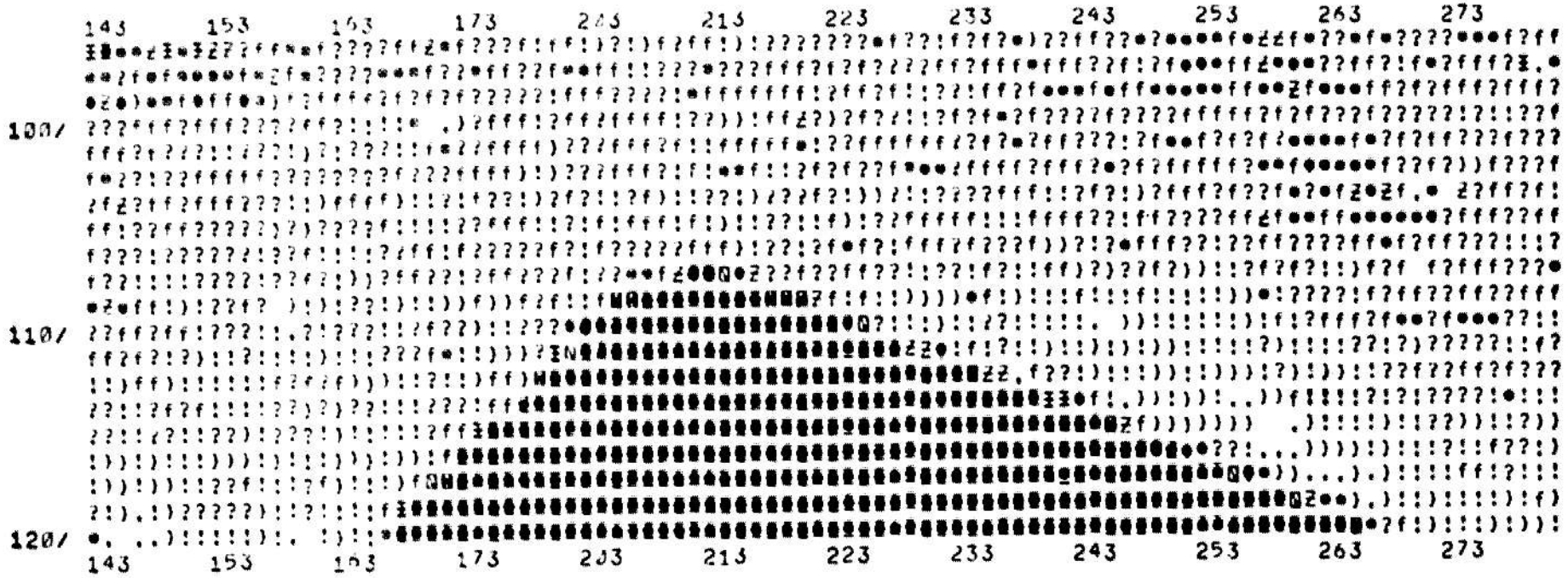
Fig. 1.11b Sketch of Cube on Grid Background

exploit differences in spatial frequency content. It follows from the definition of texture that the energy corresponding to the grid lines will be concentrated in a higher frequency range than that corresponding to the cube.

It is well-known (see Chapter 6) that simple defocusing is equivalent to passing an image through a spatial low-pass filter. This effect is illustrated in Figure 1.11c. As a result of slight defocusing, the energy that had been concentrated in the well- defined grid lines has been diffused evenly over the background. The predominently low-frequency energy concentrated in the cube has not been noticeably affected. This operation has destroyed the distinct structure of the grid lines without significantly affecting the structure of the desired boundary. Furthermore, the demise of the grid lines has left a clear intensity separation between the cube and the diffused background. This contrast can now be enhanced by concentrating the quantization levels on the range of intensities found between the cube and the background. The combined effect of focus and intensity accommodation is strikingly shown in Figure 1.11d.

1.4 ACCOMMODATION IN A PERCEPTUAL SYSTEM

We have shown several examples of how accommodation can overcome some practical problems frequently encountered



0| 0123456789
1| abcdefghij
2| klmnopqrst
3| uvwxyzABCD
4| EFGHIJKLMN
5| OPQRSTUVWX
6| YZ*λ

Fig. 1.11c Defocused Image of Cube Corner

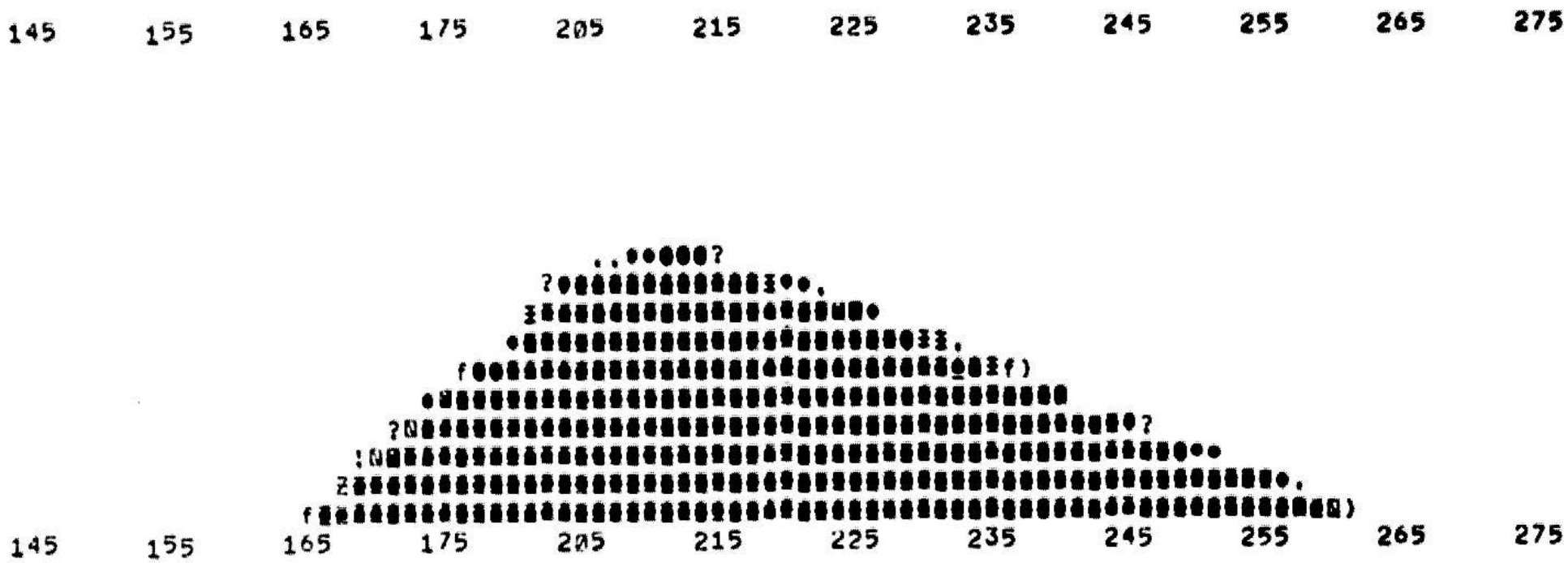


Fig. 1.11d Combined Effect of Focus and Intensity Accommodation on Cube Corner

in machine perception. To fully utilize these effects, they must be integrated into a comprehensive strategy for visual perception.

I.4.1 AN ACCOMMODATIVE VISION SYSTEM

I.4.1.1 GOALS OF SYSTEM

The ultimate goal of the hand-eye system is to complete a task with the minimum effort consistent with high reliability. To achieve this goal the vision system must attend only to those parts of a scene that are relevant to the current task. It must extract the minimum level of detail that provides a sufficient description to complete the task objectives.

A more modest initial goal is to obtain a system that will perform simple tasks, reliably, in a variety of environments. The current vision system adapts its level of effort to the difficulty of the environment in order to insure reliable operation.

I.4.1.2 DESIGN PHILOSOPHY

The vision system is composed of a flexible and overlapping variety of processing options that permit cost-effective solutions to many perceptual problems. The

general strategy is to apply cheap operators over the entire scene in order to first extract the reliable, easy-to-see features. If the scene is simple, like Figure 1.2b, this coarse structure will generally suggest probable matches linking the objects in the image with corresponding prototype models known to the machine. These guesses can then be verified by using the suspected models to hypothesize additional features, whose presence can be specifically sought in the sensory data. (This mode of information acquisition is reminiscent of the parlor game, twenty questions.) If, however, the scene contains a cluster of objects or the image is troubled by low contrast and noise, the simple routines will not provide enough information to obtain definitive matches. In these cases, the strategy is to return to the world for another and harder look.

Fortunately, it will not often be necessary to reprocess the entire image. Typically, there will be a small region of uncertainty where more detail would be helpful. Furthermore, the recognition program will often be able to make use of its a priori knowledge and of what has so far been found, to suggest a specific feature whose presence would resolve a particular recognition ambiguity. The entire perceptual system can then be accommodated to maximize the likelihood of detecting that feature.

I.3.1,3 ROLE OF ACCOMMODATION IN SYSTEM

The role of sensor accommodation is most easily discussed in the context of a specific system. Figure 1.12, a simplified version of an emerging design, is provided for this purpose.

As in the initial hand-eye system, recognition is based on the representation of objects in a scene by line drawings. To acquire an edge, the image is again scanned for an intensity discontinuity. The appropriate accommodation during this acquisition phase, depends on the task requirements and on what is already known about the environment. In the absence of a specific search goal (ie. a task has not yet been posed), or of detailed knowledge, one cannot do better than an exhaustive search for all edges. The best image for this purpose is one covering the widest field of view and the widest range of intensities. First, the chances of missing a prominent feature because of an arbitrarily limited source of information are minimized. Secondly, strong edges will be emphasized by the low intensity and spatial resolution in a broad image. Fine detail (including weaker edges), whose presence only adds to the problems of acquisition, will not be seen.

Suppose, on the other hand, that a specific item is sought and some knowledge of the environment is already available. The machine may then be able to limit the search

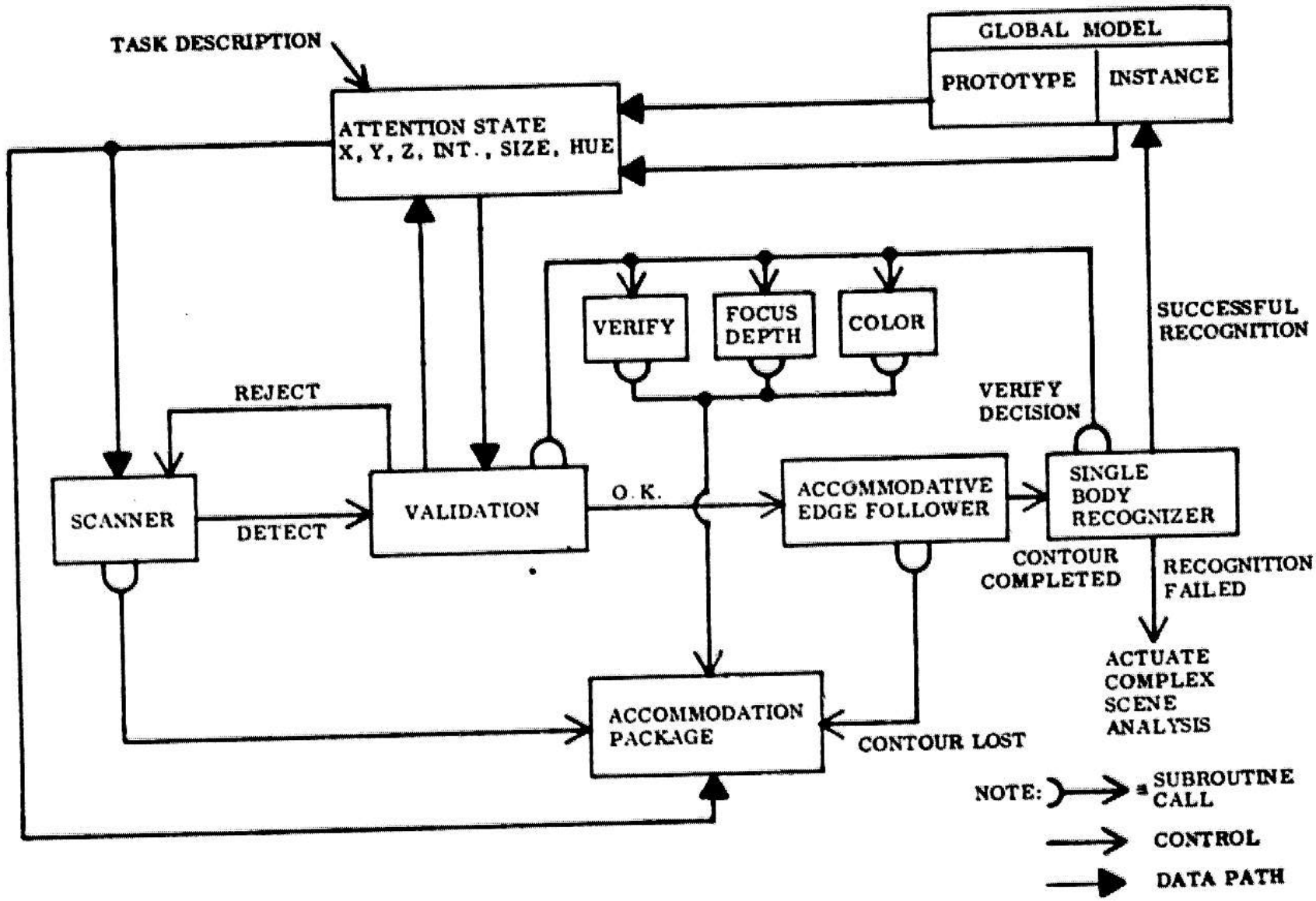


Fig. 1.12 Elementary Accommodative Vision System

by tuning its sensors (as in Figures 1,10a-d) to enhance a prominent visual characteristic of that item. The characteristics that influence accommodation define the system's "ATTENTION STATE".

When an intensity discontinuity is detected, the associated region is first subjected to a series of validating tests (described in Chapter 5). These tests determine whether the region fulfills enough of the criteria given in the attention state to justify the expense of edge extraction. Currently, color and minimum size can be checked. Size is a default requirement used to discriminate against discontinuities produced by an isolated noise phenomenon.

If the validation state rejects too many regions for the same reason, the accommodation can be tightened to discriminate against that characteristic. On the other hand, if no discontinuities are detected, the accommodation must be made less discriminating and more sensitive to low contrast.

After the acquisition is validated, the accommodative edge follower (Chapter 5) tracks the edge attempting to complete a closed contour. If the edge is lost in a region of weak contrast (eg. shadow), the camera can be re-accomodated to enhance the contrast in areas directly adajacent to where the edge was last seen. Considerably more effort can be applied to recovering the

lines that show evidence of global structure and throws away those whose existance cannot be so justified. This program can also call on the VERIFIER to see whether additional data can be found in the image to support these inferences. The completed line drawing is processed by the COMPLEX BODY RECOGNIZER (Falk [1970]) This program attempts to segment the lines into isolated bodies. Depth information obtained from camera focus (Chapter 6) can be helpful in this operation. The separate bodies are then identified using the hypothesize and verify techniques of SIMPLE.

I.4.1.4 ORGANIZATION OF CONTROL

I.4.1.4.1 CONTROL PHILOSOPHY

The principal control paths of this system can be isolated and modeled as in Figure 1.13. The basic control philosophy can be described as HIERARCHICAL EVALUATION. Briefly, it holds that:

The performance of lower level programs starting with those that handle the raw image data should be evaluated in two ways:

1. by the reasonableness of the immediate results, as determined by the system's low level expectations, and

2. by the reasonableness of the results of all higher level programs whose performance depends on the

edge at these specific points than would be practical or desirable to apply to the bulk of the scene.

When the contour has been closed, lines are fit to the edge points. The vertices, defined by the intersection of these lines, are then examined by the SIMPLE BODY RECOGNIZER (Falk [1972]). It determines whether they could correspond to an isolated, recognizeable body. If so, the appropriate object model is put forth as a tentative recognition hypothesis. This hypothesis can be verified by looking for the predicted interior edges. The EDGE VERIFIER (Chapter 4) is able to overcome the characteristically low contrast of these features by concentrating sensor resolution on the precise path over which they are expected and then accumulating global statistical evidence.

If recognition is not possible on the basis of this preliminary description, a more complex branch of the system (not shown) is activated. The camera is first accommodated for the general conditions in the region enclosed by the initial contour. A sensitive local operator (Heuckel [1969]) is then applied over this region to detect all interior edge points. This operator is likely to pick up pieces of weak edges, as well as isolated patches of texture. Pre-processing will usually be needed to massage this data into smooth line drawings. The LINE DRAWING COMPLETER (Grape [1969]) pieces together segments of sketchy

lines that show evidence of global structure and throws away those whose existance cannot be so justified. This program can also call on the VERIFIER to see whether additional data can be found in the image to support these inferences. The completed line drawing is processed by the COMPLEX BODY RECOGNIZER (Falk [1970]) This program attempts to segment the lines into isolated bodies. Depth information obtained from camera focus (Chapter 6) can be helpful in this operation. The separate bodies are then identified using the hypothesize and verify techniques of SIMPLE.

I.4.1.4 ORGANIZATION OF CONTROL

I.4.1.4.1 CONTROL PHILOSOPHY

The principal control paths of this system can be isolated and modeled as in Figure 1.13. The basic control philosophy can be described as HIERARCHICAL EVALUATION. Briefly, it holds that:

The performance of lower level programs starting with those that handle the raw image data should be evaluated in two ways:

1. by the reasonableness of the immediate results, as determined by the system's low level expectations, and

2. by the reasonableness of the results of all higher level programs whose performance depends on the

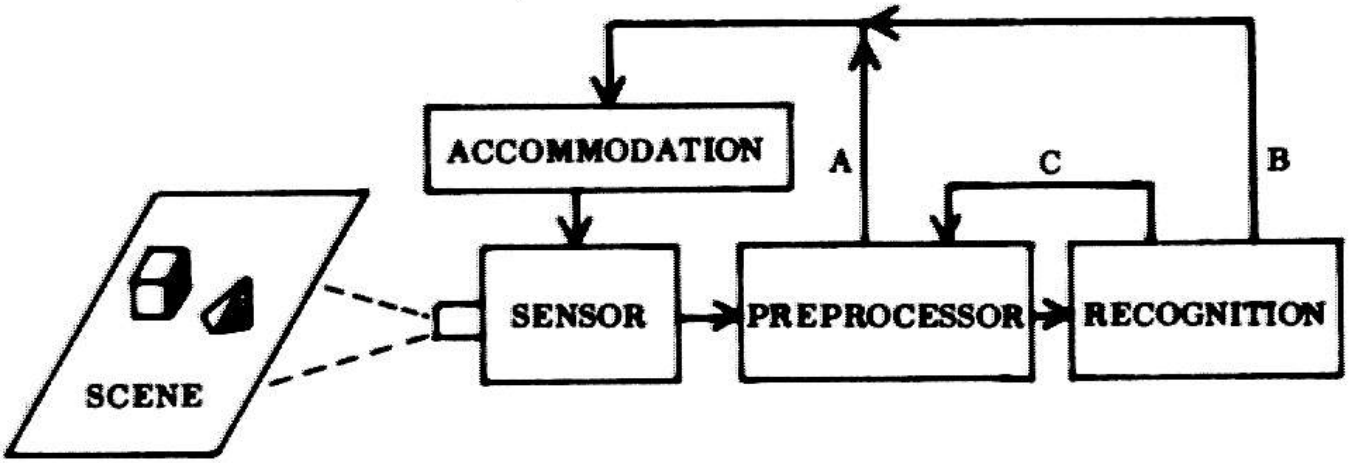


Fig. 1.13 Feedback Oriented Vision System

earlier results.

If the results at any level are found untenable, the system returns to a previous level of processing, modifies parameters and/or processing sophistication, and again attempts to construct a viable representation of the visual world. The cause of failure will often suggest specific ways in which the lower level routines can improve their performance.

Hierarchical evaluation is no longer a novel idea. Fischler [1968] applied simple shape and continuity tests to the output of a curve-fitting operation. Anomalies detected at this level prompted modifications in parameters (eg. sampling interval, intensity threshold, etc.) of the contour follower which provided the initial edge points.

This control was on the level of Loop C in Figure 1.13. In our system, it represents feedback from the line completer to the edge follower. Our present work extends this idea to Loops A and B. The visual sensor is, for the first time, controlled directly by the requirements of the system functions that must use the data. (Loop A corresponds to accomodation requests by the edge follower and by the line completer when it needs more detail. Loop B is used to tune the image for selective acquisition and later to help the feature verifier validate a recognition decision.)

These paths allow accommodation to be an integral part of the overall perceptual strategy, eliminating the

need to manually tune the camera.

1.4.1.4.2 ACCOMMODATION PHILOSOPHY

Accommodation is a specific application of hierarchical evaluation. The immediate goal is the success of a particular perceptual function (Figure 1.14). Accommodation is derived from models of the interaction of camera parameters with scene characteristics. These models in conjunction with the computer's a priori and acquired knowledge of the environment predict the composition of an image. Accommodation is initially set so that the predicted image satisfies the known information requirements of the function. The appropriateness of the initial image will depend on the accuracy and detail of the computer's information.

The performance of the function with this accommodation is evaluated. Accommodation problems are suspected when the results do not confirm prior expectations. A diagnosis is performed to determine whether re-accommodation is likely to reconcile the difference. If so, information gained from the initial failure is used to refine specific accommodations. This cycle is repeated, iteratively improving the accommodation, until either the function succeeds or the diagnosis routine is satisfied that accommodation is not at fault. (In the

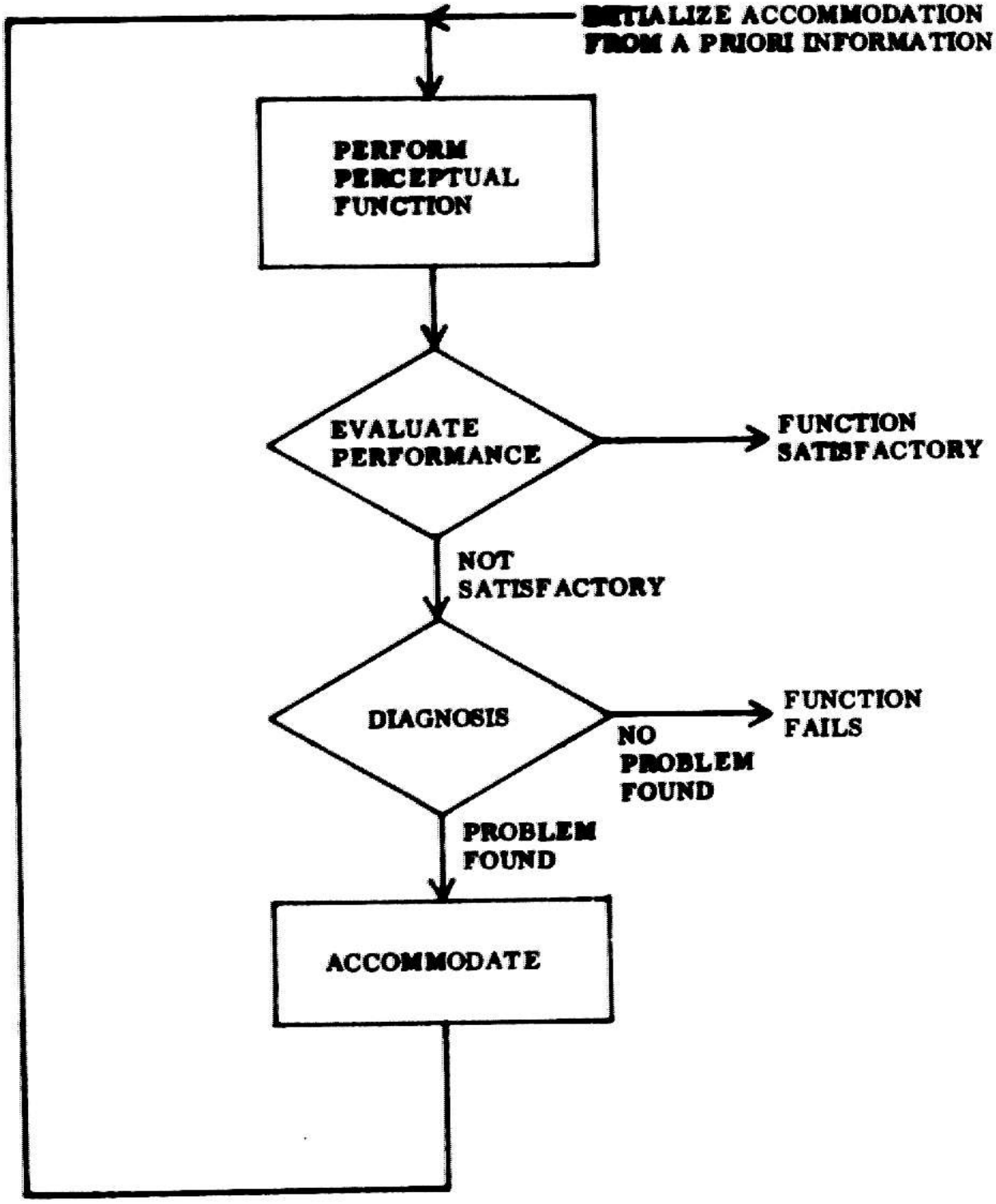


Fig. 1.14 Performance Feedback Paradigm for Iteratively Optimizing Accommodation

latter event, the original expectations must be questioned.) The process of optimizing accommodation by iteratively improving a performance function will be called PERFORMANCE FEEDBACK.

Performance feedback in the context of the edge verifier implies that the camera parameters should be tuned to enhance local contrast at the point where the edge is expected. (Contrast is the image characteristic most directly related to the success of the verifier.) If the edge is then detected, the choice of accommodation is confirmed. Otherwise, accommodation is refined.

In the context of a specific perceptual function, the most appropriate criteria by which to evaluate accommodation is the overall performance of that function. To emphasize this philosophy of accommodation the performance loop can be viewed from another perspective. In Figure 1.15 the optimization of accommodation is shown to be the principal goal. The verifier can be viewed in the above example as an elaborate criterion by which accommodation is optimized.

In this capacity, the verifier defines a complex set of decision criteria to which a suitable accommodation must conform. It is usually possible to isolate individual components of this decision, such as the desirability of high contrast. However, it is not feasible to base a final decision on a single aspect because of the simultaneous and

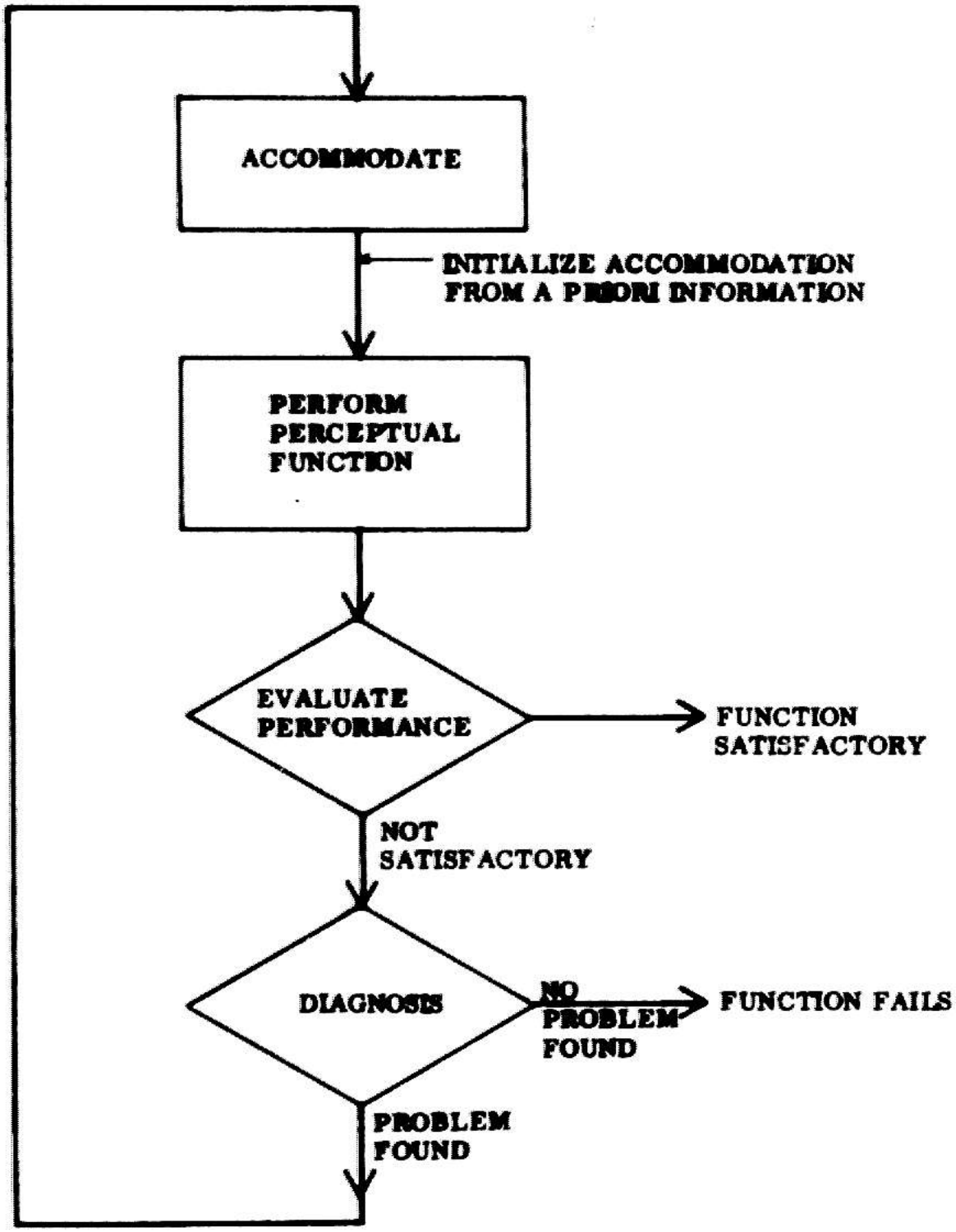


Fig. 1.15 Inverted Performance Feedback Loop

often conflicting demands of other factors, such as the need to minimize noise.

Performance feedback is, of course, only the most immediate level of hierarchical evaluation. Continued success at each subsequent level of processing reinforces confidence in previous results (and in the accommodations used to obtain those results). For example, successful line fitting confirms the goodness of the local contrast accommodations used by the edge follower. (If there are gaps in the edge, extraction can be repeated at those points with locally optimized accommodation.) We have also seen how the results of the validation test (Figure 1.12) can be used to refine the accommodation used for acquisition.

The evaluation hierarchy extends, as far as the perceptual process is concerned, to the level of a recognition decision. (The recognition will, of course, itself be judged by the success of subsequent behavior predicated on that decision.) If the final edge description fails to correspond with any known object model, the verifier can investigate questionable edges with more sensitive accommodations. If, on the other hand, the edges correspond to an object other than the one being sought for a task, then the accommodation used for initial acquisition could be refined.

1.4.2 ADVANTAGES OF AN ACCOMMODATIVE VISION SYSTEM

An effective way to demonstrate the advantages of a new system is to contrast it with known limitations of previous approaches. The strict, hierarchical structure of the initial hand-eye vision system (Figure 1.3), like most early attempts at pattern recognition, followed the general paradigm outlined in Figure 1.16. Most of the emphasis in this early work was placed on the problem: given a suitable representation of an object, identify that object in terms of a set of prototype models.

1.4.2.1 OPERATIONAL LIMITATIONS

The process of obtaining that suitable representation from the raw digitized image data was considered to be a tedious preliminary to the "interesting" recognition aspects. Accordingly, the environment was purposefully constrained to minimize any difficulty. (The cubes used in hand-eye work, for example, were uniformly white and presented to the camera against a sharply contrasting black background.)

Lacking accommodation, the experimenter would manually tune the system. He would adjust the lighting and camera sensitivity to obtain an image on the television monitor that, to him, best emphasized the object he wanted

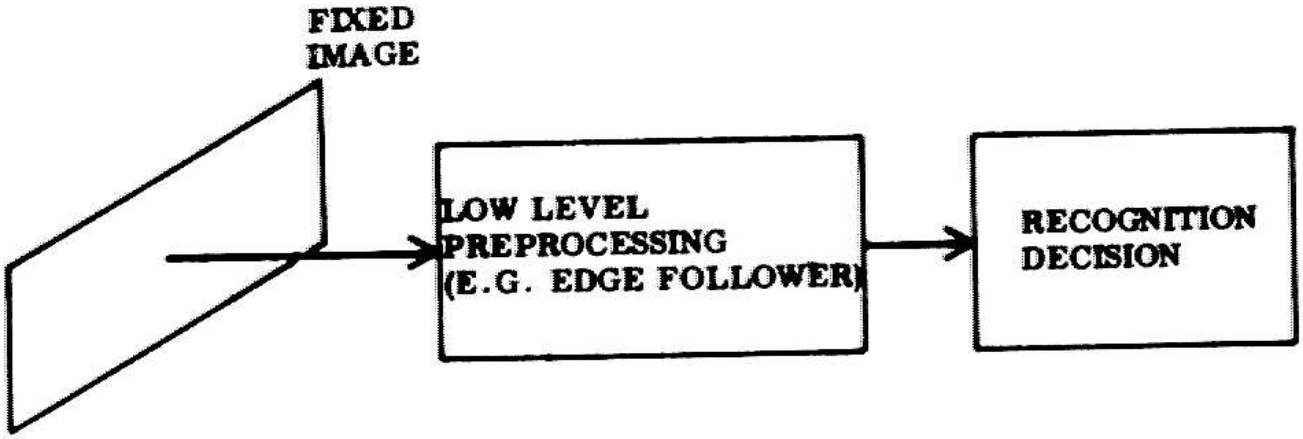


Fig. 1.16 Hierarchical Pattern Recognition Paradigm

the machine to recognize.

Recognition is also the ultimate criterion used to evaluate the effectiveness of automatic accommodation. One is tempted to draw the superficial analogy that HIERARCHICAL EVALUATION is, in a sense, a bootstrapped simulation of manual accommodation. In fact, there are essential differences, and these differences account for the superiority of the automatic method.

A man consciously perceives a scene in its totality. His access to scene characteristics at isolated local points is unavoidably biased by his total impression. Machine perception, however, depends on the local intensity distributions at specific raster points. The experimenter adjusts the camera to achieve a single image that optimizes his global impression. However, this global optimality is necessarily a compromise that guarantees a non-optimal image in particular localized areas.

It was soon realized that the image that gave the best visual impression on a television monitor was not necessarily the best source of visual information for a recognition program. Many frustrated workers found that, despite their best efforts, there was only one effective way to tune the camera: run their programs and play with the various sensor parameters until an image was obtained on which the algorithms would function properly. A slight

improvement was the use of graphics to display the computer's numeric impression of the intensities in a particular region of interest. This enlarged the scope of practical manual adjustments to include setting the quantization window to emphasize a selected edge.

The basic problem remained that the computer was limited to the use of a single image. Consequently, the effectiveness of the refined manual technique was limited to relatively simple scenes (such as a white cube on a black cloth) where:

1. the important edges were all in the same contrast range,

2. the total level of detail contained in an image sharp enough to see the desired features, would not overwhelm the processing capacity of the computer.

(M. Kelly [1970] tried to use this technique to arrive at an appropriate image in which to recognize facial features of people. His conclusion: "too time-consuming and so error prone as to be impractical".)

Automatic accommodation, by contrast, makes it practical to obtain many images of a single scene. Each image can be optimized for local data in a particular part of the scene, according to criteria that suit a particular perceptual function.

With hindsight, the early preoccupation on the recognition aspects of computer vision reflected a widespread avoidance of many of the more difficult problems in making a machine see. In particular, starting with an image which emphasizes what is being sought, eliminates a substantial amount of irrelevant information. This aspect of selective attention must be automated, before machine perception can be practical in a real world environment. One of the principal goals of this thesis is to enable a machine to obtain such images by itself.

The trial and error techniques of manual tuning were crude forms of accommodation. However, the criteria that were used to manually evaluate images and the performance of algorithms were too ill-defined to be formalized for use by a machine.

I.4.2.2 FUNCTIONAL LIMITATIONS OF A NON-ACCOMMODATING SYSTEM

1. No concept of Attention:

The high level decision process has no control over the sensors. Thus, the machine cannot exercise any discrimination over what objects will appear in an image and ultimately be presented to the recognizer. Selective attention can be realized on such a system only by closing the accommodation loop manually. The alternative is a constrained environment that contains only the objects one

wants the machine to recognize. Accommodation removes the need for many of the environmental constraints (eg. high contrast blocks, uniform backgrounds, strong lighting, etc.) required by earlier systems.

2. Inappropriate level of image detail:

The amount of detail required in a picture depends on the current perceptual function (edge following, texture analysis, color etc.), on the (dynamic) state of knowledge, and on the contents of a particular region of the scene. No single image will suit every application. It will always contain TOO MUCH or TOO LITTLE detail for any particular function.

The goal of accommodation is to obtain the simplest picture with sufficient detail. People, on the other hand, are inclined to manually adjust the camera to obtain the sharpest possible image. A sharp picture will, of course, enhance edge boundaries, making it easier to trace cubes against homogeneous backgrounds. But, in a less contrived situation, the extra textural detail contained in a sharp picture may do more to obscure the desired features than any help the enhancement may provide. Too much detail is also very uneconomical in terms of core storage and sensory channel utilization.

3. Inappropriate level of effort:

A hierarchical vision system is characterized by a fixed set of routines. These are applied in a fixed hierarchical order to transform a raw image into a set of recognized objects. There is no capability to adapt the level of processing to the difficulty of the scene. Consequently, it is often necessary to extract superfluous detail over most of the image in order to obtain sufficient detail in the weaker areas. Very sophisticated processing is, however, too expensive to apply over the whole scene. Furthermore, the system would be unable to cope with the volume of information that would be obtained by this indiscriminate acquisition of fine detail. Thus, there will always be occasional regions where the data extracted by some standard level of processing is insufficient.

By contrast, the feedback organization of our current system encourages the use of cheap operators over those parts of a scene that contain adequate contrast. Because of accommodation, these regions usually comprise the majority of the scene. More sophisticated processing is then applied only in the extremely troublesome areas that resist the effects of accommodation. Usually, these will be regions where a feature, expected on the basis of a model suggested by other visible details, has not yet been seen. In these circumstances, expensive processing, such as contour averaging, can be tailored to emphasize what is specifically being sought. This is far more efficient than bulk, high

power processing. Furthermore, the context, established by a specific expectation, provides an effective way to discount the irrelevant details and random noise that will inevitably be picked up by a sensitive operator.

A related advantage is the ability to efficiently utilize special purpose functions. Properties, such as color and depth, can be obtained at selected points where the usefulness of such knowledge justifies the cost of obtaining it.

In summary, the flexibility afforded by accommodation allows economical strategies to be formulated. These strategies utilize the information on hand in the context of the current task to select what additional information to look for. Accommodation helps the computer to see that information. These advantages are incompatible with the older hierarchical system organization shown in Figure 1.16.

1.5 ORGANIZATION OF THESIS

The remainder of this thesis is organized into seven chapters:

Chapter 2: We outline the current accommodative capabilities available to our system. Analytic expressions are developed that formalize how these parameters influence the image characteristics.

Chapter 3: We utilize the models developed in Chapter 2 to investigate how the accommodations interact with the scene characteristics to determine what will appear in the image. We introduce heuristic and analytic criteria of what constitutes an appropriate image for a task. The use of accommodation to reconcile the characteristics of the sensors and the requirements of a task is demonstrated by example.

Chapters 4-7: We discuss in detail how accommodation improves the performance of the perceptual functions that comprise the system shown in Figure 1.12. In each chapter we describe the intended operation of a particular function and devise a heuristic criterion for evaluating its performance. We next develop the theoretical and practical limitations that affect performance and reliability. The characteristics of the most appropriate image are defined for each function in terms of these limitations and any relevant a priori information about the scene that may be available. Finally, diagnostic routines are developed that can recognize departures from the desired image characteristics and initiate the appropriate remedial accommodations. The diagnostics are applied in a sequence likely to cause accommodations to be tried in a cost-effective order.

These components are integrated according to the performance feedback paradigm (Figure 1.14). The resulting

systems demonstrate the ability of accommodation to improve the reliability and performance in a variety of practical perceptual functions.

The unifying context of a complete vision system did not exist at the time this work was completed. Consequently, the advantages of and the requirements for integrating these functions into an overall system were not considered beyond the conceptual level in this introduction.

Chapter 8: We summarize our results and review the principal advantages of accommodation. The generalization of these advantages to other perceptual systems is discussed. We close with suggestions for extending the current work in the unifying context of a sophisticated vision system.

CHAPTER II: ANALYTIC MODELS OF ACCOMMODATION

In this thesis we are concerned with developing perceptual strategies which effectively utilize our system's accommodative capabilities. To approach this problem formally requires

1. a precise definition of the characteristics that make an image appropriate for a specific task and

2. analytic models that predict how the various accommodative parameters influence these image characteristics.

In this chapter we will model the influence of focus, lens-length, lens-aperture, quantizer digitization window, vidicon target voltage, and color filters on the composition of an image. Sobel [1970] studied the effects of camera orientation (pan, tilt) and lens magnification. His models are necessary to shift from coarse coverage of a wide field to high resolution in a local area, depending upon task requirements.

II.1 NEED FOR PREDICTIVE MODELS

The accommodative parameters determine which aspects of the scene the camera will emphasize. Models are needed to guide the search through the space defined by these

parameters, Without models accommodation can be evaluated only in terms of the performance of a task. The cost of this criteria and the complexity of the space make blind optimization unfeasible. Models indicate the most effective accommodations for acquiring specific information. They also establish the optimality of an accommodation for a particular task.

A second justification for accommodative models is the occasional need for absolute comparisons of photometric quantities obtained with different accommodations. In this situation, it is necessary to parametize the relation between an observed quantity and the value of the corresponding property in the scene, expressed on an absolute scale. Sobel [1970], for instance, used his orientation model to coordinate views of the world, obtained at various camera positions with respect to an absolute reference frame.

II,2 DESCRIPTION OF CAMERA (HARDWARE) CAPABILITIES

Our visual sensor (shown in Figure 2.1a) is a standard vidicon television camera (Cohu [1964]). The video output is sampled 333 times/horizontal scan line and quantized into 16 discrete levels. The camera has been modified (Figure 2.1b) to provide computer control of the functions enumerated in Chart 2.1. The pan-tilt head, lens

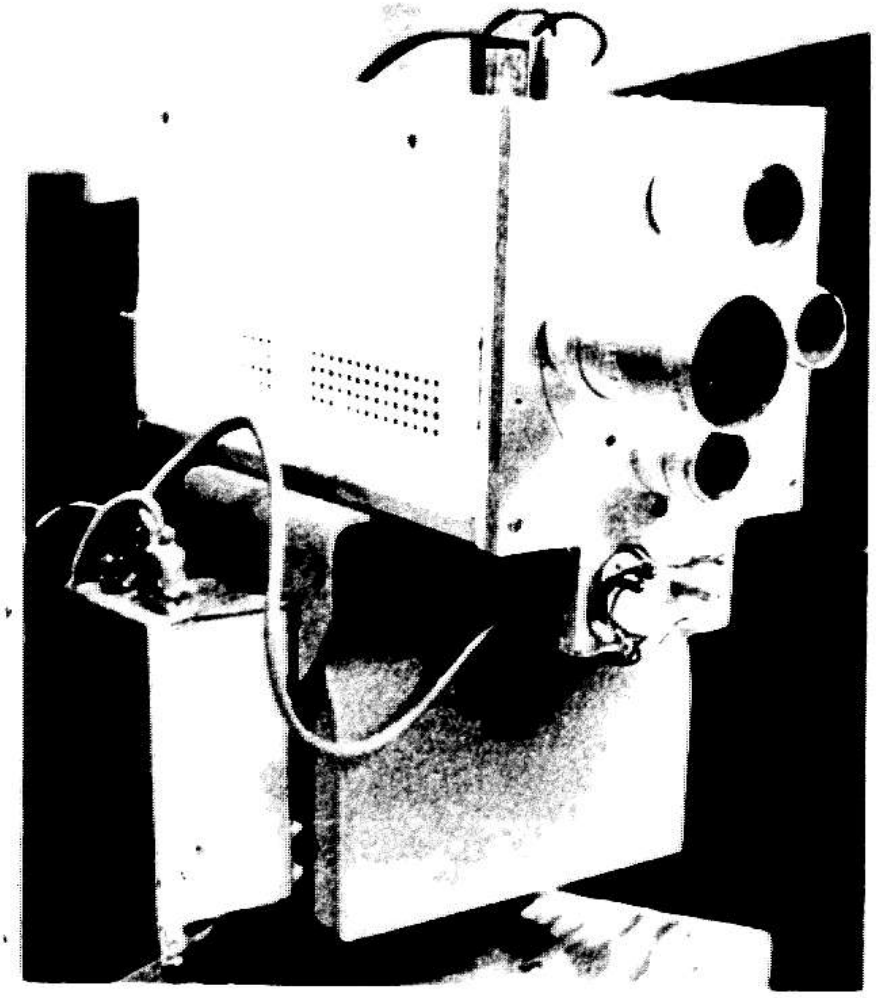


Fig. 2.1a Instrumented Conu Television Camera

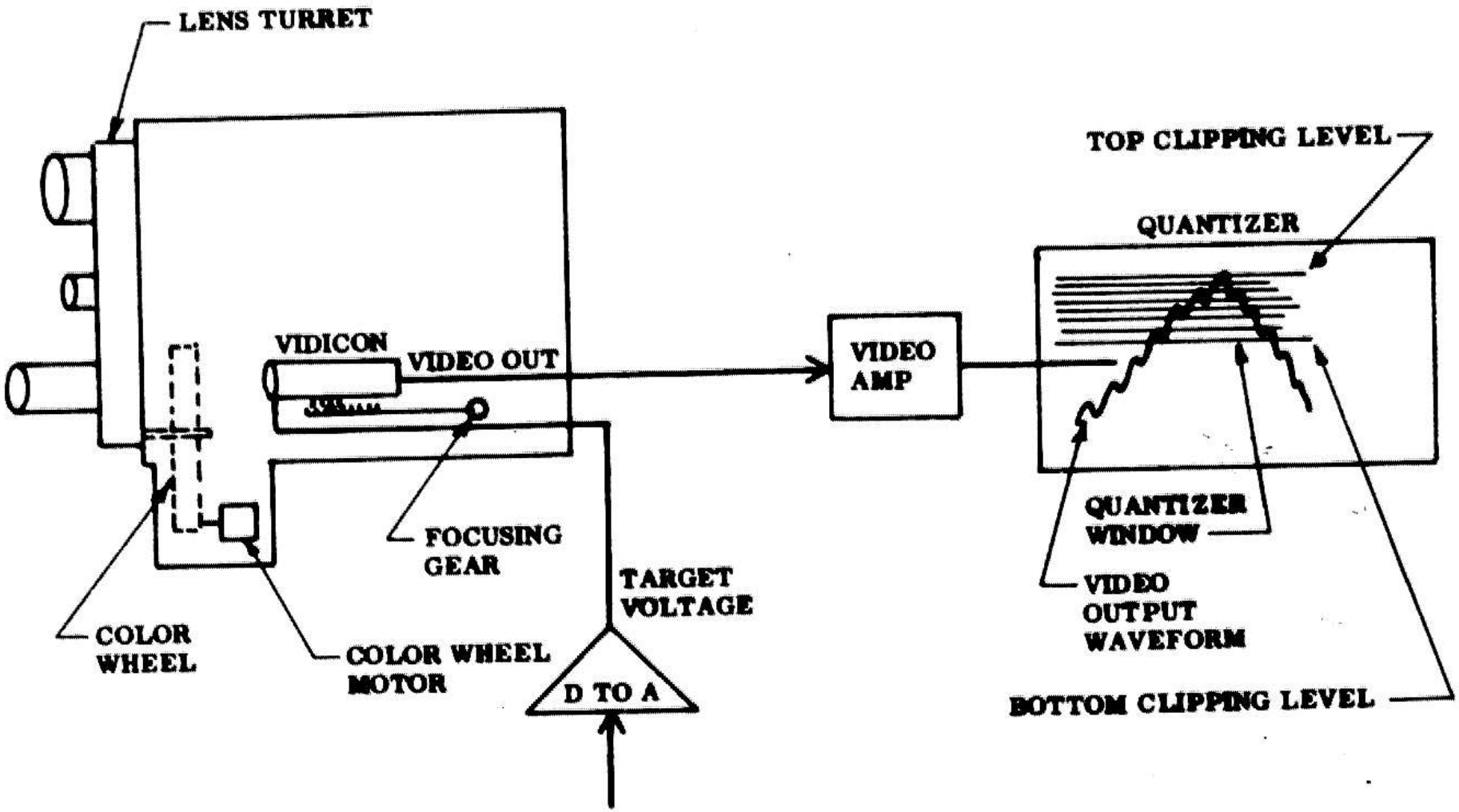


Fig. 2.1b Camera Accommodations Under Computer Control

Accommodations	Accommodation Capabilities
Orientation: Pan-Tilt head	$0^\circ < \text{Pan} < 360^\circ$, $-45^\circ < \text{Tilt} < 0^\circ$
Optical: Lens turret	Lens selection: 1", 2", 3", 4"
Iris	1" - 3" lens $1.4 < f^\# < 22$ 4" lens $2.8 < f^\# < 22$
Color filter wheel	located between lens and vidicon choice of 4 filters: red, green, blue, clear (see spectral response curves, Figure 2.17)
Focus	vidicon moved in and out from lens
Electrical: Sensitivity (vidicon target voltage)	$7 < E_T < 40$ volts
Quantizer	signal range Bvolt < v < Tvolt digitized into 16 levels: Bvolt = $0, \frac{1}{8}, \frac{2}{8}, \ldots, \frac{7}{8}$ Tvolt = $\frac{1}{8}, \frac{2}{8}, \ldots, \frac{8}{8}$

Chart 2.1

Accommodation Capabilities

turret, and focus motor drive are standard manufacturer options that were augmented with feedback pots and control logic to allow automatic servoing. The color wheel allows random access selection of any of four filters in 1/5 of a second. The filters are physically placed between the lens turret and the vidicon.

The camera's auto-target circuit was modified to allow computer selection of 64 discrete target voltages between 7 and 60 volts. This circuit's original function was to maintain a uniform signal level. It did this by adjusting the target voltage according to the average level of illumination, reaching the vidicon. This voltage level is now used as an upper limit to protect the vidicon in the event of program failures. It also serves as a reference by which to set sensitivity in the absence of any knowledge about a scene. From 1/4 to 1/2 second (about 10 television frames) are required before an image stabilizes after a change of target voltage.

The quantizer provides a 16 level window , from 1/8 to 1 v. wide, over the 1 v. working range of the video amplifier. The bottom of this window can be positioned only at integral multiples of 1/8 v. from 0 to 7/8 v. The window can be changed in one milli-second.

Many of these accessories have been available ever since the camera was interfaced with the computer. However, little emphasis was placed on utilizing them under

program control prior to the present work.

II,3 ELEMENTARY THEORY OF CAMERA ERATION

Over the years, photographers and television engineers have developed a body of theory that bears, at least peripherally, on this current work. This theory has seldom been related to the problems of computer vision. The purpose of this section is to provide an integrated interpretation of those results, used in the applications-oriented chapters that follow.

II,3.1 ELECTRICAL CHARACTERISTIC

II,3.1.1 VIDICON

The heart of the television system is the vidicon tube. It uses photoconductivity to convert an optical image into an electron flow. The basic theory of vidicons is well-known (Fink [1957]). This tube is infamous, however, for its non-linearities and other quirks which are constantly uncovered in practice. The tube appears to exhibit unpredictable hysteresis effects depending, for instance, on such obscure factors as the spectral composition of recently observed images.

The vidicon is very sensitive for differential type

intensity comparisons (as used in edge following) but difficult to use for obtaining reliable photometric measurements of absolute light intensity (as required for color perception, for example). We are presently considering requirements for a new camera in light of this experience.

II.3.1.1.1 PRINCIPLES OF OPERATION

In this section we will briefly derive a photometric transfer function that summarizes the vidicon's ideal mode of operation. With reference to Figure 2.2a, light is focused through the transparent conducting signal plate at the front of the tube and is absorbed by the target. The target is a thin layer of photoconductive material. The local resistivity of each elemental area decreases inversely with the amount of light flux falling upon it. An electron beam, magnetically deflected and focused, is swept periodically over the back face of the target. At each point, charge is deposited until the local surface has been reduced to cathode potential. This charge is held between scans by the capacitive effect of the signal plate. A small amount will leak off, due to the photoconductivity of the target. On the next scan, the electron beam will deposit sufficient charge to drive the surface back to cathode potential. This sequence causes a displacement

current to flow through the load resistor R. The resultant voltage is capacitively coupled to a linear video amplification chain.

Figure 2.2b is an equivalent circuit model of the vidicon (Eppler [1964]). Consider the elemental area $\partial A1$ represented by R_1, C_1. The electron beam charges C_1 through the parallel combination of R_L and R_1. Since $R_L << R_1$, the capacitor will charge to

$$e_1(t) \approx E_T (1 - e^{-t/R_2C}) \qquad (2.1)$$

(E_T is a constant, nominally in the range from 5 to 40v.) If the scanning beam moves slowly enough, C_1 will be fully charged. If ∂t is the time the electron beam is positioned on $\partial A1$, this condition requires that

$$\Delta t \geq 5 R_L \times C \qquad (2.2)$$

when the beam passes on to the next area (eg, the rotary switch advances), C_1 will begin to discharge through R_1

$$e_1(t) \approx E_T e^{-t/R_1C} \qquad (2.3)$$

R_1 is determined by L_1, the light incident upon ∂A_1,

$$R_1 \sim K_1(L_1)^{-\gamma} \qquad (2.4)$$

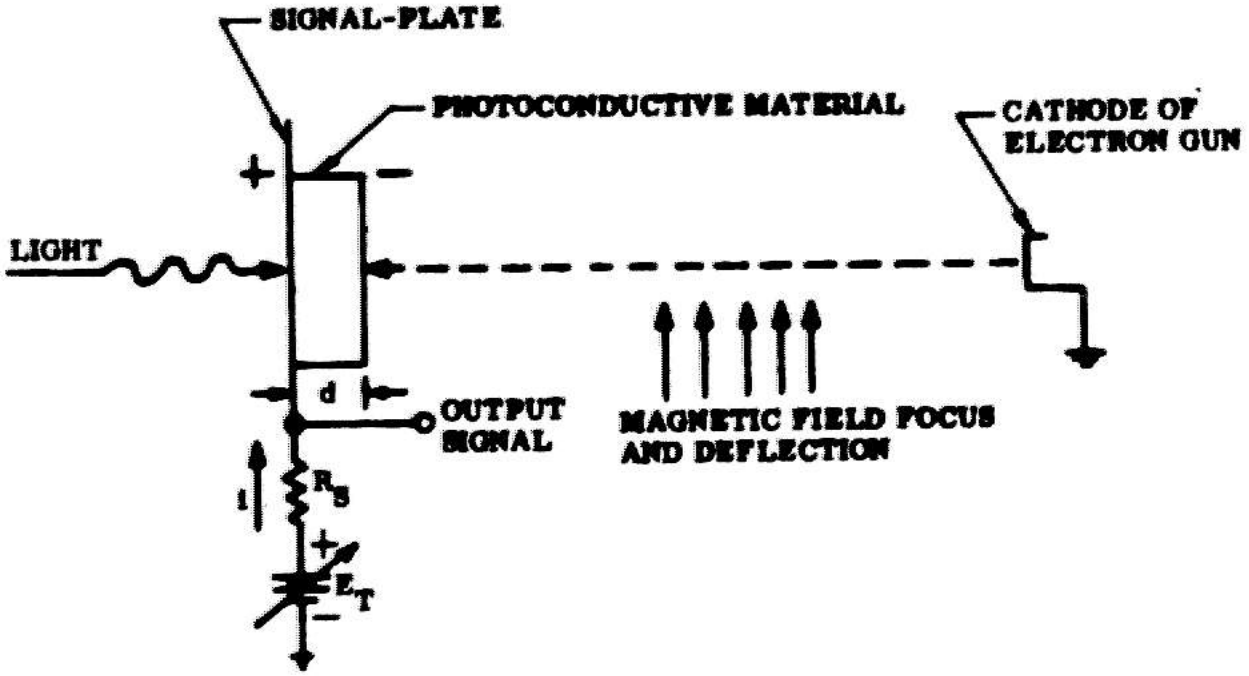


Fig. 2.2a Diagram of Vidicon Camera Tube

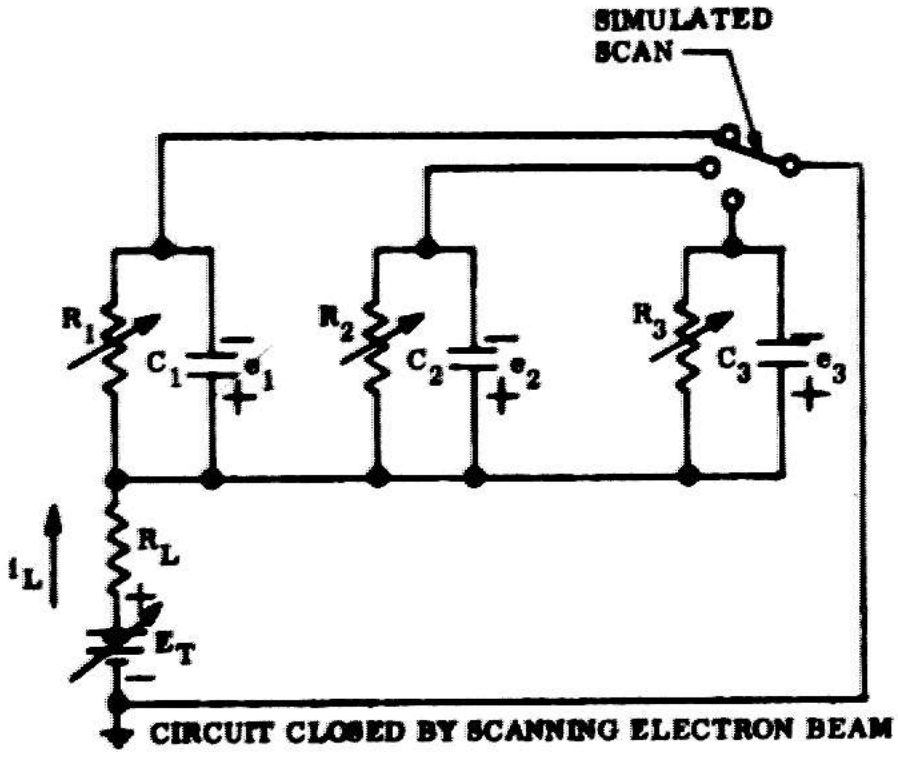


Fig. 2.2b Operation of the Vidicon Camera Tube

Gamma is a fairly constant characteristic of the tube nominally about 0,65. (It is always less than 1, since photoconductivity is a primary carrier generation process).

The effective resistance of R1 depends on the mean light flux during the time, Ts, between scans. Thus, the response of the vidicon represents an integration over Ts. Under the assumption that Ts << R1C, the charge dQ1 which is lost from dA1 between successive scans, is given by

$$\Delta Q_1 = C[e_1(0) - e_1(T_S)] \approx C E_T(1 - e^{-T_S/R_1C}) = \frac{E_T \times T_S}{R_1} \quad (2,5)$$

This charge must be replaced on the next scan, resulting in a signal current through R2 of

$$i_L = \frac{\Delta Q}{\Delta t} = \frac{E_T \times T_S}{R_1 \times \Delta t} \quad (2,6)$$

For later reference, we define E_T/R_1 to be the photocurrent i_p, excited by the incident light flux. From Equations 2,4 and 2,6, the video output current is thus proportional to

$$i_L \sim K_2 \times E_T \times L_1^{\gamma} \quad (2,7)$$

This equation is an elementary photometric model. It relates a light level L1 to a signal current iL. A linear

gain function can be applied to this current to obtain the voltage that is digitized for entry into the computer. This simple model unfortunately breaks down because of peculiarities in the photoconductive effect (RCA REVIEW [1951]).

The Cohu camera uses an RCA-8507A vidicon. The light transfer characteristics of this tube are shown in Figure 2.3, (adapted from the manufacturer's application sheet). The slope of lines on this log-log plot is what we have called gamma. It closely approximates .65 over the 411 range in target voltages and 10,000:1 range of incident light, covered in the figure. (Figure 2.3 was originally parametized only in terms of dark current. The target voltages that are shown were obtained from Figure 2.4. They represent average values at the indicated dark current.)

Signal out vs. target voltage at a constant illumination level also plots on log-log paper as a straight line. This is characteristic of a power function, rather than the linear relation expected from Equation 2.7. Thus, from empirical evidence Equation 2.7 is superceded by

$$i_L = K_3 \times L^{0.65} E_T^{\alpha} \quad , \quad K_3 = 2.798 \times 10^{-9} \qquad (2.8)$$

α can be calibrated from the slope of the best least square fit of log i_L vs. log E_T. α was calibrated with a flat-white specimen of brightness L=20ft-c. α=1.4

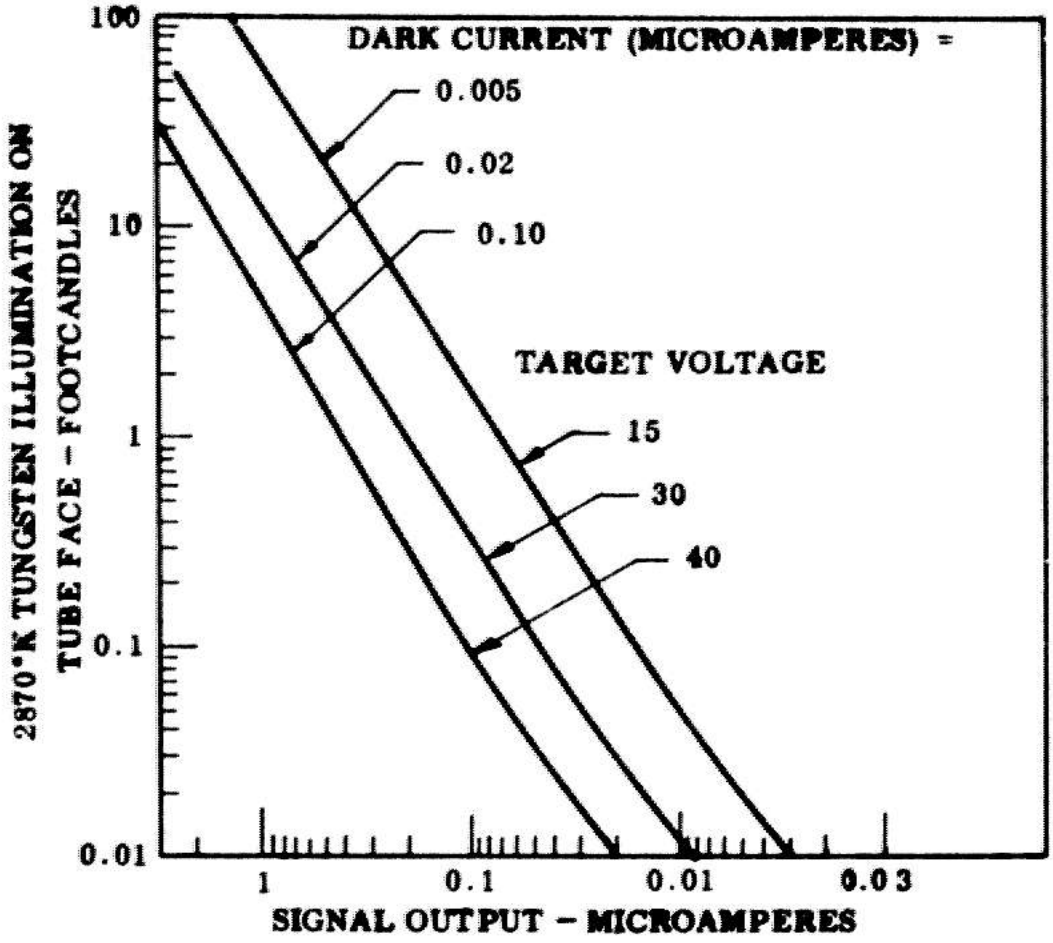


Fig. 2.3 Light Transfer Characteristics

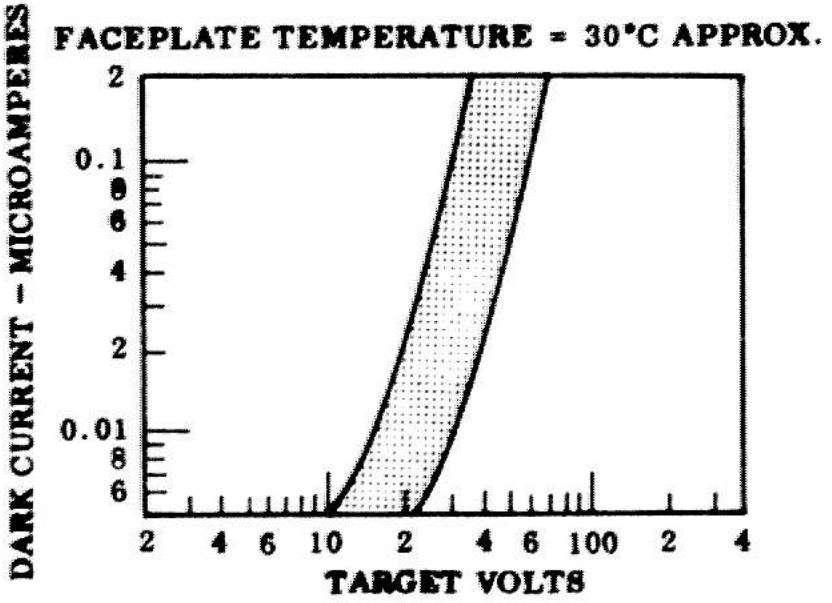


Fig. 2.4 Range of Dark Current

provided a consistently good global fit over the range $10<E_{\alpha}T<35$. Locally, slope varied unrepeatably from 1.2 to 1.6. A variation of this magnitude was also encountered as the brightness of the specimen was varied between $5<L<25$ ft.c. The mean value of α tends to increase toward 1.6 when L drops below 10ft.c. The incremental value of α at a particular light level and target voltage can be established experimentally, using Equation 2.9,

$$\alpha = \frac{\log (V_2) - \log (V_1)}{\log (E_{T_2}) - \log (E_{T_1})} \qquad (2.9)$$

A constant α is not sufficient for photometric accuracy. However, the value $\alpha=1.4$ (used with Equation 2.8) does provide a satisfactory, relative ordering of intensities accumulated over a wide range of target voltages.

α is even worse behaved for colored specimens. The spectral sensitivity of the camera is a complex function of target voltage (see Chapter 7). As a result, the effective value of α varies from 1 to 5, depending on object color. This variation defies concise analytic statements.

In this thesis, we will assume the use of white objects and thus a nominal value of $\alpha=1.4$, unless stated otherwise.

The instantaneous video signal represents the difference of the light level at a raster point from a

reference dark level. (The dark level is generated within the vidicon, when it is blanked during retrace.) This signal is a.c.-coupled to a linear amplifier chain. The overall transfer function, relating the voltage at the input of the digitizer to the light flux density, incident on the vidicon faceplate, can be expressed in the form

$$v_{out} = K(L)^{\gamma} E_T^{\alpha} \quad , \quad K = 0.0095 \qquad (2.10)$$

K and α have been calibrated for a scene illuminance of 20ft.c., f# 1.4, and a .5 (neutral) density filter: K=.0095, α=1.429. (To reiterate, L in Equation 2.10 is the light flux density that actually reaches the vidicon surface, after all optical attenuation.)

The amplifier has a 1.5 v. dynamic output range. This range is clipped by the digitizer to a maximum absolute window of 0 to 1 v. Equation 2.10 determines the range of illumination that will produce signals lying in this window at any E&T.

II.3.1.1.2 INFLUENCE OF TARGET VOLTAGE ON IMAGE

The camera sensitivity affects the composition of an image in two ways:

1. It establishes the range of scene brightnesses that will be linearly encoded by the quantizer.

2. It affects the signal gain and, thus, determines the minimum contrast that can be resolved. These effects are not independent; the minimum resolvable contrast is constrained by the dynamic range. If E$_t$T is turned up too high, the gain in contrast will be nullified by clipping in the quantizer.

In practice, E$_t$T is also constrained by the maximum voltage the vidicon can tolerate at prevailing light levels without blooming. This condition destroys the detail near the scene highlights. The camera can also sustain permanent physical damage when operated in this state.

This limitation is most significant in a scene, encompassing a wide range of brightnesses. The brightest object establishes the highest allowable target voltage. This limits both dynamic range (eg. the dimmest object that can be perceived) and minimum resolvable contrast.

In our system, the auto-target circuit physically constrains E$_t$T. It is adjusted to override the computer-selected target voltage when it exceeds a safe level. The auto-target was designed to maintain a constant average signal level over the entire scene. The effective maximum target voltage will thus vary as $L^{(-gamma/\alpha)}$ (see Equation 2.10). (The maximum voltage levels are about 35$_v$. at 20 ft.c. (a flat-white object in ordinary flouresent room illuminations) and about 20v. at 160 ft.c. (illumination, supplied by a 1000 watt, diffused

xenon arc source,) Unfortunately, this scheme does not provide adequate protection from small, glossy highlights. Such highlights can concentrate a considerable amount of energy onto a small area of the target without significantly affecting the global average.

II,3.1.1.3 NOISE SOURCES

To utilize the camera effectively, one must be aware of its limitations. There are four principal noise sources (Schreiber [1964]):

1. Signal Shot Noise

The measurement of light levels is basically a process of counting photons or, in practice, of counting electrons dislodged by photons. Statistical fluctuations limit the accuracy with which the actual light level can be measured. This inherent source error is known as shot noise.

The Poisson process is the standard model of particle arrival assumed in the analysis of signalling processes. With steady illumination the number of photons (N) that arrive in a time period (T) determine a ratio, N/T, that will fluctuate about the mean photon arrival rate, α, in a manner predictable from Equation 2.11.

$$\text{Prob (N arrivals in time T)} = \frac{(\alpha T)^N e^{-\alpha T}}{N!} \qquad (2.11)$$

For large counts ($\alpha T \gg 1$), the number of particles arriving in time T is closely approximated by a Gaussian distribution of mean (αT) and standard deviation $(\alpha T)^{1/2}$.

The basic noise characteristics are set by the number of discrete particles contained in each sample. The vidicon target converts photons into electrons with a constant quantum efficiency. We will pursue our analysis in terms of the number of carriers, comprising the resultant photocurrent. Rewriting Equation 2.6 to express signal current i_s in terms of the photocurrent, i_p, gives Equation 2.12

$$i_s = \frac{i_p \times T_s}{\Delta t} \qquad (2.12)$$

The total charge storage required to maintain this current is $(i_p T_s)$. Using e, the standard coulombic charge, the number of discrete charge carriers is thus

$$N' = \frac{i_p \times T_s}{e} \qquad (2.13)$$

From the approximations admitted by large Poisson distributions

$$\sigma^2_{N'} = \frac{i_p \times T_s}{e} \qquad (2.14a)$$

but

$$i_s = \left(\frac{e}{\Delta t}\right) N' \qquad (2.14b)$$

Thus

$$\sigma^2_{i_s} = \left(\frac{e}{\Delta t}\right)^2 \sigma^2_{N'} = \frac{e i_s}{\Delta t} \qquad (2.15)$$

Typical values for a weak signal are:

L=1 ft. c.

i_s=1 micro-amp. (nominal value from Figure 2.3 at E.T=20v.)

Δt=1.705*(10**-7)*

(57 micro-sec./333 samples per horiz. line)

(e=1.6*10**-19 coulomb/electron)

Using these values, Equation 2.15 yields a noise current of:

$$\sigma_{i_s} = \left[\left(\frac{1.6 \times 10^{-19}}{1.705 \times 10^{-7}}\right) 10^{-7}\right]^{1/2} = 3.06 \times 10^{-10} \text{ amps} \qquad (2.16)$$

This signal is especially vulnerable to the effects of additive noise prior to any amplification, because the total particle count is then lowest. Let us consider the other dominant low-level noise components.

2. Dark Current

The vidicon target, in the absence of light, behaves

like a leaky resistor. This dark resistance induces a so-called dark current, which acts like a constant bias in the presence of a light-induced signal. Fluctuations in this dark current limit the smallest signal variation that can be detected.

Let i_d be the average level of dark current which is additively combined with signal current i_s. Let i_d' be the leakage current in each elemental area of the target. i_d' is analogous to the role of i_p in Equation 2.12 (but results from thermal rather than light generated carriers). Thus,

$$i_d = \frac{i_d' \times T_S}{\Delta t} \qquad (2.17)$$

i_d' tends to be uniform over the target and to increase directly with temperature.

There are two principal causes of fluctuations in i_d:

1. Scan non-linearities account for variations in Δt. This variability is usually not a significant factor. Magnetic deflection systems are typically better than 95% linear over an entire frame.

2. Shot noise is also inherent in the dark current. Because the mechanism of dark current generation is completely analogous to the effects of a low level light source, applied uniformly over the target, we can adapt

Equation 2.15 to write

$$\sigma^2_{i_d} = \frac{e \times i_d}{\Delta t} \qquad (2.18)$$

In Figure 2.4 we note that the level of dark current is strongly affected by the target voltage. At a target potential of 20 v, (as used in the calculation of signal shot noise), the worst level of dark current is about 2*(10**-8)amps. Substituting in 2.18 and taking a square root yields the noise current.

$$\sigma_{i_d} = \left[\left(\frac{1.6 \times 10^{-19}}{1.705 \times 10^{-7}}\right) 2 \times 10^{-8}\right]^{1/2} \simeq 1.36 \times 10^{-10} \qquad (2.19)$$

The dependence of i*d on target voltage is an important factor to be considered, when deciding optimal sensitivity settings for the camera. An analytic expression of the relationship was derived from an empirical fit to the data in Figure 2.4. The values of dark current (above the level of 2*(10**-8) amps.) are essentially linear on this log-log plot with a slope of 3.97. The appropriate functional form is thus

$$i_d = B \times E_T^{\beta} \qquad (2.20)$$

The calculated best fit for the worst case of dark current

is

$$i_d = 1.37 \times 10^{-13} \times E_T^{3.97} \quad , \quad E_T > 20\ V \qquad (2.21)$$

In the range of E.T from 10v, to 20v,, a passable approximation to a linear fit can be made with a line of slope B=2.0,

Dark current acts as a pedestal under the signal. This limits the range of brightnesses that can be simultaneously included in the 1v. window of the quantizer. The problem is compounded, because low illuminations require a high target voltage to increase gain. This introduces high dark current. Gain and dark current both tend to bias the brighter parts of a scene out of the usable 1v. window.

3. Thermal noise

The output signal voltage is produced by a current flow through the target resistor R.L. This flow induces thermal noise given by Equation 2.22

$$\sigma_{i_r}^2 = \frac{2 \times K \times T_e}{R_S \times \Delta t} \qquad (2.22)$$

K=Boltzman's constant, 1.38*(10^-23) Joules/deg.kelvin

T.e=300 degrees Kel. (nominal)

R.s=(10^6) ohms

From 2.22 the thermal noise current level is found to be:

$$\sigma_{i_r} = \left(\frac{2 \times 1.38 \times 10^{-23} \times 300}{10^{6} \times 1.705 \times 10^{-7}} \right)^{1/2} = 2.16 \times 10^{-10} \qquad (2.23)$$

This value is independent of target voltage but, like dark current, is temperature sensitive.

4. Preamplifier noise

The principal noise source in normal operation appears to be fluctuations in the bias current of the preamplifier. The FET input stage in our camera's amplifier chain contributes an rms noise current on the order of 2.4*10^(-9) amps. (The mean is taken over read out time st.) This noise is a combination of normal shot noise (Sevin [1965]) and induced stray currents, picked up in the noisy electrical environment of a computer room. It is independent of target voltage and light level, but it does depend on such non-optical aspects of the environment as temperature and electromagnetic activity.

The four noise sources are uncorrelated; the overall noise power is consequently given by

$$\sigma_i^2 = \sigma_{i_s}^2 + \sigma_{i_d}^2 + \sigma_{i_r}^2 + \sigma_{i_a}^2 \approx \sigma_{i_a}^2 \qquad (2.24)$$

The significance of this noise is established by comparing it with the maximum signal level (1v.) at the

input to the quantizer. The transfer impedance of the amplifier chain can be inferred from Equations 2.8 and 2.10.

$$G = \frac{K}{K_3} = 0.34 \times 10^6 \qquad (2.25)$$

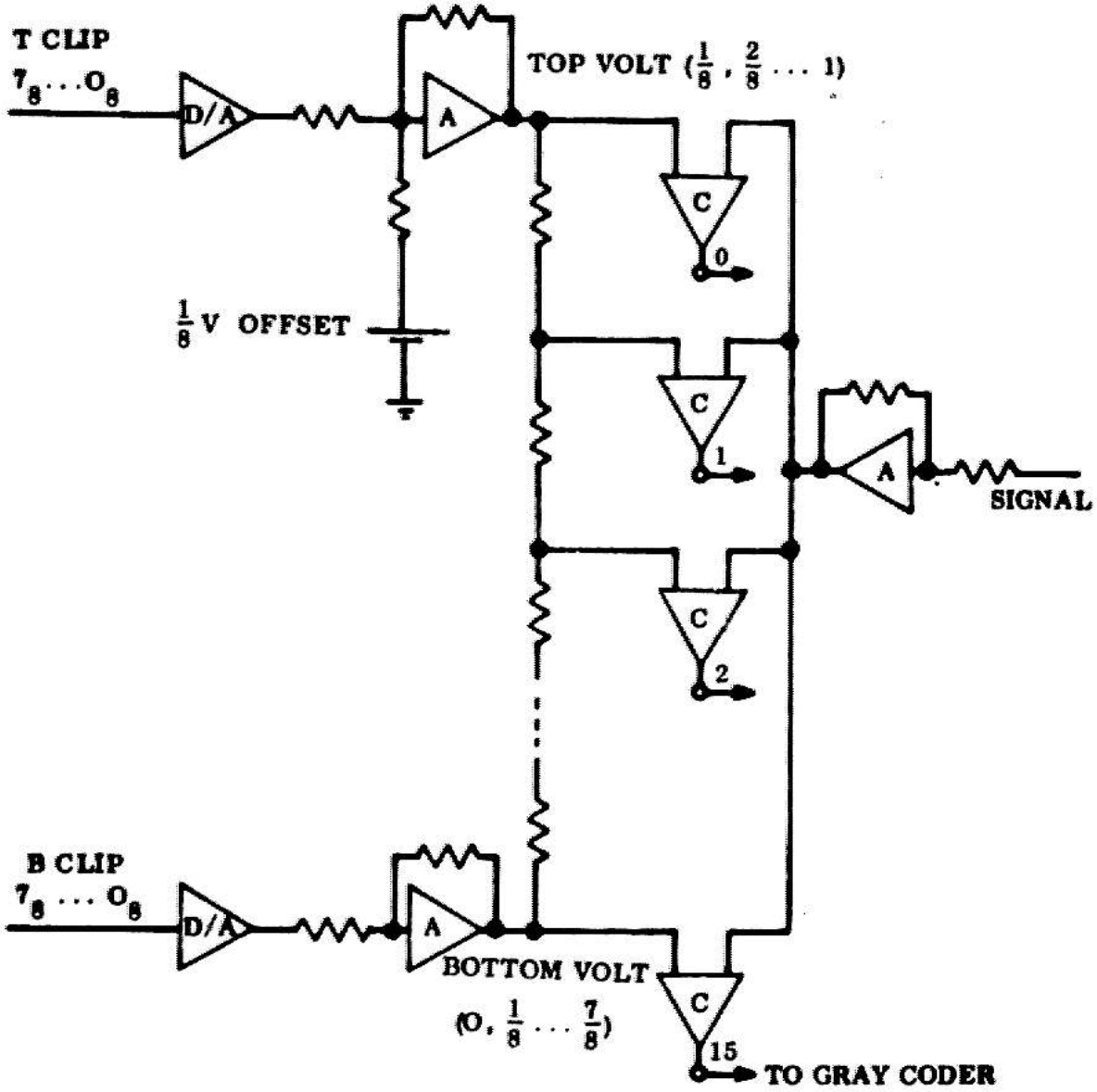
On the assumption that amplifier noise is dominant, the noise level at the input to the quantizer is then (roughly) $8 \cdot 10^{-6} \cdot (-3)v$. The strongest signal is 1v. (limited by saturation). Consequently, the signal/noise is about 125.

II.3.1.2 QUANTIZER

II.3.1.2.1. HARDWARE FEATURES

The quantizer sketched in Figure 2.5 was designed by Wichman [1967]. It converts a selected portion of the 1v. dynamic signal range of the video-amplifier into a 4 bit digital number in real time. (The parallel comparator string was built with Fairchild micro-logic and offers overall conversion times of under 50 nano-seconds/sample.) The outputs of the 15 comparators are subsequently gray-coded to avoid race conditions. The computer translates this encoding into a 16 level intensity in the range from 0 to 15.

The conversion window is selected under computer control. It is defined by two 3 bit numbers



LEGEND D/A - 3-BIT DIGITAL TO ANALOG CONVERTER
A - OPERATIONAL AMPLIFIER, FAIRCHILD μL709
C - INTEGRATED DUAL COMPARATOR, FAIRCHILD μL711

CIRCUIT DETAILS CAN BE FOUND IN WICHMAN [1967]

Fig. 2.5 Simplified Schematic of Fast TV Digitizer With Adjustable Quantization Window

(Tclip,Bclip). These numbers are translated by D/A converters into the voltage levels at the top and bottom of the reference ladder.

(Equations 2.26a,b relate these top and bottom ladder voltages to Tclip and Bclip.

$$\text{Top-Volt} = \left[\frac{1}{8} \blacklozenge (7 - \text{Tclip})/8\right] \qquad (2.26a)$$

$$\text{Bottom-Volt} = \left(\frac{7}{8} - \frac{\text{Bclip}}{8}\right) \qquad (2.26b)$$

Tclip, Bclip are both integers, satisfying the constraint

$$0 \leq \text{Tclip} \leq \text{Bclip} \leq 7 \qquad (2.26c)$$

From these equations it is seen that the top of the ladder can be set to any integral multiple of 1/8v, in the range 1/8v.<top volt<1v. The bottom voltage can be set from 0 to 7/8v, in similar increments.)

With the window wide open (Tclip = 0, Bclip = 7) the 16 intensity levels available in the computer correspond to coarse 1/16v, steps over the full 1v, video range. For the narrowest window each level represents 1/16 of a selected 1/8v, total range. Equation 2.26d relates the 16 relative levels, $0 \leq I \leq I+15$, to corresponding voltages in terms of the absolute range selected by Tclip, Bclip.

$$V(I) = \frac{I}{16} (\text{Top Volt} - \text{Bottom Volt}) + \text{Bottom Volt} \qquad (2.26d)$$

The ability to accommodate the quantization window greatly enhances the effectiveness of the four available bits. The quantizer can resolve intensities in selected 1/8v. ranges to the equivalent of 112 levels (about 6 1/2 bits) over the total 1v. range. This figure is based on the fact that out of the 16 possible levels in any window, the lowest and highest values indicate only that the sample is out of range (low and high respectively). The 14 middle values are valid voltage representations in the sense of Equation 2.26d. There are 8 possible 1/8v. ranges, each contributing 14 definitive levels, a total of 112. (Alternatively, it could be said that the specification of the window contributes three additional bits of information to the 14 linear levels inside the window.)

It is possible to simulate the action of a 6 1/2 bit digitizer over the full dynamic range. A scene is observed through each of the eight narrowest window positions. The window number and value are recorded, when the intensity is in the range $0 \leq I \leq 14$. As stated, this procedure is very inefficient. It should seldom be necessary, however, to sample all eight windows; there is rarely important information in all ranges. Furthermore, fine detail is not necessary except to resolve specific ambiguities. In such

cases, the specific narrow window that brackets the initial data (obtained with coarse quantization) can be selected directly.

II.3.1.2.2 INFLUENCE OF QUANTIZATION ON IMAGE

The influence of quantization depends on how the intensity information in an image is to be used. There are two basic types of intensity measurements:

1. Differential: In applications, such as edge following and acquisition, it is important to know that two intensities differ. How much they differ is less significant. The 16 intensity levels in any window can be used directly without calibrating them to any absolute scale (as with Equation 2.26d). Differential data is consequently less costly to obtain. Prior to the present work, the camera was only used in this way.

The principal influence of the quantization window width on differential intensity measurements is to effect a trade-off between dynamic range, on the one hand, and intensity gain and resolution, on the other. A wide window provides coarse resolution over a broad range of intensity. A narrow window provides fine discrimination in a selected range. The details of this trade-off are summarized by four considerations:

1. Dynamic Range. A narrow window will not detect any edge, both sides of which lie above (or below) the window boundary. Recall the loss of edge B in Figure 1.7.

2. Gain. When two intensities are contained within a quantization window, narrowing that window will enhance the differential contrast between them. The effect is equivalent to pure gain, since any noise that may be present on either signal will also be amplified leaving the ratio of contrast/noise deviations unchanged.

3. Noise limiting. If the window width is narrower than the contrast between two intensities, such that one registers as out of range low and the other is out of range high, then the edge transition will be effectively sharpened by intensity thresholding. This hard limiting eliminates noise on each side. These effects are shown for an intensity profile of an edge in Figure 2.6.

4. Contrast Resolution. Two intensities may be lumped into the same quantization level and so be indistinguishable when viewed through the coarse resolution of a wide quantization window. Narrowing the window may improve resolution enough to detect an edge.

2. Absolute: For tasks such as color analysis or edge detection, using statistical methods, differential measurements are not sufficient. It is important in these cases to obtain the actual luminosity of the specimen

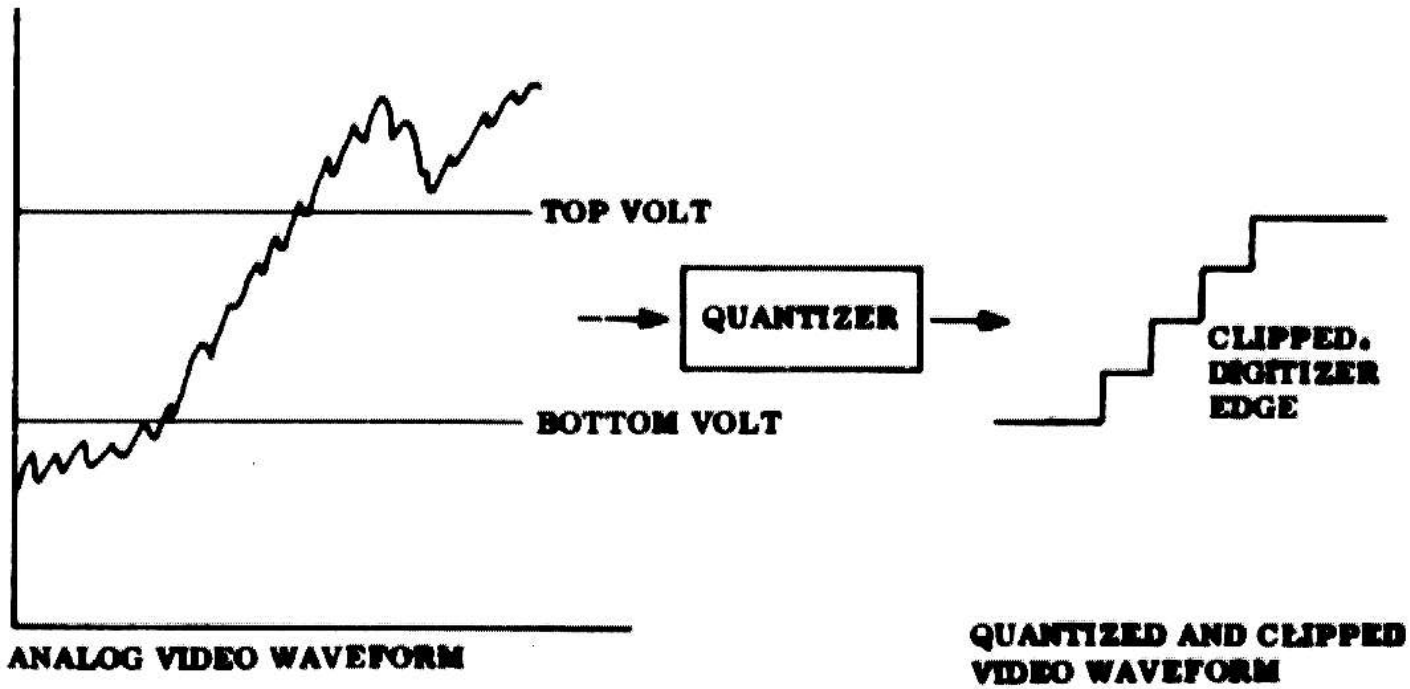


Fig. 2.6 Thresholding and Noise Limiting Effects of Quantization

expressed on an absolute brightness scale. (Absolute measurements are needed, for example, to compare intensities obtained with different color filters. The brightness of colored objects is strongly affected by the color of the filter. A single window that is wide enough to contain the range of intensities from all filters may provide inadequate resolution. The alternative is to use a window optimized for each filter. This would preclude a direct comparison of the observed values.)

Absolute measurements, by definition, require that all intensities be in the linear digitization range (1...14) of the quantization window. Narrowing the window reduces the number of intensity samples in an image that can be validly encoded. There is no concept of contrast enhancement associated with a reduction in window width. The additional resolution simply reduces the uncertainty introduced into the measurement by quantization.

II.3.1.2.3 QUANTIZATION UNCERTAINTY

An individual intensity sample is an estimate of the mean brightness of an object. In the absence of quantization a Gaussian distributed variable will differ from its true mean, u , by an amount less than σu with confidence given by

$$C = \Phi\left(\frac{\partial u}{s_d}\right) \quad , \quad s_d \equiv \text{standard deviation of } u \tag{2,27}$$

Phi(y) is the well-known normal error integral. Its value is the proportion of total area under a normalized-Gaussian curve contained within +/- y units of the mean.

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-y}^{y} e^{-x^2/2}\, dx \tag{2,28}$$

Our present problem is to determine how the introduction of quantization will affect the accuracy and confidence with which the true mean can be estimated.

Intuitively, with very fine quantization, the uncertainty of the quantized signal must approach the inherent uncertainty of the original source, given by Equation 2,27. On the other hand, with very coarse quantization, the uncertainty introduced by the quantization interval must dominate. In the latter case the quantizer can be thought of as contributing additional noise to that already present in the original signal.

These ideas have been formalized by Ross (Susskind [1957]). He determined the accuracy with which the statistics (mean, mean square etc.) of a signal could be recovered from a quantized representation. Specifically, Ross compiled a probability density function from quantized intensity samples. To evaluate the accuracy of this

reconstruction he calculated its moments and compared them with the corresponding ones obtained from the density function of the original time series.

The probability density of the amplitude of a time function relates the percentage of time over some measurement interval that the function spent at each amplitude level in its domain (see Figure 2.7). All statistical parameters of the original time function (eg. the average or mean, the mean square etc.) can be found from its density function.

Quantization can be modeled as a sampling operation, applied to the intensity range of a signal. Ross based his analysis on clever analogies drawn between this model and the well-understood process of time sampling. He reasoned that probability density is a function of amplitude in the same sense that the original function is related to time. Sampling rate, 1/T, for a time signal is determined by the frequency content of that signal. The proper quantizing fineness, 1/q, should thus be determined by the analogous frequency content of the amplitude probability density function.

The following is a sketch of Ross' results. It is my feeling that the general engineering community is not sufficiently aware of the fundamental nature of this work.

Quantizing Theorem: Let $W(x)$ be the probability density of time function $x(t)$. $W(\alpha)$ denotes its Fourier

transform. Let q be the width of a quantization interval, if

$$W(\alpha) = 0 \quad \text{for} \quad |\alpha| \geq \frac{\pi}{q} \tag{2.29}$$

then w(x) may be completely recovered from quantized samples of x(t).

Proof of Quantizing Theorem.

Figure 2.8a compares the continuous probability density function w(x) that was obtained from x(t) with the discrete density function w'(x) developed from digitized samples of x(t). A quantizer converts all values of x(t) within a continuous quantization range into a single discrete value x+o(t) corresponding to the midpoint of the segment. Therefore the distribution function w'(x) of the quantized signal will be zero at all values of x except integral multiples of q. The value of w'(x) at nq represents the fraction of the values of x(t) contained in the interval $nq-(q/2) \leq x \leq nq+(q/2)$. (A quantizer effectively lumps all the density under w(x) from $(nq-(q/2)$ to $nq+(q/2))$ into a single impulse at nq.)

To complete the analogy of quantization as a sampling process it is necessary to devise a function whose value at the sample points nq is the area under w(x) in the surrounding interval of width q. Figure 2.8b suggests that this function can be obtained by convolving w(x) with a

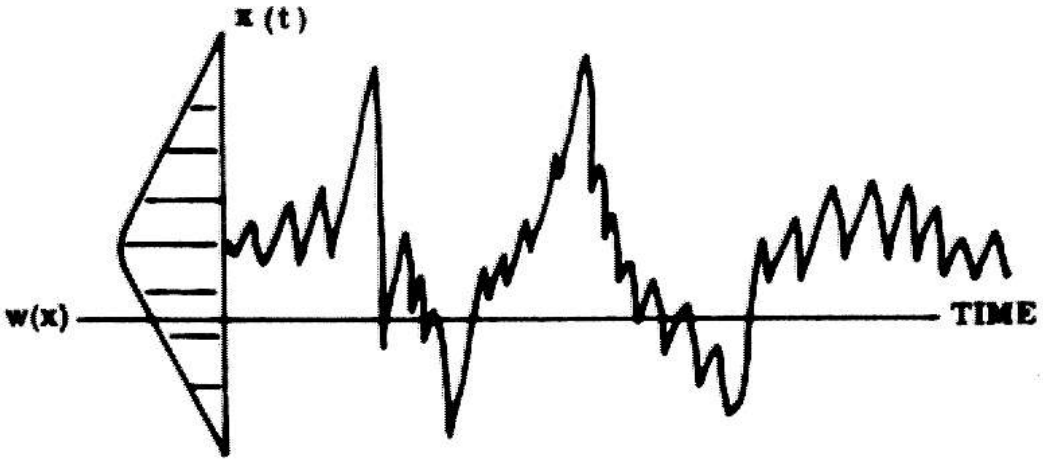


Fig. 2.7 Probability Density of a Random Signal

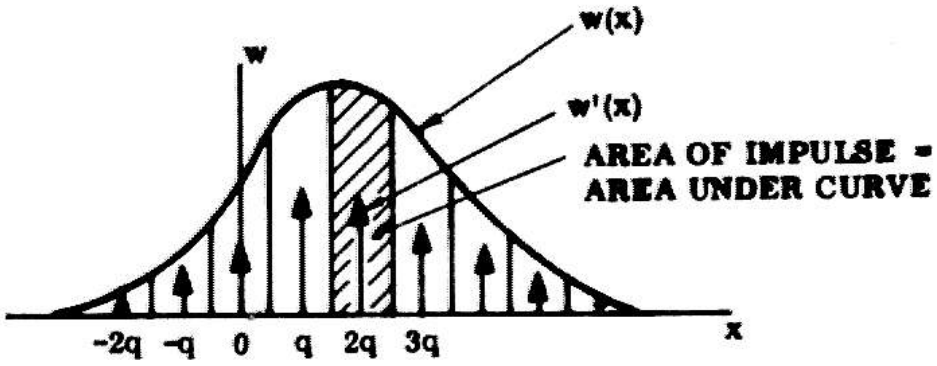


Fig. 2.8a **Sampling the Probability Density Function**

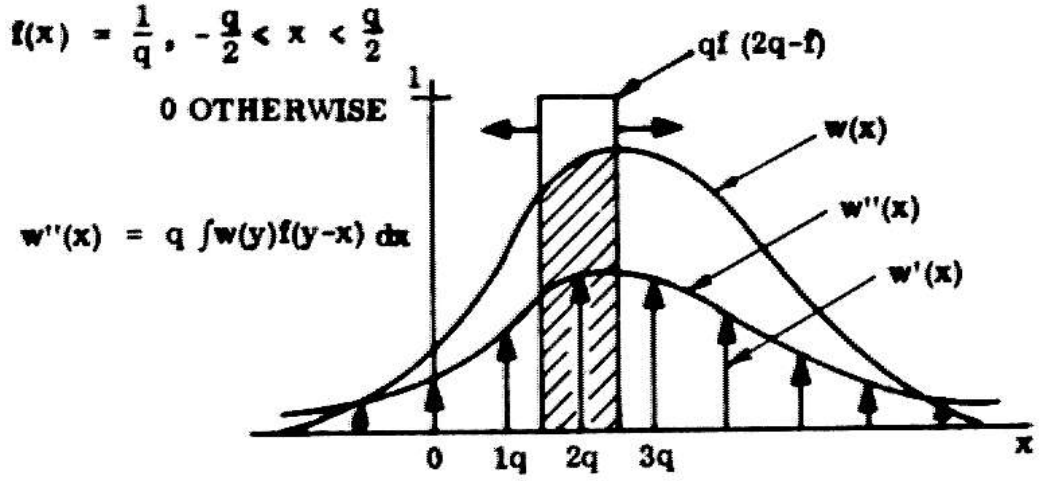


Fig. 2.8b **w'(x) From w(x) by "Fuzz and Sample"**

rectangular pulse, qf(x). Then the resulting "fuzzed" function, w"(x), can be sampled at integral intervals nq, n = 0,1,2,..., to obtain the quantized distribution w'(x).

With this model, one can determine the maximum quantization interval which will allow adequate recovery of the desired signal statistics. Sampling a function in the time domain with a periodic impulse train is known to generate a periodically repeated spectrum of the sampled function in the frequency domain. If the sampling rate is at least twice the highest frequency found in the (band-limited) signal, then the periodic spectra will not overlap. The exact spectrum of the original signal can then be isolated by low pass filtering. If the signal is not band-limited, or the sampling rate not frequent enough, then aliasing of the spectrum will compromise recovery of the original signal.

In quantization, the amplitude axis (of the density function) rather than the time axis is sampled. We are thus concerned with isolating the spectrum of:

$$w'''(x) = w(x) \circledast f(x) \quad . \quad \circledast \equiv \text{convolution} \qquad (2.30)$$

From the periodically repeated spectral kernal

$$W'''(\alpha) = W(\alpha) \times F(\alpha) \qquad (2.31)$$

(Note that the factor of q, introduced in Figure 2.8b, is cancelled by the spectrum of the periodic sampling pulses with which 2.31 is convolved.)

The analysis is completely analogous to that developed for time sampling. Samples, spaced by q in the amplitude domain, will have separations of $2\pi/q$ in the frequency domain. $F(\alpha)$ is not band-limited.

$$F(\alpha) = \frac{\sin \alpha \times \frac{q}{2}}{\frac{\alpha q}{2}} \tag{2.32}$$

Thus, $W'''(\alpha)$ can be exactly recovered providing condition 2.2^9 applies.

$W(\alpha)$ is obtained from $W''(\alpha)$ by dividing it by $F(\alpha)$. (The division can be performed since $|F(\alpha)|>0$ for $|\alpha|\leq\pi/q$.) Finally, $w(x)$ is retrieved from $W(\alpha)$ by inverse transformation.

Application of Quantizing Theorem to Gaussian Signals

In this thesis we are concerned almost exclusively with digitizing Gaussian distributed signals. The spectrum of a Gaussian waveform is also Gaussian and thus not strictly band-limited as required by the quantizer theorem. It is well-known, however, that, for practical purposes, the dynamic range of such signals is effectively 3 standard deviations on each side of the mean. We would expect to do reasonably well at recovering the first few moments, if an

adequate number of quantization intervals are spaced over this dynamic range. Ross calculated the theoretical effects of foldover from finite quantization of a normally distributed signal. He reached the intuitively satisfying conclusion that with a quantization interval one standard deviation in width, the first four moments of the quantized distribution differed by less than $3*(10^{-5})\%$ from the corresponding moments of w(x). This resolution provides only six quantization levels over the dynamic range of the distribution. Yet, finer quantization will not significantly improve the statistical characterization of this signal.

The results of Ross' work can be summarized by noting that the effects of quantization contribute two distinct modes of noise:

1. aliasing caused by undersampling and

2. the fuzzing effect of a finite quantization interval.

The first result is relevant to the selection of an appropriate quantization window. The theorem states that the width of a quantization interval (1/16 the width of the entire window) should be on the order of 1 standard deviation to avoid aliasing effects. With this quantization fineness Ross showed that the exact density, w(x), of the original time signal, x(t), could be theoretically recovered. However, an expensive deconvolution

is required to remove the effects of noise source 2, the fuzzing function. In practice, quantized signals are not subjected to this processing. Therefore, what is observed is the density function

$$w'''(x) = \int_{-\infty}^{\infty} w(y) f(x - y)\, dy \qquad (2.33)$$

rather than w(x). w'''(x) is the same density function that would be obtained if an independent source of random noise with distribution function f(x) were added to the process, described by w(x). In other words a finite quantization interval, q, has the same effect on the recovered moments as would a source of uniformly distributed noise of mean 0 and mean square (power) of $q^2/12$, summed into x(t) before the moments were calculated. (This mean square value is based on

$$\int x^2 f(x)\, dx = \frac{1}{q} \int_{-q/2}^{q/2} x^2\, dx = \frac{q^2}{12} \qquad (2.34)$$

(This result is well-known to statisticians as Shepard's correction to the 2nd moment calculated from grouped data. It is the error that will be observed by calculating the 2nd moments from W'''(ω)=W(ω)•F(ω) rather than first dividing to remove the effects of the fuzzing function.)

Implications for Noise Reduction by Averaging.

The model of a quantizer as a source of uniform random noise is only valid, when the basic quantization interval is small enough to eliminate aliasing. This constraint has important implications for accommodative strategies. Temporal averaging of the quantized signal will reduce the combination of quantization and source noise (described by w'''(x)) by 1/sqrt(N). If the quantization interval does not satisfy the aliasing constraint, however, the effectiveness of this averaging must deteriorate. In the limit when the interval width becomes greater than the maximum signal deviation, no improvement by averaging is possible, since all samples will be quantized to the same value. (The theoretical relation between averaging effectiveness and quantization width warrants further investigation. Ross' model seems like a promising approach to use.)

Averaging cannot correct the loss of detail, caused by under-sampling. Aliasing uncertainties can be removed only by using finer quantizer resolutions. Narrowing the window beyond the requirements of the quantization theorem will reduce the fuzzing uncertainty (Equation 2.34). However, unlike averaging, it will not refine the basic uncertainty in the source.

II.3.2 OPTICAL CHARACTERISTICS

In this section we briefly derive some well-known results from photographic optics that are necessary to understand the optical accommodations of our camera. Readers wishing more detail are directed to any standard photographic text (eg. Larmore [1965], the primary reference for the discussions that follow).

We will describe the imaging properties of an elementary camera, consisting of a simple convex lens. In actuality, our camera's lens contains two elements, but the results of our analysis depend only on parameters that can be modeled by an equivalent lens.

Figure 2.9 illustrates a single lens camera. The region from object to lens will be known as object space and x, the object distance. Correspondingly, y is the image distance in image space. x and y are related by the focal length, according to the well-known Gaussian lens equation

$$\frac{1}{x} + \frac{1}{y} = \frac{1}{f} \qquad (2.35)$$

(This equation can be derived directly from the geometric optics of a simple lens: rays through the lens center are not bent, while rays parallel to the principal axis are bent to pass through the focal point. A derivation in which the lens is modeled by introducing phase delay into the wave

front is given in Goodman [1968].)

The light gathering power of a lens is characterized by its focal ratio or f#, defined as

$$f_{\#} = \frac{f}{d} \tag{2.36}$$

The lateral magnification, M, is given by

$$m = \frac{h_1}{h_2} \tag{2.37}$$

From similar triangles CAB and CA'B' in Figure 2.9, M can be expressed as

$$m = \frac{y}{x} \sim \frac{f}{x} \tag{2.38}$$

II.3.2.1 EFFECT OF OPTICS ON IMAGE BRIGHTNESS

In a strict sense brightness is a subjective term, applied to an object's appearance, as seen by the eye. For quantitative measurements we define

1. luminous intensity as the total light emitted by a source,

2. luminance as the light emitted or reflected per unit area, and

3. illuminance as the light/unit area incident on a

surface. To illustrate the use of these terms and introduce more standard definitions imagine a point source at the center of a 1 foot sphere. The luminous intensity of the source is S candle power. It emits light flux that propagates through space. The unit of flux is the lumen. 4π lumens=1 candle power. The illuminance (light/unit area), falling on the inside surface of the sphere, is $I=S/4\pi(R^2)$. This equation is the familiar inverse square law. If S=1 candle power, then I=1 foot candle. Expressed in terms of the luminous flux density, this illuminance is equivalent to 1 lumen/square foot.

II.3.2.1.1 PHOTOMETRY OF A BROAD SOURCE

It is seldom necessary to look at a point source. The light flux (in lumens), reflected from a surface S per unit area, is known as the luminance E_s of that surface. In order to establish image brightness in terms of the brightness of an object, we temporarily make three simplifying assumptions (Figure 2.10a):

1. The object is distant and will be imaged in the focal plane of the lens.

2. The object is perpendicular to the axis.

3. The surface is a diffuse reflector.

First, the illuminance at the lens will be found. If A_s is the area of S, $E_s \cdot A_s$ will be the total emitted flux

in lumens. If an imaginary hemisphere were constructed about this source (Figure 2.10b), then all of the flux emitted by S must be intercepted by the hemisphere. Let the intensity (ie, flux density) at the center of the hemisphere (the position of the lens) be I₁.

A diffuse reflector appears equally bright from all directions. Lambert's Law states that since the projected area of the source decreases as the cosine of the angle between the viewing direction and the perpendicular, the flux density must also decrease (as cos (theta)) to maintain the appearance of constant illuminance. Thus at any off axis position on the hemisphere $I = I_1 \cos\theta$. I_1 is defined by the requirement that the total surface integral of I equals $E_S \cdot A_S$. The integral is straightforward under the assumption that the linear dimensions of the source are $\ll x$,

$$E_S A_S = 2\pi I_1 \int_0^{\pi/2} x^2 \sin\theta \cos\theta \, d\theta = \pi I_1 x^2 \qquad (2.39)$$

Therefore, the illuminance at the lens is

$$I_1 = \frac{E_S A_S}{\pi x^2} \qquad (2.40)$$

If the diameter of the lens D is also $\ll x$, the flux through the lens will simply be

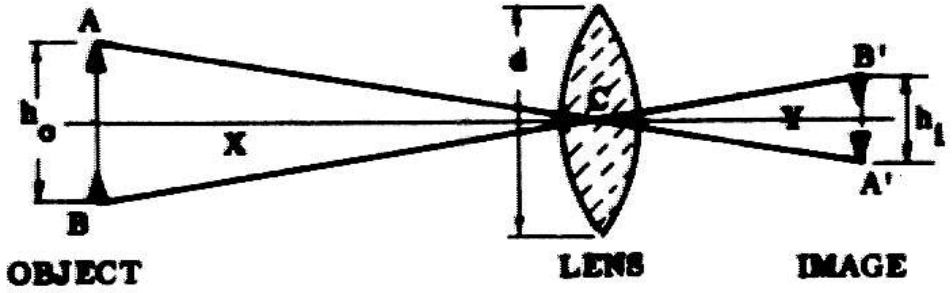


Fig. 2.9 Single Lens Camera

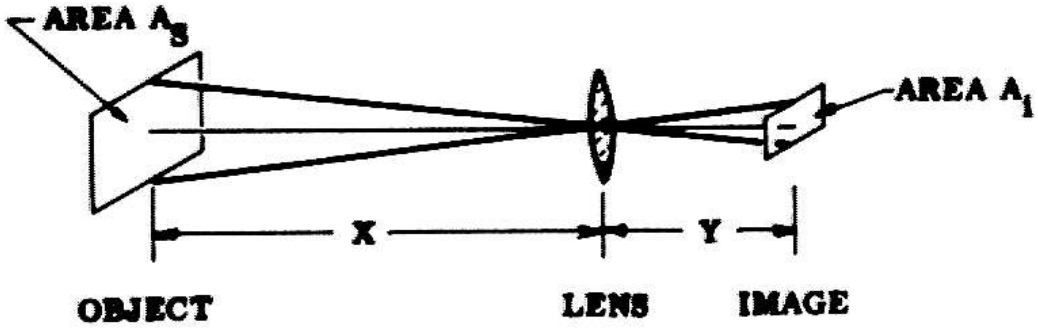


Fig. 2.10a Image Brightness of an Extended Source

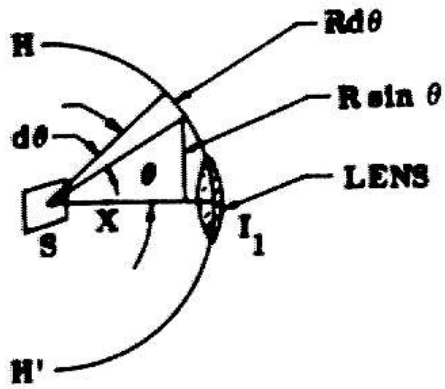


Fig. 2.10b Light Collected by Lens

$$F = \frac{I_1 \pi D^2}{4} \tag{2.41}$$

When all of this flux is focused onto the image of S, the illuminance of the image is

$$I_1 = \frac{F}{Ai} = \frac{I_1 \pi D^2}{4Ai} = \frac{E_S A_S D^2}{4 Ai x^2} \tag{2.42}$$

where Ai is the area of the image, Ai is related to As by the square of the linear magnification ratio,

$$Ai = m^2 \times A_S = \frac{f^2}{x^2} A_S \tag{2.43}$$

where f is the focal length of the lens. Substituting into 2.42 yields the desired image illuminance

$$I_1 = \frac{E_S D^2}{4f^2} \tag{2.44}$$

Recall that the focal ratio (f#) was defined as f/d. Thus,

$$I_1 = \frac{E_S}{4f_\#^2} \tag{2.45}$$

The most significant results of the above relationships for our work are:

1. The image brightness depends on the source brightness and the lens f#, but not on the object distance

x.

2. Brightness at constant f# is independent of magnification.

In concluding this section, it is appropriate to relax two of the constraints involved in this derivation so that the results will be more applicable in practice:

1. Off axis sources: S could be placed off axis (We will do the analysis for one dimension.) at an arbitrary angle to the lens (see Figure 2.11). Assuming I+1 is the flux density at distance P, perpendicular to S, then by Lambert's Law the flux, approaching the lens at angle phi from this perpendicular, is simply

$$I_2 = I_1 \cos\phi = \frac{E_S A_S}{\pi p^2} \cos\phi \qquad (2.46)$$

The projected area of the lens along this ray is

$$A'_L = \frac{\pi D^2}{4} \cos\theta \qquad (2.47)$$

Hence the total flux, absorbed to form the image is (from Equations 2.46 and 2.47)

$$F = I_2 A'_L = \frac{\pi D^2}{4} \frac{E_S A_S}{\pi p^2} \cos\phi \cos\theta \qquad (2.48)$$

Again all of this flux is focused onto the image of the projected area of the source, A+s cos (phi). The

magnification of the lens in terms of p will be

$$m = \frac{f}{x} = \frac{f}{p \cos \theta} \tag{2.49}$$

therefore

$$A_i = \left(\frac{f}{p \cos \theta}\right)^2 A_S \cos \phi \tag{2.50}$$

$$I_i = \frac{F}{A_i} = \frac{\dfrac{D^2 E_S A_S \cos \phi \cos \theta}{4p^2}}{\dfrac{f^2}{p^2 \cos^2 \theta} A_S \cos \phi} \tag{2.51a}$$

$$I_i = \frac{E_S D^2}{4f^2} \cos^3 \theta \tag{2.51b}$$

Note that I↓i is independent of phi. Thus, in the case of an edge of a cube (see Figure 2.12) the differences in brightness of the two adjacent faces will only be a function of their reflectances and the incident illumination. In the common case where the cube is a homogeneous body, the edge will be manifest entirely by differences in the directional nature of the illumination field.

2. Extended sources. Equation 2.40 which defined I↓i depended on the fact that the object's dimensions were <<x. Optical systems, however, are linear. Consequently,

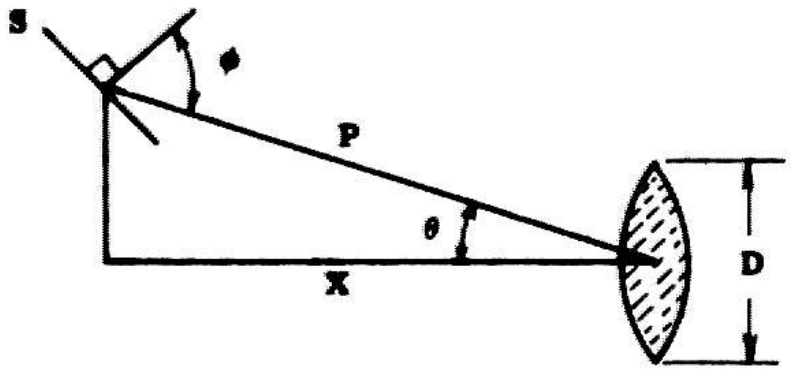


Fig. 2.11 Brightness of an Off-Axis Source

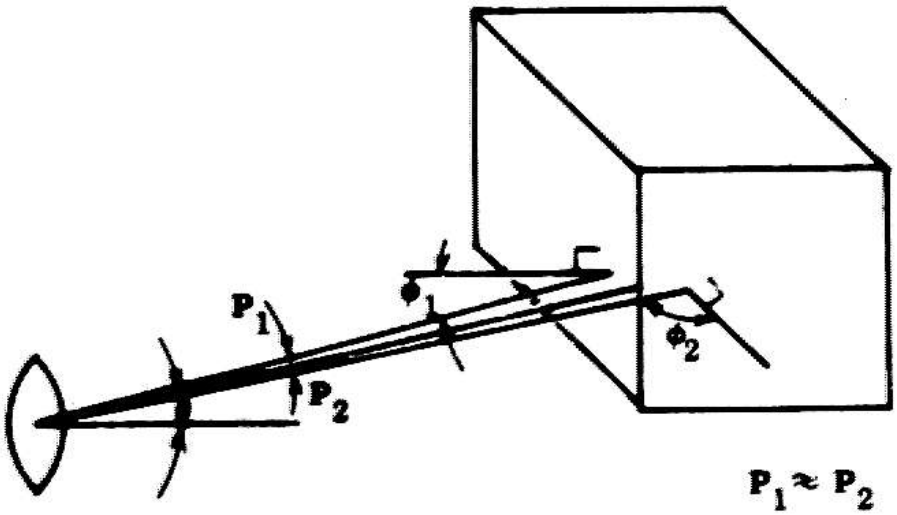


Fig. 2.12 Contrast of Interior Edge of a Cube

extended images can be analyzed, at each point, using the principle of superposition and the results derived above.

II.3.2.2 EFFECTS OF OPTICS ON IMAGE SHARPNESS

Limiting Spatial Resolution

In optical instruments, the main limitations on spatial resolution result from diffraction and lens aberrations. However, the resolution in a digitized television image is limited primarily by the raster sampling rate and the bandwidth of the video amplifier. In our system, each horizontal scan line is sampled 333 times. (This rate is consistent with the 10 mc. bandwidth of the video amplifier.) The horizontal scanned area is 1/2". Thus, the size of the basic resolution cell on the face of the vidicon is .5/333=1.6*10↑(-3)". Because diffraction effects are insignificant on this scale, all subsequent analysis will be based on the simplified models of geometric optics.

II.3.2.2.1 FOCUS

Depth of field

The depth of field is the distance between the nearest and farthest points on the optic axis for which all objects are in "satisfactory" focus. A workable

definition of "satisfactory" focus can be based on the geometric model illustrated in Figure 2,13. A point source of light, out of focus, is imaged as a circular disk, (known in photography as a circle of confusion). The diameter of this disk is a measure of unfocus. In particular, the image of a point can be taken to be in focus, when the diameter of its circle of confusion falls within a single resolution cell on the image plane.

In Figure 2,13, a point source on the axis between x+1 and x+2 will be imaged between y+1 and y+2. In this range, the size of the disk, defined by the rays in the image plane, will be less than the diameter c. Let c be the diameter of a resolution cell, ie. the maximum allowable circle of confusion. Then x+1 and x+2 define the near and far field limits of focus for image distance y.

x+1 and x+2 can be expressed as functions of x (the nominal depth we are focused for), f(the focal length of the lens), d(the diameter of the lens), and c. The derivation is based on simple geometric arguments. From the similarity of triangles in the image space

$$\frac{c}{d} = \frac{y - y_1}{y_1} = \frac{y_2 - y}{y_2} \qquad (2,52)$$

We use the lens formula to transform each of the y's into a corresponding x in object space. Substituting in Equation 2,52 and simplifying yields

$$\frac{c}{d} = \frac{x_1 f - xf}{x_1 x - x_1 f} = \frac{xf - x_2 f}{x_2 x - x_2 f} \qquad (2.53)$$

These equations can then be solved for x+1 and x+2:

$$x_1 = \frac{xdf}{df - c(x - f)} \qquad (2.54a)$$

$$x_2 = \frac{xdf}{df + c(x - f)} \qquad (2.54b)$$

The desired results follow directly

$$\text{Far Depth} = D_1 = x_1 - x = \frac{cx(x - f)}{df - c(x - f)} \qquad (2.55a)$$

$$\text{Near Depth} = D_2 = x - x_2 = \frac{cx(x - f)}{df + c(x - f)} \qquad (2.55b)$$

$$\text{Depth of Field} = D_1 + D_2 = \frac{2xdfc(x - f)}{d^2 f^2 - c^2 (x - f)^2} \qquad (2.55c)$$

The significance of these equations for our purposes are:

1. The depth of field decreases as d increases. The effect of a bigger lens aperture is to widen angles a-0+1'-b and a-0+2'-b in Figure 2.13. Consequently, the size of the circle of confusion, corresponding to a point source a given

distance from the focused range, will be increased.

2. The depth of field will also decrease as the focal length of the lens is increased. This fact is most easily seen by rewriting the lens equation in the form

$$(x - f)(y - f) = f^2 \qquad (2.56)$$

This equation has been plotted in Figure 2.14 in a normalized form by labeling both axes in units of f. Consider the image distances y+1 and y+2 corresponding to an arbitrary interval of depth in object space, for example, ∞ to 2f. With a 1" lens (eg. f=1) these depths lie on the horizontal asymptote of the figure. Thus, only an insignificant difference exists between the y-coordinates of their respective best image planes. This situation typifies a broad depth of field. Now consider the effect of a 4" lens focused at these two ranges. Utilizing the normalized axes, 10" and 20", now plot at 2.5 and 5 respectively. This produces a much greater disparity in the two focal distances (remember that the y axis is also scaled by f). As a result, it is much harder for both points to be in focus at the same time.

3. Depth of field varies inversely with range. This fact can also be demonstrated using Figure 2.14. Decreasing the image distance moves the operating points on the horizontal axis towards the sharply sloped region of the

curve.

Depth of Focus

Depth of focus refers to the allowable tolerance in image distance over which the image of an object remains in acceptable focus. In Figure 2.15, distance D represents the depth of focus for an allowable circle of confusion diameter c. From the figure, we see that this interval is symmetrically placed about the point of best focus. By similar triangles

$$\frac{c}{\frac{1}{2}D} = \frac{d}{y} \qquad (2.57a)$$

for $x \gg f$, $y \sim f$, therefore

$$D = \frac{2 \times f \times c}{d} = 2 \times c \times f_{\#} \qquad (2.57b)$$

We see from Equation 2.57b that depth of focus is directly proportional to f# (eg. f/d defines the shape of the focus triangle in Figure 2.15).

Depth of focus measures how sharply the best image plane is defined in image space and is thus a useful criteria for bounding the precision of automatic focus algorithms. In a related application the chosen point of best focus can be used with the lens equation to provide a crude estimate of depth. In this context it is important to

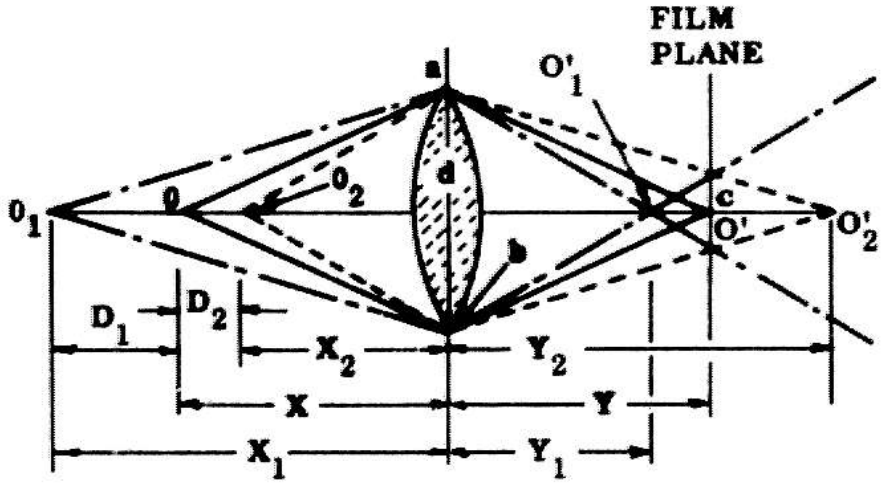


Fig. 2.13 Diagram Illustrating the Depth of Field

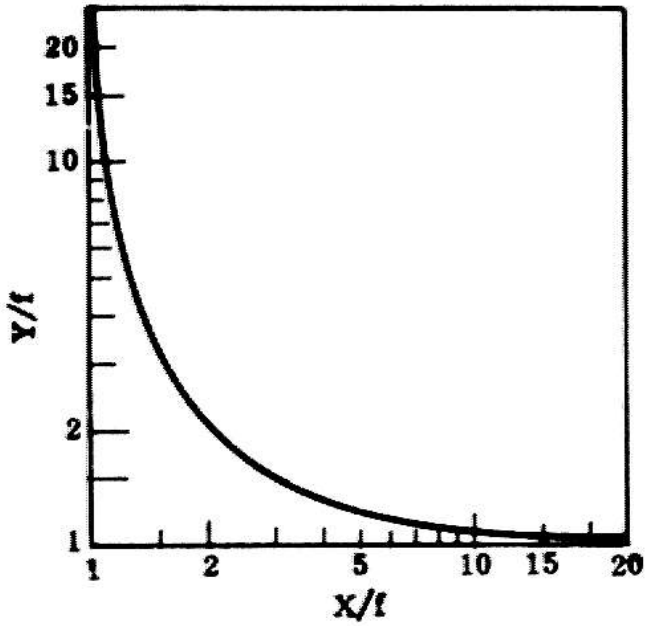


Fig. 2.14 Lens Equation: $(X-f)(Y-f) = f^2$

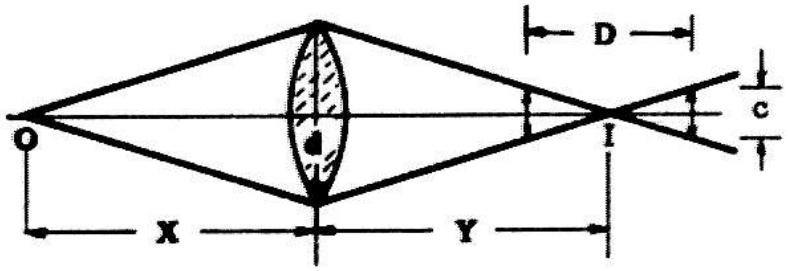


Fig. 2.15 Diagram Illustrating Depth of Focus

relate the uncertainty that D represents in the image plane to an equivalent uncertainty in object space. The lens equation can be used to transform the end points of interval D into object space. For d>>c, there is an algebraic identity between this reflected interval and the depth of field predicted by Equation 2.55c for the same f# and image plane (y).

Field of View

The field of view of a television camera is the volume of space enclosed by the rectangular cone shown in Figure 3.2, truncated to the selected depth of field. The cone is defined by projecting the image plane through the lens. At each object distance the inverse magnification function of the lens is

$$m^{-1} = \frac{x}{y} = \frac{x - f}{f} \qquad (2.58)$$

This factor, applied to the dimensions of the image plane, will determine the width and height of the scene that will be in view (The half angles, α and β, subtended by the projected image plane, are sometimes known as the horizontal and vertical viewing angles.). Field of view is a function of x (the range at which focused), f(focal length), and d(iris).

II,3.2.3 COLOR

A television camera responds only to the total light energy that is absorbed by its image tube. Two surfaces, easily distinguished by the eye on the basis of a color difference, may appear with the same tonal value when reproduced in black and white. This situation was illustrated in Figure 1.10a for the case of three colored blocks. The brightness, I, at which colored objects will be rendered depends intrinsically on three factors:

1. the spectral composition of the source illumination $S(\lambda)$,

2. the spectral reflectance of the object's surface, $O(\lambda)$, and

3. the spectral sensitivity of the vidicon $W(\lambda)$ (see Figure 2.16).

These factors are related by Equation 2.59.

$$I = \int_{\lambda} S(\lambda)O(\lambda)W(\lambda)\,\partial\lambda \qquad (2.59)$$

λ in practice is limited to the range $3500A^{\circ}<\lambda<7000A^{\circ}$ by the sensitivity of the camera.

Color filters modify the spectral composition of the light passing through them. Let $F(\lambda)$ describe the spectral transmission of a filter. The effect of putting this filter in the optical path is given by

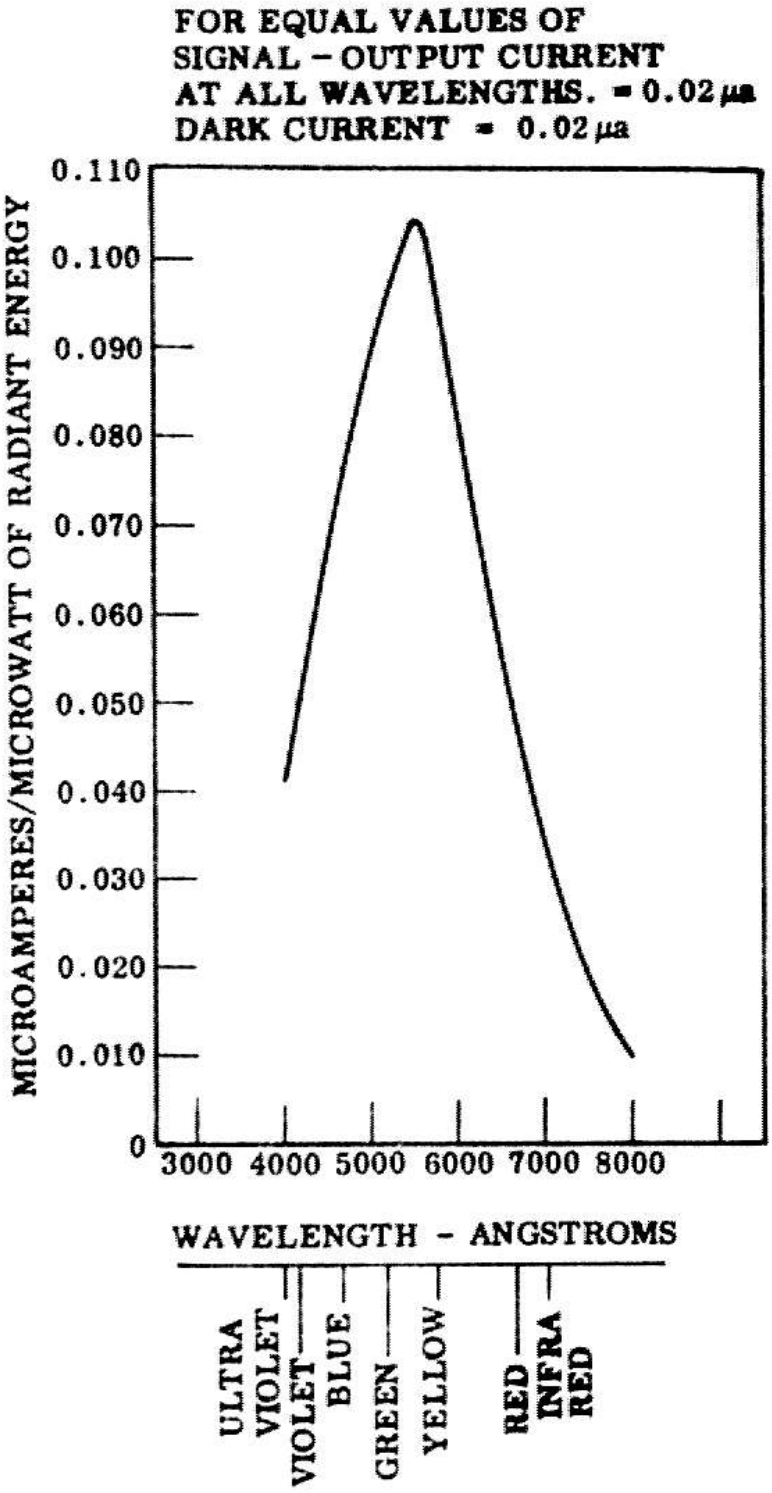


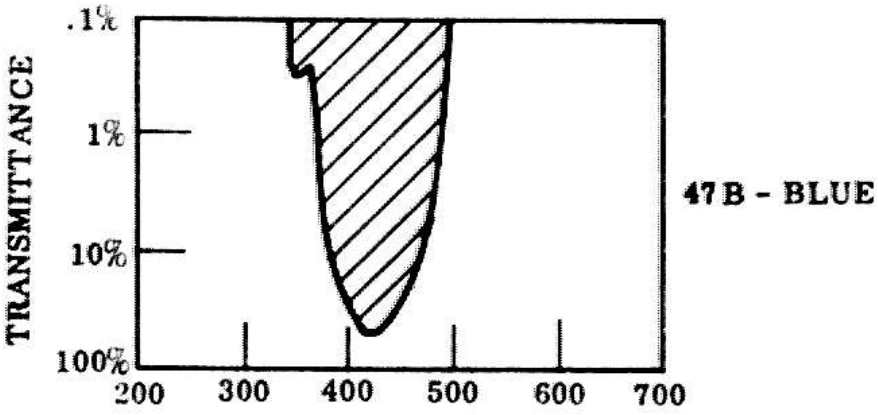
Fig. 2.16 Typical Spectral Sensitivity Characteristic of Vidicon

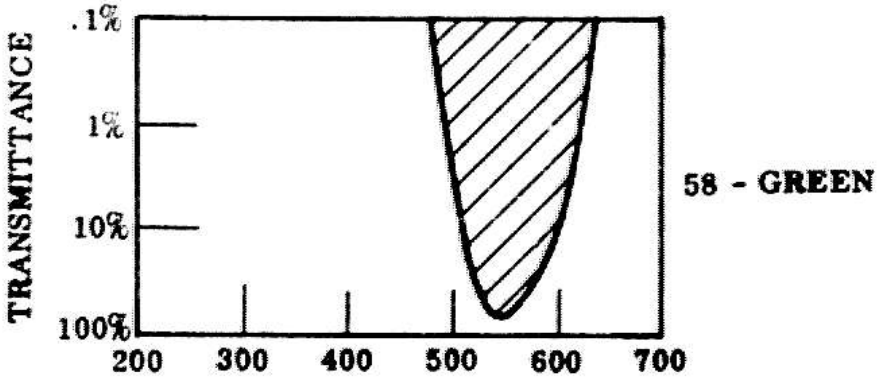
$$I = \int_{\lambda} S(\lambda)O(\lambda)F(\lambda)W(\lambda)\,\partial\lambda \qquad (2.60)$$

Color filters can be used, for example, to emphasize selected colors (as in Figure 1.10b-d) or to restore the contrast between two colors eliminated by strict black/white tonal rendering. In Chapter 7, the relative brightnesses obtained through the various filters are used to recognize color.

As a general rule, a colored specimen can be lightened, relative to the average brightness level of a scene, by using a filter of similar color (ie, a matched filter). Conversely, an object can be darkened by using a filter of a different color or blackened by using a filter of the complimentary color (see Chart 3.3).

The color wheel of our camera holds three color selective filters (in addition to a neutral path). These filters partition the complete spectrum into three broad bands centered in the red, green, and blue regions. The transmission characteristics are shown in Figure 2.17. Further details can be found in the Kodak Filter Manual [1968].





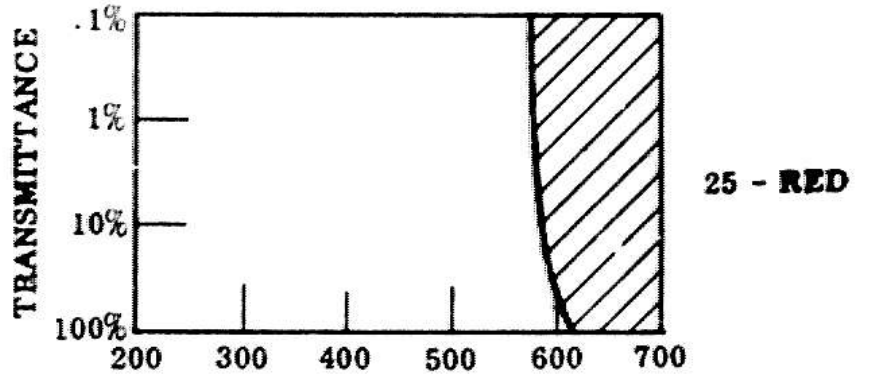


Fig. 2.17 Color Filter Characteristics

CHAPTER III: INTERPRETATION AND USE OF MODELS

The picture taking process can be parametized as in Chart 3.1. The appearance of a scene is characterized by two-dimensional patterns of properties, like those in column a. The camera transforms this scene into an image described by the characteristics in column c. Because of the physical limitations of the camera, only those characteristics of the scene, in the range selected by the accommodation parameters of the camera, will appear in the image. The camera, in effect, is an adjustable multi-dimensional window on the world.

In this section we will model the interaction of the camera accommodations (Chapter 2) with the scene characteristics to determine the composition of an image. In the following section we will use these results to correlate accommodations with the information requirements of specific tasks.

III.1 SUMMARY OF SCENE CHARACTERISTICS

In this study a scene is limited to table-top environments similar to that pictured in Figure 1.1. The dimensions of the usable hand-eye work space are drawn in Figure 3.1. Typical objects have lateral dimensions in the range from 3/4" to 6". Detail, whose extent is less than

Scene Characteristics	Camera	Image Characteristics
Brightness (absolute dynamic range)	Sensitivity	Intensity Window
	Clips (quantizer)	Contrast
Color	Color Filter	Color Tonal Rendering
Depth	Iris Focus	Spatial Window field of view depth of field
Spatial Extents	Lens	Level of Textural Detail 1. Magnification 2. Spatial resolution
Texture Detail	Pan-Tilt	
Spatial Noise (e.g. dust, chipped paint)	Temporal Noise 1. shot 2. dark current 3. thermal 4. amplifier	Spatial Noise (combined results of scene and camera noise)
a	b	c

Chart 3.1

Descriptive Parametization of Imaging Process

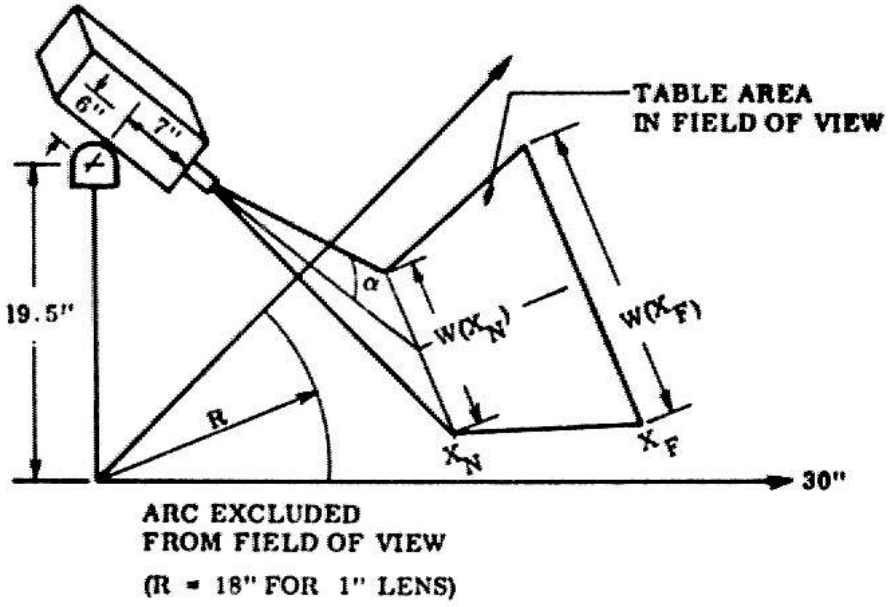
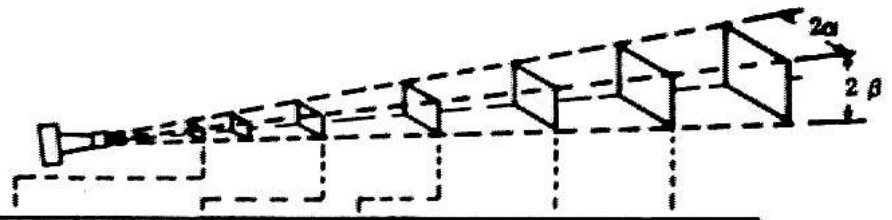


Fig. 3.1 Dimensions of Table Work Space



LENS	10 W × H	20 W × H	30 W × H	40 W × H	50 W × H
1	5 × 3.75	10 × 7.5	15 × 11.25	20 × 15	25 × 18.75
2	2.5 × 1.87	5 × 3.75	7.5 × 5.62	10 × 7.5	12.5 × 9.37
3	1.67 × 1.25	3.33 × 2.5	5 × 3.75	6.67 × 5	8.33 × 6.25

NOTES: WIDTH OF FIELD (W) = $\frac{\text{LENS TO SUBJECT DISTANCE (D)}}{2 \times \text{LENS FOCAL LENGTH (INCHES)}}$ ALL UNITS IN INCHES

HEIGHT OF FIELD (H) = 0.75 × W

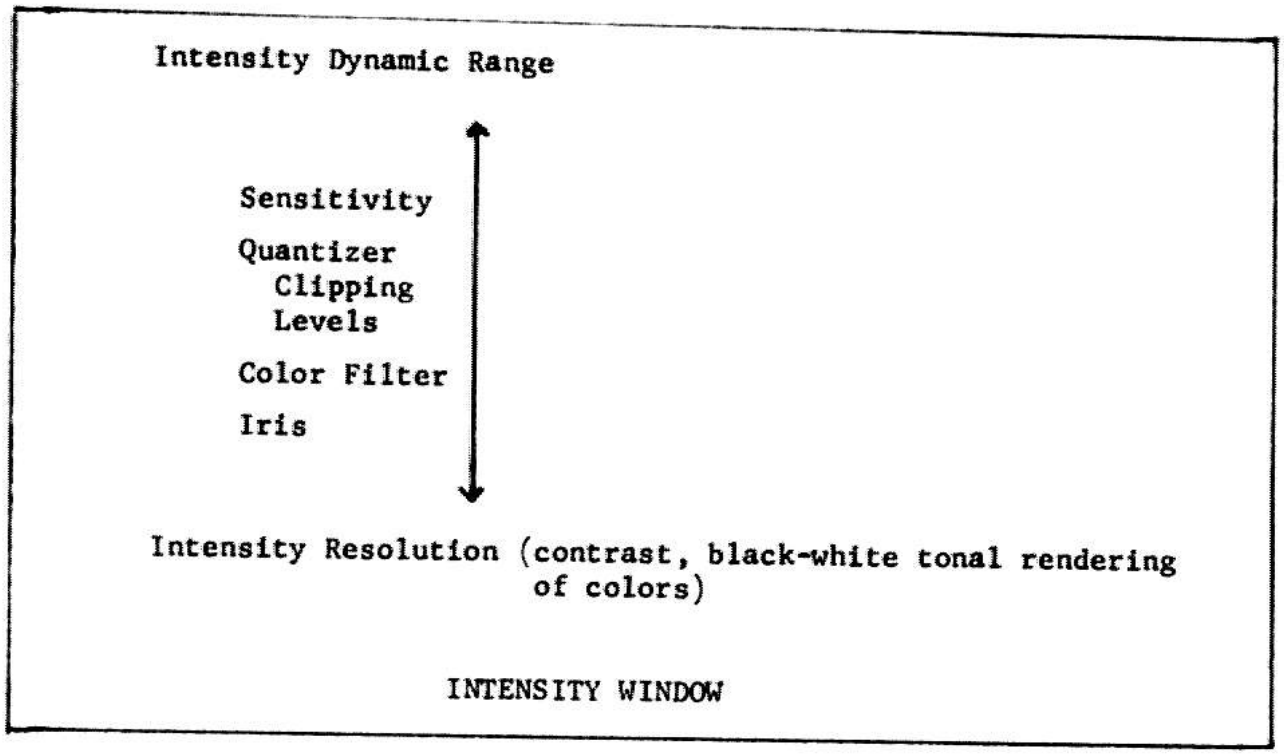
Fig. 3.2 Field of View

2/10", is classed as texture.

The luminance of objects in these scenes covers a 160:1 dynamic range (from 1 ft.c. for a black object in flourescent room lighting to 160ft.c. for a white object, directly illuminated by the 1000 watt quartz-iodine source). The range of object reflectance (lightest to darkest) is about 25:1. Hence the range of effective illumination energy is less than 8:1. (If illumination possibilities included sunlight, the overall dynamic range would be considerably extended. Furthermore, if a single illumination source were used, an additional 20:1 variation could be expected due to directionality.)

III.2 SUMMARY OF IMAGE CHARACTERISTICS

Image characteristics are summarized by the concept of a window. This concept defines the range of intensities, or depth, or color tones, etc., distinguishable in the picture. The resolution of the detail contained in each window is inversely related to the width of that window. We have already seen several examples of this basic trade-off. Chart 3.2 summarizes the two principal sets of interdependencies among image characteristics and includes the relevant accommodations which influence how each trade-off is resolved.



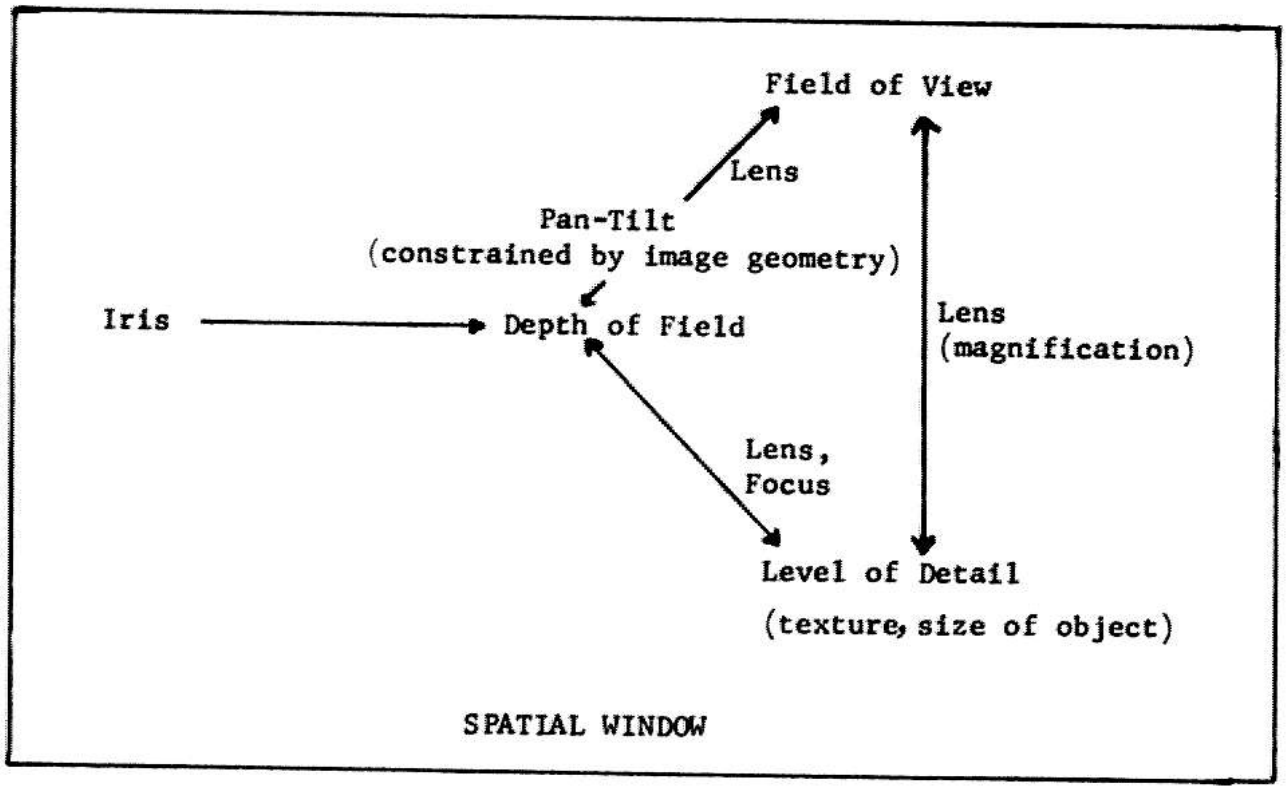


Chart 3.2

Window Characterization of Accommodation

III.3 SUMMARY OF ACCOMMODATIONS

III.3.1 INTENSITY WINDOW

The range of scene brightnesses at which contrast will be evident is selected by four adjustments:

1. sensitivity: controls the gain of the vidicon. It sets the range of absolute illuminance (on the vidicon faceplate) which will produce signals in the dynamic range of the video amplifier;

2. iris: controls the amount of light from each point in the scene that will reach the vidicon;

3. color filter: biases the spectral composition of light reaching the vidicon. This biasing affects tonal rendering and contrast in the image,

4. clips: selects a limited portion of the amplifier range,

The combined effect of the first three parameters is expressed analytically by Equation 3.1,

$$v = K\left(\frac{d^2}{4f^2}\int_{\lambda} L(\lambda)F(\lambda)W(\lambda)\,d\lambda\right)^{\gamma} \times E_T^{\alpha} \qquad (3.1)$$

d - iris diameter
f - focal length
F(λ) - filter characteristic
W(λ) - camera spectral response
γ - vidicon gain, γ = 0.65
α ~1.4

This equation constitutes a comprehensive photometric model of the camera. It expresses the overall transfer function between the spectral illuminance, $L(\lambda)$, of a scene element and the voltage at the output of the video amplifier.

Only illuminances that produce voltages in the range from .008 to .992v. can be linearly quantized. This limitation sets an outer window. The quantizer can, of course, select a much narrower window within this range.

III.3.2 SPATIAL WINDOW

The dominant accommodation that influences the spatial window is the focal length of the lens. In brief, the larger the lens is, the higher the magnification will be. Higher magnification implies a smaller field of view with more visible fine detail. The field, seen by the camera from its orientation in Figure 3.1, is bounded by the intersection of the rectangular viewing cone (Figure 3.2) and the table surface. The geometry of this intersection is a function of camera tilt. The analysis is complicated by the fact that the tilt axis does not lie on the principal ray of the lens.

Sobel [1970] has derived a parametized collineation transformation that maps any point in the image plane into a corresponding point on the table surface. This transformation, applied to the four corners of the image

plane, defines the trapezoidal area of the table-top (Figure 3.1) that will be in view for a given lens and camera angle.

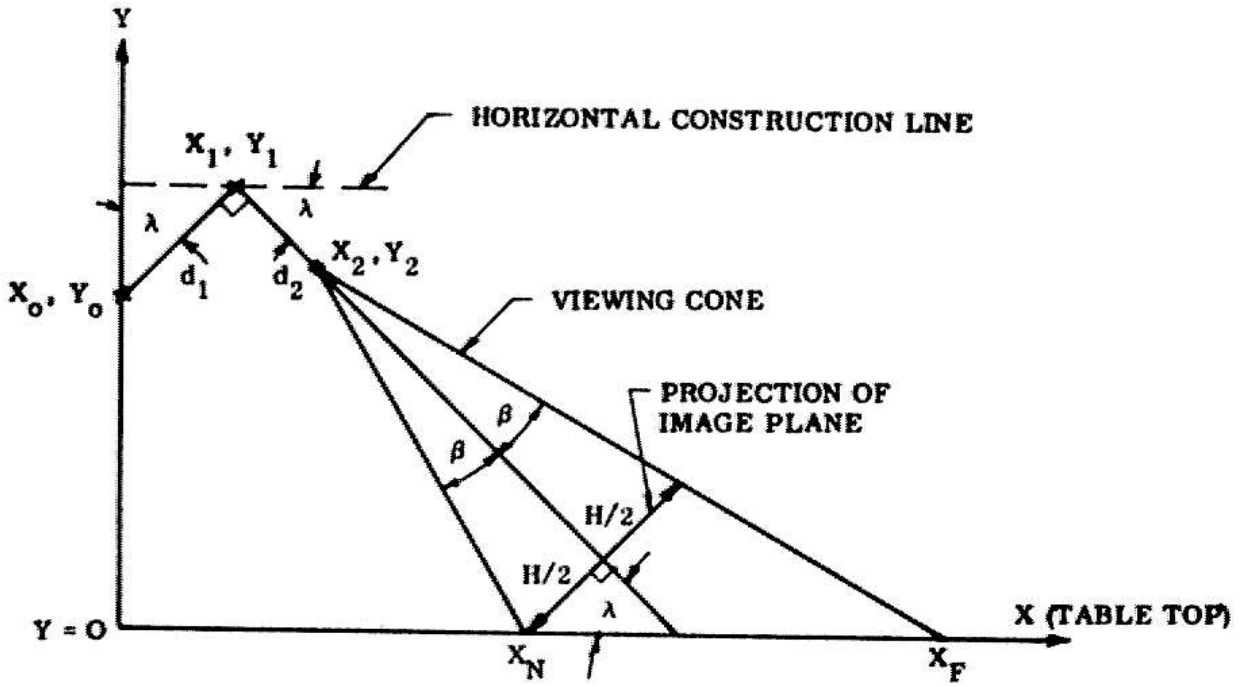
The geometry of the transformation with respect to tilt angle is illustrated in Figure 3.3. This figure is a two-dimensional, vertical, cross-sectional cut through the principal ray. It shows the near and far table limits that will be in view for any tilt angle. With simple geometry it can be shown that

$$\left.\begin{matrix} x_n \\ x_f \end{matrix}\right| = (y_0 + d_1 \cos\lambda - d_2 \sin\lambda) \tan\left(\frac{\pi}{2} - \lambda \mp \beta\right) + (d_1 \sin\lambda + d_2 \cos\lambda) \tag{3.2}$$

The derivation of these equations is sketched in Figure 3.4.

The area enclosed by the triangular region consisting of X_{in}, X_{if} and the lens center, defines a potential field of view in the same sense that the orientation of a human head and neck does. What is actually seen depends on where within this conical envelope one looks. The human can concentrate on a particular point in space by panning his eyeball to line up the appropriate ray with his fovea. He then acccommodates his lens to focus on the desired depth.

The computer has similar capabilities. The television interface provides the option of inputting an arbitrary rectangular portion of the entire field of view,



X = Y = O = BASE OF CAMERA SUPPORT

X_o, Y_o = TILT AXIS = 0, 19.5 IN.

d1 = RIGID DISPLACEMENT FROM PRINCIPAL RAY (~6 IN.)

d2 = RIGID DISPLACEMENT TO PRINCIPAL POINT (~7 IN.)

} INTERNAL CAMERA GEOMETRY

X_2, Y_2 = LENS CENTER (OF PROJECTION)

H = PROJECTED HEIGHT OF IMAGE PLAN (HEIGHT H)

λ = TILT ANGLE

X_N = NEAR TABLE INTERSECTION

X_F = FAR TABLE INTERSECTION

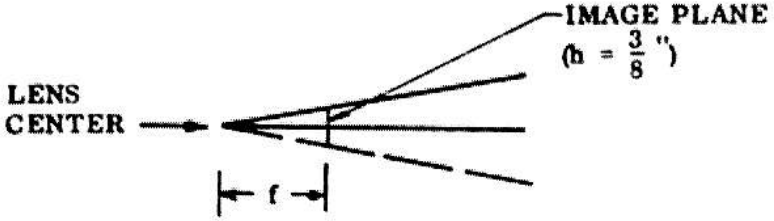
Fig. 3.3 Determining Area of Table-Top in Field of View

LENS CENTER

1. X_2, Y_2 FOUND BY SIMPLE TRANSLATION ALONG DISPLACEMENTS d_1, d_2.

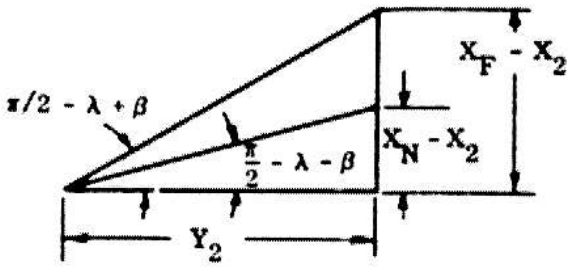
$$X_1 = X_o + d_1 \text{ SIN } \lambda$$
$$Y_1 = Y_o + d_1 \text{ COS } \lambda$$
$$X_2 = X_1 + d_2 \text{ COS } \lambda = X_o + d_1 \text{ SIN } \lambda + d_2 \text{ COS } \lambda$$
$$Y_2 = Y_1 - d_2 \text{ SIN } \lambda = Y_o + d_1 \text{ COS } \lambda - d_2 \text{ SIN } \lambda$$

2. GEOMETRY FOR DETERMINING VERTICAL VIEWING ANGLE β



$$\text{TAN } \beta = \frac{h}{2f} : \beta \approx 10'36' \text{ FOR 1" LENS}$$

3. GEOMETRY FOR FINDING X_N, X_F



$$\text{TAN } (\tfrac{\pi}{2} - \lambda - \beta) = \frac{X_N - X_2}{Y_2}$$

$$\text{TAN } (\tfrac{\pi}{2} - \lambda + \beta) = \frac{X_F - X_2}{Y_2}$$

4. SUBSTITUTE RESULTS FROM 1, 2, INTO 3.

$$X_N = Y_2 \text{ TAN } (\tfrac{\pi}{2} - \lambda - \beta) + X_2 =$$
$$X_N = (Y_o + d_1 \text{ COS } \lambda - d_2 \text{ SIN } \lambda) \text{ TAN } (\tfrac{\pi}{2} - \lambda - \beta) + d_1 \text{ SIN } \lambda + d_2 \text{ COS } \lambda$$
$$X_F = (Y_o + d_1 \text{ COS } \lambda - d_2 \text{ SIN } \lambda) \text{ TAN } (\tfrac{\pi}{2} - \lambda + \beta) + d_1 \text{ SIN } \lambda + d_2 \text{ COS } \lambda$$

Fig. 3.4 Sketch of Derivation of X_N and X_F

This situation, as Sobel points out, can be viewed as an analogy of eye rotation. The vidicon is then focused on a specific depth. The focal length and iris together determine a depth window, δz, about the point of best focus, according to Equation 2.55c.

The intersection of this depth window with the viewing cone determined by the lens geometry forms a composite spatial window illustrated in Figure 3.5. (Range is customarily measured along the principal ray. The viewing angle, 2β, is small enough so that the entire area of the plane it subtends at each depth can be assumed to be in focus when the center point is in focus. This assumption is known in optics as the paraxial approximation.)

III.3.3 LEVEL OF DETAIL

Accommodation, as we have seen, provides the option of obtaining coarse detail over a wide window or fine detail over a narrow window. The finest detail that can be seen in a digitized image depends on four factors:

1. spatial sampling density,
2. lens magnification,
3. sharpness of focus, and
4. detail contrast.

High magnification, sharp focus, and high contrast

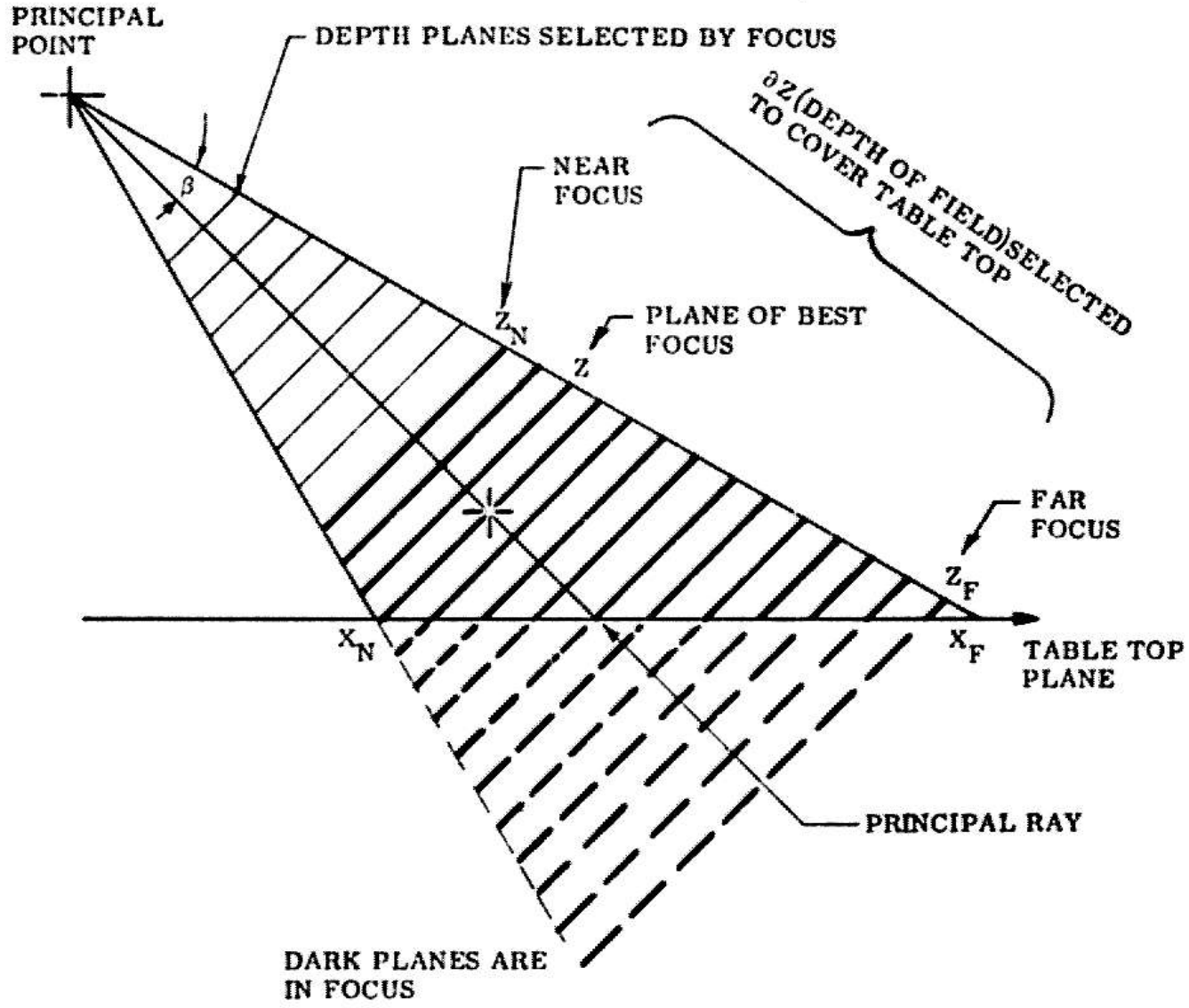


Fig. 3.5 Depth Dimension of Spatial Window

all lower the minimum detail size necessary for threshold visibility at a given sampling density. Quantitative relationships are developed in Chapter 6, in terms of the limiting amplitude and spatial resolution of the camera.

The relation of focus and contrast to detail size was illustrated in Chapter 1. By excluding the table-top from the depth window, the contrast between the unwanted grid lines and the table-top was reduced.

III.4 APPLICATION OF MODELS TO PERCEPTUAL TASKS

We now have a quantitative and a descriptive understanding of how each accommodation influences the appearance of an image.

The extent to which the analytic results can be used depends on how precisely one can define the concept of an appropriate image. This ability will vary with the task. In the following two sections examples of qualitative (heuristic) and quantitative use of the accommodation models will be considered.

III.4.1 HEURISTIC APPLICATION OF ACCOMMODATION MODELS

In Chapter 1, we discussed a general perceptual strategy for utilizing this basic capability. Gross features are initially sought over a wide field of view. If more

detail is needed to interpret this coarse structure, it can be obtained by focusing attention on the specific part of the scene where the additional information was needed. In effect, the more a vision strategy knows about a scene, the better it can anticipate what else will be found. Furthermore, these expectations can be utilized to narrow the accommodation window in order to maximize the likelihood of finding a specific item of interest.

For many tasks we may heuristically "know" how to tune the camera to obtain an appropriate image. Yet we may be unable to express these criteria analytically. It is clear, for example, that an appropriate image for initial reconnaissance should provide coverage of a large area of the table and coverage of a wide range of brightnesses. These criteria seem appropriate, because they facilitate a thorough search. Furthermore, the low resolution which accompanies wide windows tends to minimize the chance of seeing small, unimportant disruptions of the background (eg. dust specks). If these were found, they would require some effort to reject. The desired image characteristics correspond to wide spatial and intensity windows, obtained with the following accommodations:

1. quantization interval set as wide as necessary to encompass the scene's dynamic range,

2. sensitivity set according to average scene brightness to minimize the number of clipped intensity

samples,

3. short lens to maximize field of view,

4. iris set small enough so that depth of field encompasses work space (but no smaller to avoid loss of signal/noise), and

5. clear color filter to eliminate spectral bias.

The accommodations listed above are based on a qualitative interpretation of the analytic results developed in Chapter 2. When the criteria of a desirable image are expressed in an imprecise manner, a qualitative interpretation is the best use that can be made of these results.

Heuristic criteria also implied the use of selective windows, when a specific item was sought. The effectiveness of an appropriate color filter was demonstrated in Figures 1.10a-d. In this example, the choice of filter and clipping level was based on a qualitative understanding of the influence of these accommodations.

This understanding was built into the selective attention program. Given two sets of colors to be discriminated, the program simply consulted Chart 3.3 to see if any of its color filters provided an appropriate segmentation. If so, the clips were set to exclude intensities above or below the scene average, depending on whether the desired color was dim or bright through the selected filter.

Filter	Bright Colors	Dim Colors
Red	red yellow white	green blue black
Green	yellow green white	red blue black
Blue	blue white	red green black yellow

Chart 3.3

Color Segmentation

This program is heuristic in the sense that, while it generally increases the likelihood of finding a red cube, the increase cannot be guaranteed. There exists no analytic criterion for measuring the optimality of the resulting image. If a red cube were in shadow, for example, it would probably get clipped into the black background. Furthermore, we do not yet know what image characteristics minimize search time or maximize decision reliability.

At the present time our knowledge of the accommodation requirements of many perceptual functions is limited to qualitative criteria, like those used for initial acquisition. (In almost every task, the role of at least one accommodation is only understood at this level.) For these cases, the quantitative results of Chapter 2 have been interpreted for specific tasks and built rigidly into the programs. Given a task, these programs then compare it to a pre-specified list of prototypes. The best match is used to select the appropriate accommodation from a table, like Chart 3.3.

As in all artificial intelligence tasks, one hopes to eventually have the program learn better and more complete tables. One step in this direction is to maintain statistics of the relative effectiveness of alternative strategies in various situations. For example, two filters can sometimes accomplish the same color dichotomization. (Red and green can be segmented using

either a red or green filter.) One of these options may prove more effective, depending upon the task and the attendant circumstances. For acquisition, the color of the background was an important consideration. Chart 3.3 could be refined by experience to include this dimension.

III,4,2 ANALYTIC USE OF ACCOMMODATION MODELS

We now consider examples of perceptual functions for which some of the measures of what constitutes an appropriate image can be formalized in terms of our analytic accommodation models. In these cases, it is possible to obtain the optimally appropriate image. It is expected that, in time, most perceptual functions will be understood well enough to apply this type of analysis.

Detection Criteria

A fundamental performance measure of a visual system is the combination of contrast and size, needed to detect the presence of a homogeneous pattern, superimposed upon a uniformly illuminated background. Light measurement is a statistical process. Analytic techniques, borrowed from statistical detection theory, can be used to examine how parameters of the vision system influence the threshold of visibility.

We will first formulate the analytic approach in

terms of an ideal statistical visual detector. The analysis will then be applied to study the effects of accommodation on the performance of a real camera.

III.4.2.1 ANALYSIS OF AN IDEAL DETECTOR

Consider an ideal statistical detector that counts all light quanta collected by a simple optical system (eg. Figure 2.9), and produces a current proportional to brightness. Assume that the background light flux collected from a uniformly illuminated area is Poisson-distributed with mean N(quanta/exposure period) and standard deviation sqrt N. These fluctuations limit the minimal incremental increase ∂N, that can be detected in a given time with a pre-determined statistical confidence.

Figure 3.6a helps visualize the detection problem. It is obvious from this figure that the primary limitation on visibility is the size of ∂N (the signal) relative to the noise deviation $N^{+1/2}$. ∂N must be larger than sqrt (N) for reliable detection. The threshold contrast is given by

$$C = \frac{\Delta N}{N} = K_1 N^{-1/2} \qquad (3.3)$$

K^{+1} is a confidence factor that compares the minimum required signal to the size of the normal deviation range.

If the collected light is imaged onto a square

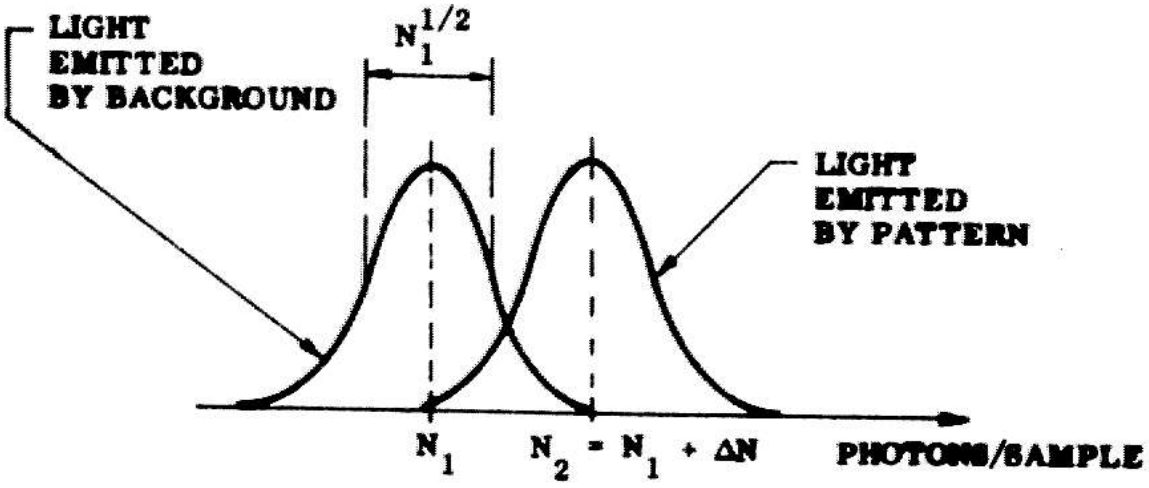


Fig. 3.6a Statistical Detection Model

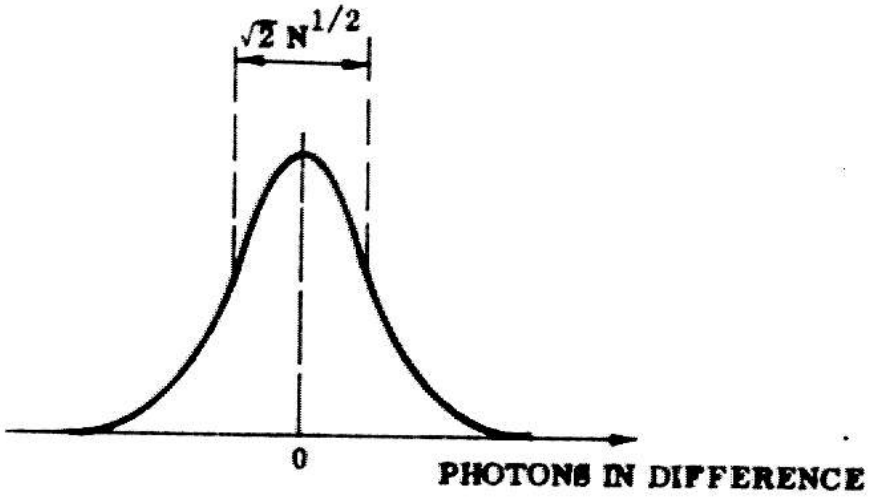


Fig. 3.6b T-Distribution of Difference of Photons in 2 Samples Drawn From Same Gaussian Distribution

summation area, h*h, the N quanta constitute an image illuminance of

$$B_1 = \frac{N}{h^2} \sim B \qquad (3,4)$$

which, in turn, is proportional to the scene illuminance B. Rewriting Equation 3.3 in terms of brightness yields a condition which must be satisfied to reliably discriminate the presence of a pattern.

$$BC^2h^2 \geq K_2 \qquad (3,5)$$

The statistical model of visual detection on which this equation rests was originally formulated by Rose [1948] in, what is now, a classic paper in the area.

III.4.2.1.1 DETERMINANTS OF IDEAL PERFORMANCE

The left side of Equation 3.5 describes the information available from the scene. Brightness B and summation area h*2 determine the size of the particle count and thus the size of the sample on which a decision will be based. Contrast C compares the differential particle count, ∂N, with the total count. These terms establish the significance of a given change in brightness in terms of the statistically expected variations for the total sample size.

The product is an expression of the familiar statistical concept that a reliable decision can be based on either a few samples of a strong signal or many samples of a weak signal.

The right side of Equation 3.5 expresses the performance limitations of the detector. It accounts for the fact that only a fraction of the information present in a scene element will be picked up by the detector and made available to a decision process. This fraction depends on how many of the available quanta will be collected by the optical system. If a real photo-detector were then used to convert the collected light into conduction electrons, the quantum yield of the conversion process would be included in K+2. This constant also includes the confidence ratio , K+1.

III.4.2.1 ACCOMMODATION FOR AN IDEAL DETECTOR

Rose conceived of Equation 3.5 as a universal performance scale. The performance potential of various imaging devices, limited by their primary quantum and optical efficiencies, could be compared with the standard established by an ideal device (whose quantum efficiency is assumed to be unity). We will use this equation to compare the performance of a given sensor in various detection tasks, using alternative accommodations.

In the ideal detector, the only adjustable

parameters are the focal length and diameter of the lens, and the exposure time. To make the influence of these parameters on K+2 explicit, consider the relations for image brightness derived from Figure 2.10a. Let the scene radiate No quanta/square foot/second. Assuming a Lambert distribution, the number of quanta captured by the lens and imaged into h+2 will, by Equation 2.44 be h+2*No*(D+2)/4(f+2) per second. In an exposure time, t, the total number of quanta absorbed and counted will be

$$N = \frac{N_0 \times t \times D^2 \times h^2}{4 f^2} \qquad (3.6)$$

To relate the physical model, expressed by Equation 3.6, to the terms used in Equation 3.5, we make the following substitutions in Equation 3.6:

1 lumen of white light = 1.3 *10+16 quanta/sec. (3.7a)

1 ft.c. = 1 lumen/square foot (3.7b)

B (brightness in foot candles) = No/1.3*10+16

= K+3*No lumens/sq. ft. (3.7c)

N = K+1+2/C+2 (from Equation 3.3) (3.7d)

The resulting equation contains an explicit expansion of K+2 in terms of the camera parameters.

$$BC^2h^2 \geq (4K_3)\left(\frac{f^2}{t\,D^2}\right) \times K_1^2 \qquad (3.8)$$

The camera parameters provide flexibility in satisfying the basic requirements for detection. Suppose it is necessary to detect a small feature. The most straightforward way is simply to have the decision process consider information from a smaller area of the detector (reduce h). h, however, is constrained by Equation 3.8. If the contrast is weak, the reduction in sample size, resulting from an uncompensated reduction in h, may be intolerable. By increasing exposure time, however, h can be decreased with no change in the overall number of samples collected. Alternatively, the iris can be opened to admit more light and the focal length increased so that h•2 corresponds to a smaller area of the scene.

III.4.2.1.3 SELECTION OF ACCOMMODATION ALTERNATIVES

A criterion like 3.8 only indicates the need to accommodate. It suggests no preference among the possible adjustments. This freedom allows us to choose the accommodation that is most favorable in the overall context of the task. We illustrate this idea by considering two specific contexts in which the detection of small detail might be necessary.

The first case entails a search for a small object that may lie anywhere on the entire table-top. We saw earlier that an appropriate image for an unconstrained

search is one encompassing a broad spatial window. In this context increasing the focal length is undesirable, because it decreases the field of view. Similarly a narrow iris is preferred to maintain good depth of field.

Temporal averaging entails an overhead in processing time, but this fact is not significant in acquisition, since a coarse sampling pattern is employed. Consequently, for our system, averaging is the preferred method of increasing signal to noise during a search.

If, on the other hand, the task was to detect a low contrast edge in a specific location, the lens accommodations would be preferable. For other applications some mixture of these accommodations may be appropriate. In the worst case, both may be needed to maintain adequate signal/noise.

In practice, the optimum mix of accommodations rests on reliability and efficiency considerations. These depend on the task and on the characteristics of the camera hardware. There has not been sufficient experience with the working system to establish formal criteria for trading-off speed and reliability. Consequently, the most appropriate parameter to accommodate in a specific application is currently chosen on a heuristic basis. However, once this choice has been made, equations analogous to 3.8 dictate what adjustments are needed to attain the desired decision confidence.

III,4,2,4.1 STATISTICAL DETECTION ANALYSIS FOR A VIDICON CAMERA

In this section we extend the statistical analysis, developed for the ideal detector, to the more realistic and useful case of a vidicon camera. This departure from ideality has several consequences.

The performance of the ideal device was limited only by the fundamental and unavoidable fluctuations in the arrival of light quanta. Physical devices can only detract from this performance, The vidicon in particular suffers from two basic faults:

1. A significant amount of additional noise is introduced in the conversion process by the physical mechanisms (dark current, amplifier noise, etc.) discussed in Chapter 2,

2. The vidicon has a fractional quantum yield. Its effect is to further reduce the available image signal/noise for a given scene quality,

As a partial compensation, our camera chain has many additional accommodation options which allow greater flexibility in adapting the available signal/noise to suit the task requirements. Because of the computer's limited processing capacity, a selectively accommodated vidicon will often be a more appropriate source of data for perception than an unselective ideal detector. (Signal/noise is, after

all, not the only criterion of an appropriate image.)

III.4.2.4.1 VIDICON PERFORMANCE IN EDGE DETECTION

We will study the effects of camera accommodation in the specific context of an (intensity) edge detection task. This particular task was chosen for several reasons:

1. Edges are the primary means of information representation in our system. The results derived in this section will be directly utilized in two important applications, the edge verifier (Chapter 4) and the edge follower (Chapter 5).

2. Edge detection provides an opportunity to exercise all of the camera accommodations in an effective way.

3. Several practical limitations in a vidicon type camera will be highlighted. We will demonstrate the ways in which accommodation can help to overcome these difficulties.

4. The second most common use of video data is to determine the brightness of a surface. This information, in turn, is usually used to compare the brightnesses of several surfaces exposed to a common source of illumination (For color vision, the relative intensities of a single surface observed through different color filters are important.). The factors that influence the ability to distinguish two

intensities separated in space or time are identical to those that determine whether an edge of equivalent contrast can be detected.

III.4.2.4.2 A T-TEST FORMULATION OF STATISTICAL DETECTION

A more convenient statistical criterion than 3.5 is desirable. We note that the detection problem is equivalent to the standard statistical question of whether or not two groups of samples (eg. one drawn from the background, the other from the presumed location of a pattern) were actually drawn from the same distribution.

The difference of the means of N samples drawn from two normal distributions is T-distributed with 2N-2 degrees of freedom. For 2N>30 this distribution is virtually identical with a normal distribution of standard deviation

$$s_d = \frac{\sqrt{s_1^2 + s_2^2}}{\sqrt{N}} \qquad (3.9$$

(where S.1,S.2 are unbiased estimators of the standar deviations of the original distributions - Figure 3.6b).

If no pattern were present, the difference in means would have an expectation of zero. The T-score is a comparison of the actual observed difference in the means with the expected devlation, S.d.

$$T = \frac{M_1 - M_2}{S_d} \qquad (3.10)$$

This ratio is used to test the (null) hypothesis that no pattern was present. The probability that there is no edge is given by

$$\text{Confid}\Big|_{\text{null}} = 1 - \Phi(T) \qquad (3.11)$$

(Phi is defined by Equation 2.28.) If the difference (M1-M2) is significantly greater than Sd, a pattern is assumed present with confidence Phi(T).

The T-test Applied to Edge Detection

The T-test can discriminate between an ideal step discontinuity separating two uniform intensity surfaces, and a completely uniform surface, in the presence of Gaussian detector noise. Intensities are sampled from small regions on each side of a suspected edge and tested to determine the likelihood that they were drawn from the same distribution. In the following sections the rules of good accommodation for edge detection will be formalized in terms of this test. In Chapter 4, the T-test will be modified to reject gently sloping surfaces, often found in the real world.

III,4,2,4,3 SIGNAL/NOISE RATIO

The significance of a given absolute difference in sample means, as reflected by the T-score, is directly dependent on the signal/noise of the original distributions. We will develop expressions for the basic signal/noise figure of the camera system as a function of the various camera parameters. The goal of accommodation for edge detection is to maximize the signal/noise figure for intensities in the local region where an edge is sought.

III,4,2,2,4 A SIGNAL/NOISE MODEL OF THE VIDICON

Statistic fluctuations in the bias current of the preamplifier is the primary noise source in a vidicon camera. Secondary sources are shot noise in the signal, in the vidicon dark current, and in the thermal current of the target resistor. Using results developed earlier in this thesis, we write the signal/noise of a brightness measurement as:

$$Z = \frac{i_s}{i_n} = \frac{i_s}{\left(i_{sn}^2 + i_{dn}^2 + i_{rn}^2 + i_{an}^2\right)^{1/2}} \qquad (3.12)$$

The amplifier mean square noise power (i*2*an) was found to be 5,66 *10*(-18) (Chapter 2). Signal current (i*s) is

defined by Equation 2.8, signal shot noise (i^2_{sn}) by Equation 2.15, dark current shot noise (i^2_{dn}) by Equations 2.18 and 2.20, and resistor noise (i^2_{rn}) by Equation 2.22.

$\bar{z}$ is a function of the target voltage E_T, the exposure or storage time, T_s, the read-out time per resolution cell, Tau, and the brightness level, L, incident on the vidicon. T_s and Tau are both fixed by the vidicon scanning mechanism. Tau, however, can be effectively increased by averaging the results obtained for several resolution cells, or alternatively, by temporally averaging the results of one cell over several frames. (Although spatial and temporal averaging have similar effects on $\bar{z}$, they are not always interchangeable in a task context.)

The influence of the readout interval (1/tau) can be factored from each of the noise terms in the denominator of $\bar{z}$. Equation 3.12 can then be written in the form:

$$Z = \frac{\sqrt{n\tau}\, K_3\, L_v^{\gamma}\, E_T^{\alpha}}{\left(A\, L_v^{\gamma}\, E_T^{\alpha} + B\, E_T^{\beta} + C\right)^{1/2}} \qquad (3.13)$$

where we have substituted

$$i_s = K_3\, L_v^{\gamma}\, E_T^{\alpha} \quad \text{(from 2.8)} \qquad (3.14)$$

$$(K_3 = 2.798 \times 10^{-9}$$

$$\gamma = 0.65$$

$$\alpha = 1.429)$$

$$i_{sn}^2 = \frac{A\, L_v^{\gamma}\, E_T^{\alpha}}{n\tau} \qquad \text{(from 2.8, 2.15)} \tag{3.15}$$

$(A = e\, K_3 = 4.476 \times 10^{-28}$

$n \equiv$ number of samples

$\tau \equiv$ read out time $= (1.7 \times 10^{-7}$ sec)

$$i_{dn}^2 = \frac{B\, E_T^{\beta}}{n\tau} \qquad \text{(from 2.18, 2.20)} \tag{3.16}$$

$(B = 2.19 \times 10^{-32}$

$\beta = 3.97)$

$$i_{rn}^2 = \frac{C'}{n\tau} \qquad \text{(from 2.22)} \tag{3.17}$$

$(C' = 8 \times 10^{-27})$

$$i_{an}^2 = \frac{C''}{n\tau} \tag{3.18}$$

$(C'' = 9.64 \times 10^{-25})$

$$C = C' + C'' = 9.72 \times 10^{-25} \tag{3.19}$$

L_v is the light actually incident upon the faceplate of the vidicon. Using Equation 2.45, we can relate L_v to the illuminance, (L), incident from the scene upon the lens:

$$L_v = \frac{L}{4 f_\#^2} \tag{3.20}$$

Signal/noise increases with light level. The numerical results generated in the following discussion will thus be based on scene illuminance, observed with the widest aperture (f#=1.4). We reformulate 3.13 in terms of scene illuminance, (L), using 3.20 with f# 1.4:

$$Z = \frac{\sqrt{n\tau}\, K_\ell\, L^\gamma\, E_T^\alpha}{\left(A_\ell\, L^\gamma\, E_T^\alpha + B\, E_T^\beta + C\right)^{1/2}} \tag{3.21}$$

where

$$K_\ell = K_3/L\,(1.4)^2 \tag{3.22a}$$

$$A_\ell = A/L\,(1.4)^2 \tag{3.22b}$$

III.4.2.4.5 OPTIMIZATION OF SIGNAL/NOISE VS. E_T

As target voltage is increased, the signal will

increase relative to both signal shot noise and constant C (thermal and amplifier noise). However, because $\beta > 2\alpha$, dark current will increase faster than the signal. Thus, signal/noise is limited at any light level by the dark current noise.

At low target voltages, dark noise is much less than amplifier noise. Thus, Z will initially increase with E_T. The optimal target voltage (E_{opt}) is that level at which the dark component dominates the total noise to the extent that a further increase in E_T will reduce Z.

Since the signal, but not the dark noise, depends on light, E_{opt} will be lowest for low light levels. At 1 ft.c. the theoretically optimal target voltage was found (numerically) to be 115v.

The target voltage is limited in practice by two considerations which do not enter Equation 3.21. There is first the possibility of damaging the vidicon by applying excessive target voltage. The maximum safe target voltage (E_{dam}) is an inverse function of light level. The absolute maximum target voltage allowed by the auto-target circuit is 45v. This limit minimizes the chance of accidental damage from local highlights that do not affect the average intensity upon which the auto-target relies.

The finite dynamic range of the quantizer imposes the second limitation. The target voltage cannot be raised so high that the amplified signal will exceed the maximum

1v. range of the quantization window. We rewrite Equation 2.10 in terms of scene illuminance (using 3.20) and solve it to obtain the E↓T that yields a 1v. output at a light level, L.

$$E_{sat} = (FL^{\gamma})^{-1/\alpha} \qquad (3.23)$$

$$F = \left| \frac{0.0095}{[4\ (1.4)^2]^{0.65}} = 0.0025 \right|$$

E↓sat is inversely related to light level, as shown graphically in Figure 3.7.

E↓sat and E↓dam are both less than E↓opt at 1ft.c. illumination. Furthermore, both decrease with illumination while E↓opt rises. Since Z is a unimodal function of E↓T, an empirical rule for setting the target voltage to maximize Z is given by

$$E = \text{minimum}\ (E_{sat}, E_{dam}) \qquad (3.24)$$

In homogeneous scenes, E↓sat will always be less than E↓dam. However, in non-homogeneous scenes, E↓dam, imposed by a scene highlight, may be lower than the E↓sat for a lowlight in which we are interested. (The rate of decrease of E↓dam with increasing L is a consideration that

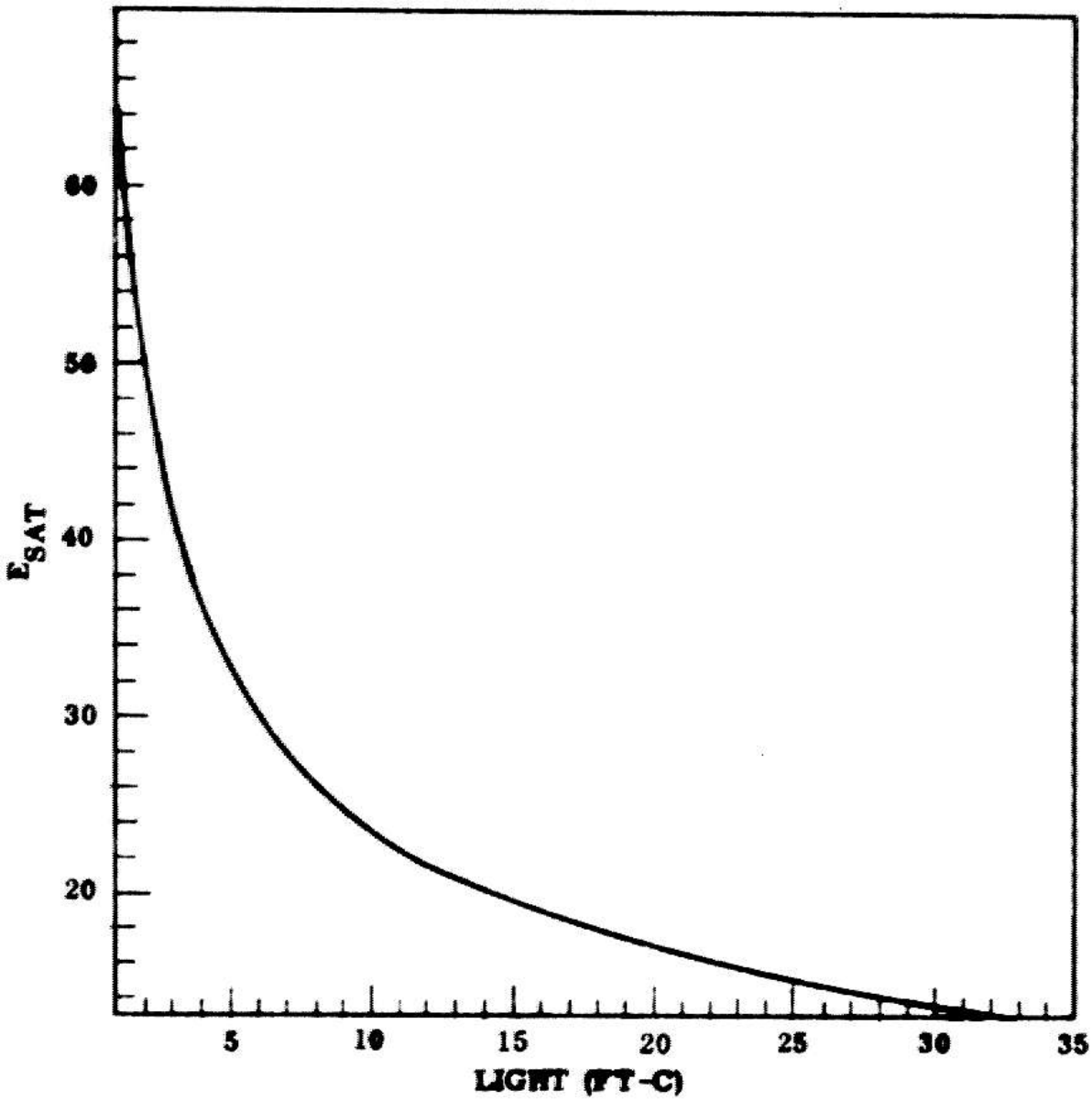


Fig. 3.7 E_{SAT} Versus L

must be taken into account when selecting the optimal light level for scenes with a wide dynamic range. Clearly, too much light could detract from the ability to see lowlights.)

The optimal signal/noise, with E*T limited by saturation, can be found by substituting Equation 3.23 into 3.21

$$Z_{sat} = \frac{\sqrt{n\tau}\, K_\ell\, L^\gamma \left|(F\, L^\gamma)^{-1/\alpha}\right|^\alpha}{\left(A_\ell\, L^\gamma \left|(FL^\gamma)^{-1/\alpha}\right|^\alpha + B\left|(FL^\gamma)^{-1/\alpha}\right|^\beta + C\right)^{1/2}} \qquad (3.25)$$

By grouping constants Equation 3.25 can be expressed more simply as

$$Z_{sat} = \frac{G}{\left(HL^{-\frac{\beta\gamma}{\alpha}} + J\right)^{1/2}} \qquad (3.26)$$

where

$$G = \sqrt{n\tau}\, K_\ell\, F^{-1} \qquad (3.27)$$

$$H = BF^{-\beta/\alpha} \qquad (3.28)$$

$$J = A_\ell F^{-1} + C \qquad (3.29)$$

Equation 3.26 shows that under saturation limited conditions, as the light level is raised, the signal/noise ratio will climb towards a finite asymptote

$$\lim_{L \to \infty} Z_{sat} = GJ^{-1/2} = 119 \qquad (3.30)$$

This limit is illustrated in Figure 3.8. At very low values of E↓T (approaching Ø), vidicon performance will deteriorate. Excessively bright lighting should thus not be used.

III.4.2.2.6 APPLICATION OF SIGNAL/NOISE MODEL TO EDGE DETECTION

The preceding discussion was tailored to the problem of detecting a weak signal against a background of vidicon noise. The basic signal/noise model of a vidicon will now be used to develop accommodation criteria for the detection of edge contrast. An edge is assumed for present purposes to be a simple step of intensity. Let the step height be expressed as a fraction, rho, of the signal, I↓s↓1, representing the mean brightness of the brighter boundary of the edge. The signal from the dimmer side is then

$$i_{s_2} = (1 - \rho)\, i_{s_1} \quad , \quad 0 < \rho < 1 \qquad (3.31)$$

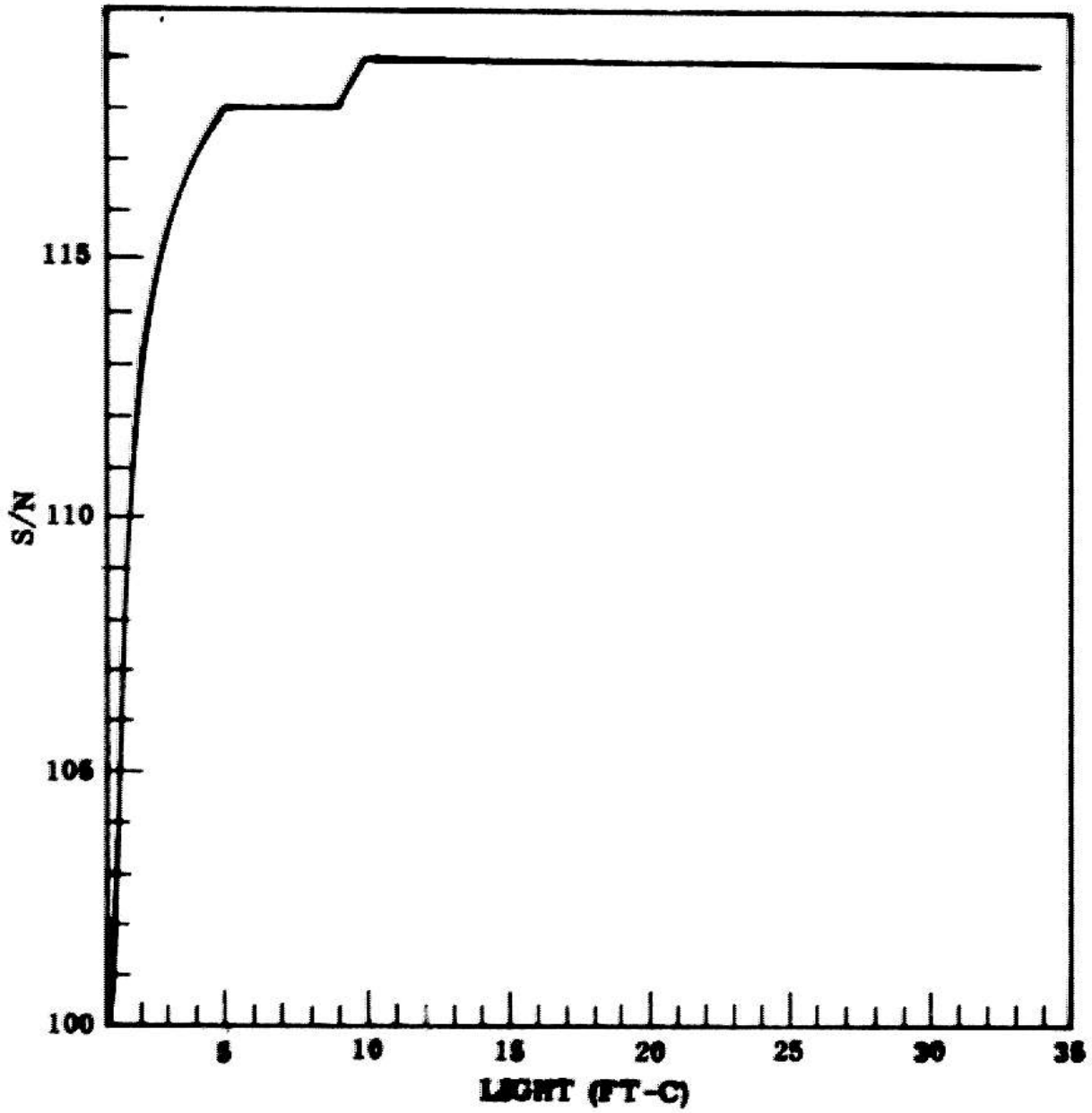


Fig. 3.8 Z_{SAT} Versus L

The T-score, (3,10), which represents the significance of an edge is expressed in terms of these signals as

$$\frac{i_{s_1} - (1 - \rho)\, i_{s_1}}{s_d} = \frac{\rho\, i_{s_1}}{s_d} = Z_c \qquad (3.32)$$

(where S+d is the expected standard deviation of an intensity difference with no edge). Equation 3,32 defines what we shall call "contrast" signal/noise, Z+c. (For clarity, we shall refer to our previous model as "detection" signal/ noise, Z+d). For weak edges, S+d is proportional to the noise on either side of the edge. In this case, Z+c is (by 3,32) directly proportional to Z+d.

We will now establish a formal relationship between contrast and detection signal/noise. The signal, i+s+1, can be expressed, according to 3.21 as

$$i_{s_1} = K_i\, L^{\gamma}\, E_T^{\alpha} \qquad (3.33)$$

The signal from the dark side of the edge is then (by 3,31)

$$i_{s_2} = (1 - \rho)\, i_{s_1} = (1 - \rho)\, K_i\, L^{\gamma}\, E_T^{\alpha} \qquad (3.34)$$

Using the denominator of Equation 3,21 as a model, we can express the respective noise in these mean brightness measurements as

$$\sigma_{i_1} = \frac{1}{\sqrt{nT}} \left(A_f L^{\gamma} E_T^{\alpha} + B E_T^{\beta} + C \right)^{1/2} \qquad (3,35)$$

and

$$\sigma_{i_2} = \frac{1}{\sqrt{nT}} \left[A_f (1 - \rho) L^{\gamma} E_T^{\alpha} + B E_T^{\beta} + C \right]^{1/2} \qquad (3,36)$$

n (the number of samples or the size of the operator) is assumed to be equal on both sides of the edge. Using Equation 3,9 we can combine Equations 3,31-3,36 to obtain an analytic expression of the contrast signal/noise,

$$Z_c = \frac{\sqrt{nT}\, \rho\, K_f\, L^{\gamma} E_T^{\alpha}}{\left[(2 - \rho) A_f L^{\gamma} E_T^{\alpha} + 2 B E_T^{\beta} + 2C \right]^{1/2}} \qquad (3,37)$$

Signal/noise is an important consideration only in the detection of very weak edges. Since limiting detection is our primary concern, we make the assumption that rho<<2, Z+c is then directly proportional to the detection signal/noise, calculated for the bright side of the edge,

$$Z_c = \rho \frac{Z_d}{\sqrt{2}} \qquad (3,38)$$

(Since rho is small, Z+c is considerably less than the corresponding value of Z+d,)

As a result of 3.38, the optimum signal/noise for edge detection is again obtained using condition 3.24. Under normal saturation-limited conditions the target should be adjusted so that the bright side of the edge almost saturates the amplifier. A lower voltage will reduce Z•d while a higher voltage will compress contrast due to clipping. More light is also effective in increasing signal/noise, up to the asymptotic limit shown in Figure 3.8.

Z•c expresses the T-score (3.10) of the intensity difference observed across an edge, Phi(Z•c) (see 3.11) thus defines the confidence that a given Z•c represents an edge. (For instance, a T-score of 3 means that with 95% confidence an actual edge is present.) Equations 3.11 and 3.38 determine the weakest edge (ie, smallest rho) that can be detected at a required confidence level, given an attainable Z•d. For example, requiring 95% detection confidence,

$$\rho_{min} = \frac{3\sqrt{2}}{Z_d} \quad . \text{ for 95\% confidence} \qquad (3.39)$$

The smallness of rho•min is an important measure of accommodation effectiveness for edge detection.

Effect of Color Filters

A black/white television system maps a complex illumination spectrum into a one-dimensional grey scale.

This mapping often obscures colored edges which were obvious in the original scene. In these cases, color filters can be used to augment the contrast signal/noise.

Figure 3.9 clarifies the effects of color filters on contrast. Several simplifications have been introduced to isolate the effects of object color.

1. I_1, I_2 represent the total luminous energy of broad sources of spectrally flat "white" light. They provide uniform illumination over the surfaces upon which they are incident.

2. The camera is aligned so that the principal ray of the lens intersects the edge perpendicularly at a point midway along it. This assumption avoids superfluous directional reflectance factors.

3. The camera has a uniform normalized spectral sensitivity $S(\lambda) = S = 1$.

The light level, L_1, reflected by the right face of the cube, is proportional by Equations 2.45 and 2.59 to

$$L_1 \sim I_1 \int_\lambda r_1(\lambda)\, d\lambda = I_1 R_1 \qquad (3.40)$$

Similarly, the left face reflects

$$L_2 \sim I_2 R_2 \qquad (3.41)$$

If the cube were homogeneous, R_1 would be identical

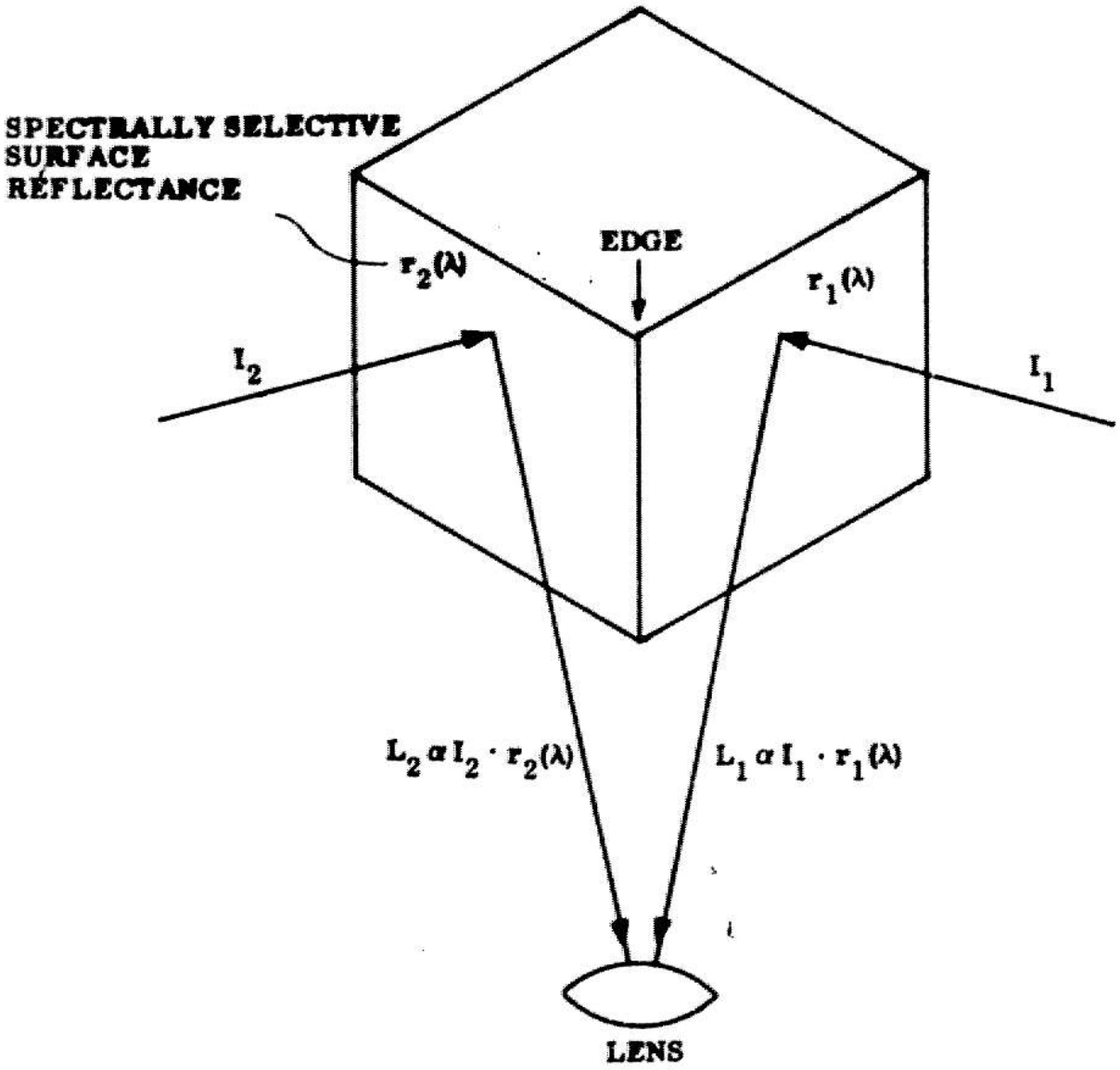


Fig. 3.9 Formation of an Intensity Edge

to R₂2. The edge would then be visible only because of differences in the directional nature of the illumination field (which, in general, could include color differences).

Color filters can affect the contrast signal/noise of the image of this edge in two ways:

1. The similarity between the spectral transmission of the filter and the spectral reflectances on each side of the edge can alter the relative amount of light received from each side.

2. The combination of filter rejection and absorption will attenuate the total light energy reachin the vidicon from both sides of the edge.

The reduction in light level will (by Equation 3.26) reduce or at best maintain the saturation limited Ẑ+d on the brighter side (as seen through the filter). The spectral bias of the filter must compensate for this reduction by increasing the effective disparity in the surface reflectances, (ie, increasing the relative step height, rho). By Equation 3.39 we see that for a color filter to enhance the detection of an edge, the relative increase in rho due to spectral effects must be greater than the relative decrease in Ẑ+d due to attenuation.

An Analytic Discussion of Filter Effects

With filter function $F(\lambda)$ the light intensity corresponding to the bright (eg, right) side of the edge

will be proportional to

$$L_1' \sim I_1 \int_\lambda F(\lambda) r_1(\lambda)\, d\lambda = I_1 R_1' \qquad (3.42)$$

Similarly, for the dim side

$$L_2' \sim I_2 R_2' \qquad (3.43)$$

Let us first examine the explicit effect of this filter on rho. In the absence of a filter, Equations 3.33, 3.40, and 3.41 can be used to express the difference signal corresponding to the edge

$$I_s = i_{s_1} - i_{s_2} \approx K_l E_T^\alpha \left[(L_1)^\gamma - (L_2)^\gamma\right] \qquad (3.44)$$

By analogy with Equation 3.34

$$i_s = K_l E_T^\alpha \left[(L_1^\gamma) - (1 - \rho)(L_1)^\gamma\right] \qquad (3.45a)$$

$$i_s = K_l E_T^\alpha \left[L_1^\gamma - \left|(1 - \rho)^{1/\gamma} L_1\right|^\gamma\right] \qquad (3.45b)$$

Define Gamma,

$$\Gamma \equiv (1 - \rho)^{1/\gamma} \qquad (3.46)$$

A relative edge signal, rho, thus corresponds to a light ratio of Gamma,

$$L_2 = \Gamma L_1 \tag{3.47}$$

Gamma can be resolved into two components corresponding to the reflectance and illumination factors of the L's.

$$L_2 R_2 = \Gamma L_1 R_1 = (\Gamma_1 L_1)(\Gamma_2 R_1) \tag{3.48}$$

where

$$\Gamma = \Gamma_1 \Gamma_2 \ , \ \Gamma_1 = L_2/L_1 \ , \ \Gamma_2 = R_2/R_1 \tag{3.49}$$

Gamma$_{i+1}$ is independent of the filter color, because white light is assumed. If the colored edge were painted on a plane surface, Gamma$_{i+1}$ would simply be 1. For the cube edge, Gamma$_{i+1}$ will be a constant representing the directional intensity difference between I$_{i+1}$ and I$_{i+2}$.

A color filter will alter the ratio of R$_{i+2}$/R$_{i+1}$ as per Equations 3.42 and 3.43. Using Equation 3.46, the change in reflectance ratio can be described in terms of relative signal contrast. The selectivity of the color filter has changed rho from

$$\rho_1 = \left[1 - \Gamma_1^{\gamma} (R_2/R_1)^{\gamma}\right] \tag{3.50}$$

to

$$\rho_2 = \left[1 - \Gamma_1^{\gamma} (R'_2/R'_1)^{\gamma}\right] \qquad (3.51)$$

The effective attentuation introduced by the color filter is found by comparing the intensity (L) of the brighter side, as observed with and without a filter. Assuming the right side is brighter in both cases, the effect of attenuation on detection signal/noise, Z•d, is obtained by solving Equation 3.26 with L•1,(given by Equation 3.40) and L•1' (given by Equation 3.42). (Equation 3.26 implies that E•T is optimized at both light levels.)

Let Z•D' be the highest detection signal/noise that can be attained on the brighter side of the edge viewed through a color filter and Z•D the corresponding figure without the filter. Equation 3.39 defines the minimum signal contrast required in each case for 95% reliable detection:

$$\rho_{min} = \frac{3\sqrt{2}}{Z_d} \qquad (3.52a)$$

$$\rho'_{min} = \frac{3\sqrt{2}}{Z'_d} \qquad (3.52b)$$

Since Z•D'≤ Z•D, rho'•min ≥ rho•min. The edge will be visible with the filter if

$$\rho_2 \geq \rho'_{min} \qquad (3.53)$$

and without the filter if

$$\rho_1 \geq \rho_{min} \qquad (3.54)$$

The color filter will provide an overall improvement in the ability to detect an edge when

$$\rho' Z'_d > \rho Z_d \qquad (3.55)$$

In general, this condition will depend on the intensity and spectral composition of the source, as well as the spectral nature of the object reflectance and filter transmission characteristics. The relative spectral sensitivity of the vidicon tube (Figure 2.16) must also be taken into account in comparing relative contrast with and without filters.

Pragmatics of Color Filter Usage

The meaning of the analytic results can be clarified by a concrete example of filter usage. Consider the boundary between red and blue regions painted on a plane surface. The camera is roughly equally sensitive to these two hues (Figure 2.16). With uniform white illumination,

the difference in grey levels across the edge is likely to be small. If Ż•c at E•sat is insufficient, contrast can be improved using a red or blue filter.

Ideally, the transmission characteristic of the filter should match one of the surface hues. Let us assume a red filter is used. The camera sensitivity is then adjusted so that the signal corresponding to the red (eg. bright) side is just below saturation, thus optimizing Ż'•0. Since the spectra of typical red and blue surfaces approximate the transmission characteristics of the corresponding filters (shown in Figure 2.17), these accommodations produce an impressive gain in contrast signal/noise.

Color filters can also be helpful when target voltage is limited by E•dam. However, the limiting highlight feature must not be one of the edge boundaries. Suppose the brighter side of the edge is colored differently than the highlight feature. A color filter can then be used to reduce the highlight brightness relative to that of the edge surface. E•dam will be increased, perhaps above E•sat, for the edge. Hence, E•T can be raised with a corresponding increase in Ż•c for the desired boundary. In this application the color of the other edge boundary is unimportant. A significant advantage can be realized even if it were the same color as the bright side (eg. for any homogeneous object). Furthermore, the highlight feature

need not be colored. Even if it were white, the dynamic range of the scene would still be selectively compressed relative to the contrast across the edge. It is always most desirable that the feature we wish to perceive be the brightest element in the scene. A matched color filter can help achieve this.

Pragmatics of Lens Accommodation for Edge Detection

In this section, various lens accommodations, analyzed earlier, will be applied to edge detection.

1. Iris

We have already established the desirability of more light to enhance signal/noise (first in Equation 3.8 with the ideal detector and again in Equation 3.21 for the vidicon). Equation 2.44 formalizes the obvious: to get more light, open the lens iris. A wider iris also decreases the depth of field. This side effect makes it easier to eliminate unwanted (textural) detail by intentional defocusing (as illustrated in Figures 1.11, a,b).

2. Focal length

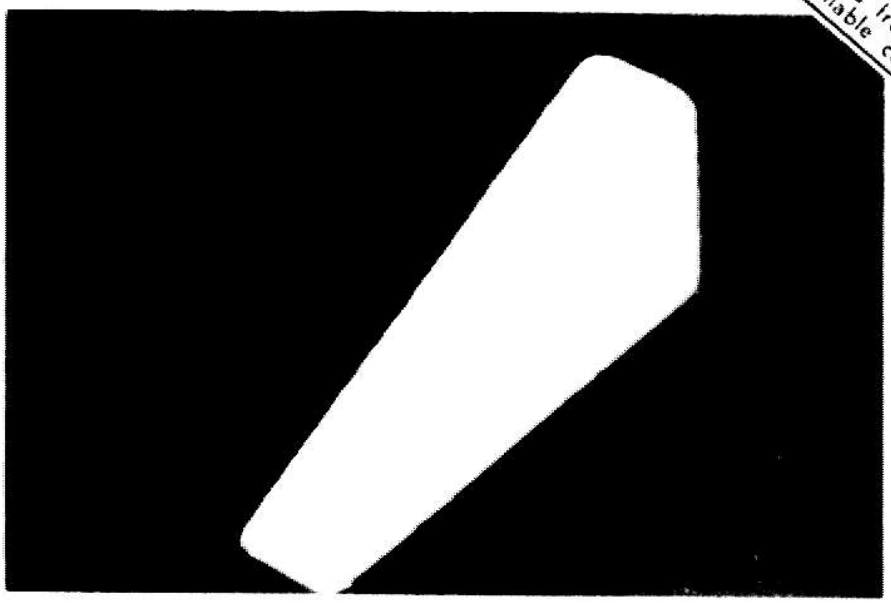
The focal length of a lens affects both depth of field (Equation 2.55c) and the image magnification (Equation 2.38). Consider first the problem of detecting an ideal edge

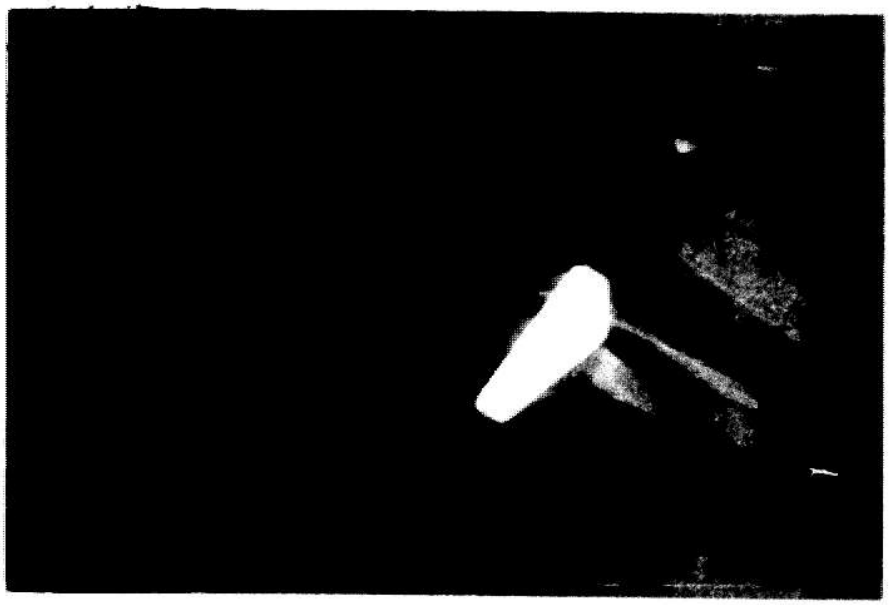
in the presence of detector noise. The higher the magnification, the more resolution cells will be spaced along the length of an edge. Thus, magnification provides the opportunity to increase detection confidence by taking larger samples (larger n in Equation 3.37) from a given area along the edge. Alternatively, more independent samples of a given size can be fit over the length of an edge. This is advantageous when seeking global evidence to confirm the presence of an expected edge.

For non-ideal edges, magnification is important for additional reasons. Figure 3.10a is a photograph of the television monitor, showing the edge of a wedge observed at high magnification. Note that cracks and paint anomalies tend to concentrate along the length of the edge. These "edge effects" distinguish the actual discontinuity with far more clarity than the contrast across the edge.

Edge effects are often found at surface boundaries of real objects in the physical world. In many cases, these anomalies provide the primary evidence which people use to confirm the presence of low contrast edges.

A long lens allows the machine to capitalize on the presence of these edge artifacts. Magnification broadens the actual discontinuity into a small surface, as stylized in Figure 3.11. Tests designed for conventional surface boundaries can then detect edge effects that would appear as one-dimensional lines (if at all) through shorter,





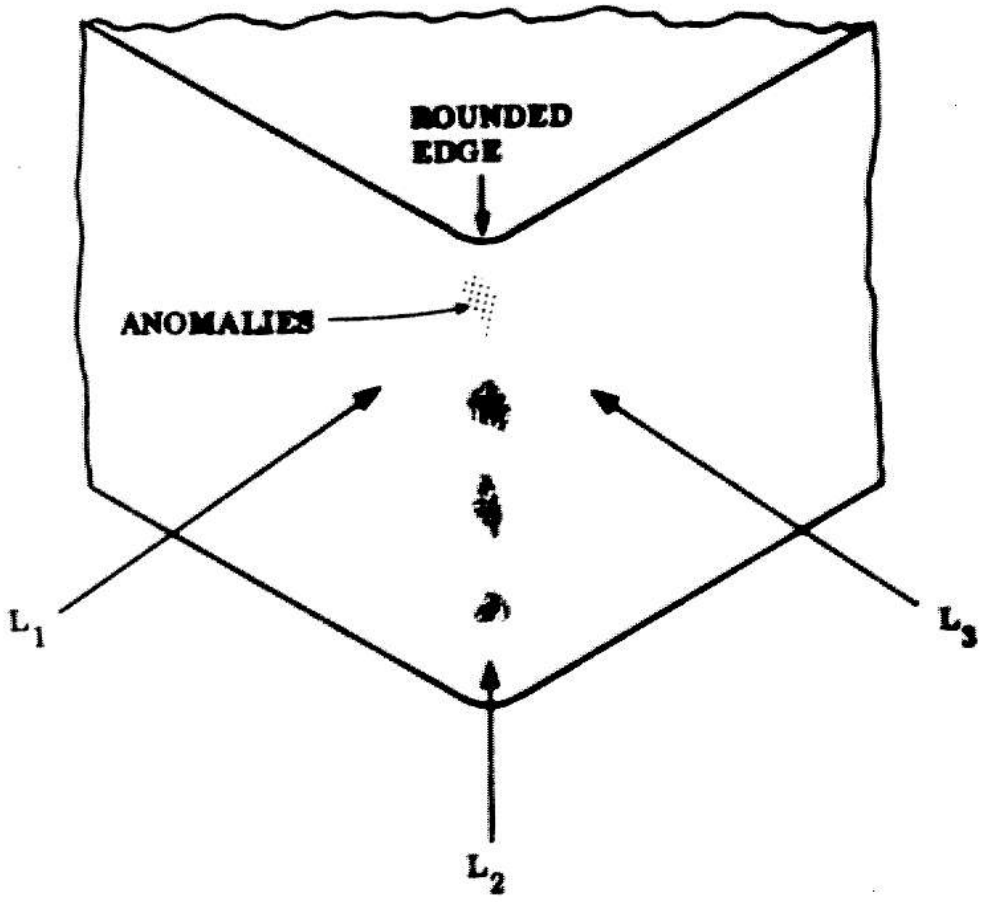


Fig. 3.11 An Imperfect Edge Under High Magnification
(Stylized Presentation)

wider angle lenses (Figure 3.10b).

As usual, the choice of lens depends ultimately on which magnification gives the best performance. For strong edges that are marred by occasional, very small anomalies, a short lens may be effective. Low magnification maps a larger area of the surface into each resolvable cell of the image. This is a form of spatial averaging that can smooth the unwanted visual noise on the edge.

In addition to anomalies, there are several other ways in which edges, appropriately magnified, can give themselves away. Specular reflections, for instance, are often concentrated at sharp edge discontinuities. Rounded edges under magnification can, on the other hand, provide enough surface area to pick up directional discontinuities that may be present in the lighting field. L_2 (Figure 3.11) could be more different from L_1 and L_3 than the latter two are from each other. These considerations are of great practical utility in finding edges.

The higher the magnification is, the more limited the field of view will be. Limiting the field of view to the immediate vicinity of the edge helps insure that the edge itself, and not some irrelevant highlight, will determine how high the target voltage can be set. This use of magnification is analogous to a person using his hand to shield his eyes from the sun to get a better view of details in a shadowed region. It is an important reason for

changing lenses in practice.

A final use of optical magnification is to reduce the influence of surface gradients relative to the sharper transition of the edge (ie. between surfaces). With sharp focus the width of a good, abrupt edge is limited by electronic bandwidth and not by optical resolution. This transition will remain about three raster units wide regardless of which lens is used. Typical surface gradients fall well within the amplifier bandwidth. Their slope (in volts/raster unit) will decrease inversely with increases in spatial magnification.

3. Focusing

We have mentioned the use of defocusing to intentionally blur unwanted fine detail when tracing intensity edges. For edge effects, on the other hand, fine detail is the primary characteristic that distinguishes the edge. In this case, sharp focus on the actual plane of the discontinuity is crucial. When texture operators are available, they will also require good focus to see fine detail.

The choice of sharp or dull focus thus depends on the type of edge. The decision should optimally be based on performance feedback: ie., the focus is adjusted to maximize the confidence of the actual edge detection test at a point thought to lie on the edge. If performance falls off

when the edge is tracked into unknown areas, the focus can be re-optimized. These techniques will be developed in Chapter 4.

Clip-setting for Edge Following

In practice, edges must be found in quantized data. It is necessary to consider the additional noise contributed by the quantization process when assessing the practical limits of our system. In this section we apply the quantization theory developed in Chapter 2 to the specific task of edge detection.

1. Clip-setting for Differential Measurements

We again distinguish the cases of differential and absolute measurements. Differential measurements are useful for rapid extraction of high contrast edges. A simple threshold criterion is applied to the difference of the intensities observed on both sides of a presumed edge boundary. If the difference is large enough, the edge is assumed to be present.

$$(i_{s_1} - i_{s_2}) > T \rightarrow \text{Edge} \qquad (3.56)$$

The principal effect of quantization noise on this test is to limit the minimum edge height that can be resolved. The most effective quantization window is thus the narrowest

one. It should be centered on the average intensity across the edge, as illustrated in Figure 3.12. If the difference of intensity across the edge exceeds 1/8v. (after sensitivity, lens and color filters have been tuned to optimized contrast) then the clips will be positioned as in Figure 2,6. In this case, the difference signal will be clipped at 15, independent of noise that may be riding on the edge. Even if the signals can not be hard clipped, the narrowest quantization window will always maximize differential separation.

The standard deviation of the camera noise is, on the average, about one quantization level at the narrowest window width (sigma = 1/112v.). A minimum threshold of at least T=3 quantization units (I↓s↓1-I↓s↓2≥3/112v.) should thus be used to avoid false detection with 95% confidence.

2. Clip-setting for Absolute (Statistical) Measurements

For weak edges, a simple intensity difference is not a sufficiently reliable test. The statistical T-test in which signals are explicitly normalized by the observed noise, is a more suitable indication of edge significance. The quantizer theorem suggests that to avoid errors the quantization levels should be no larger than the standard deviation of the signal we are trying to measure. In view of the observed noise level, the heuristic accommodation chosen for differential measurements is also best according

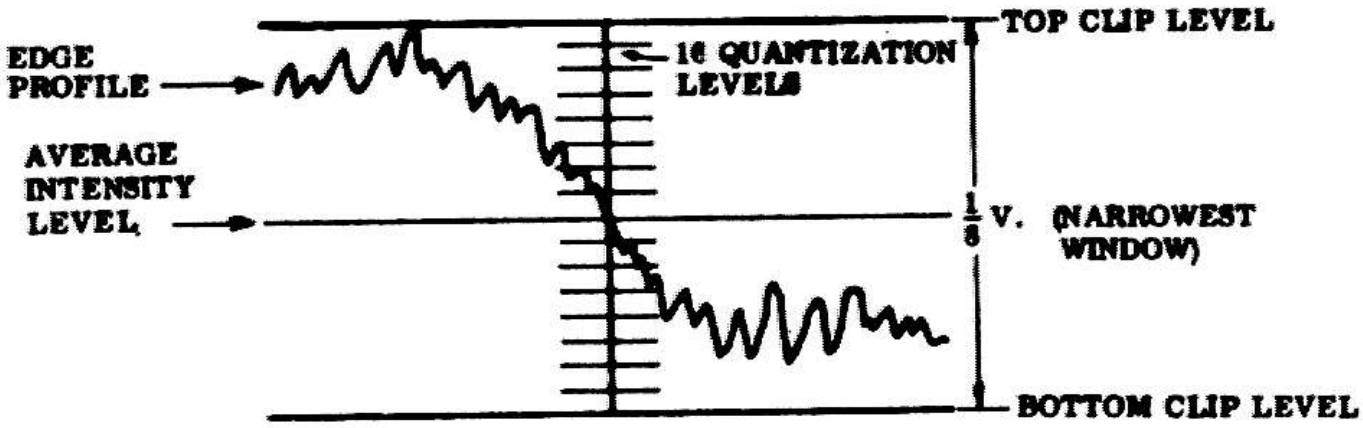


Fig. 3.12 Quantization Window for Edge Detection

to the formal requirements of the quantization theorem. (For high contrast edges, a narrow window will clip intensities, as in Figure 2.6, introducing quantizer noise. However, statistical detection is not necessary when adequate contrast is available.)

After satisfying this sampling condition the remaining quantization uncertainty is treated as ordinary independent additive noise (see Equation 2.34). The overall contrast signal/noise that would be measured at the output of the quantizer can be found by modifying Equation 3.37 to include the transfer impedance of the video amplifier and the residual quantization noise. The transfer gain of the video amplifier is G=3.4*10^(5) v./a. (Equation 2.25). The contrast signal/noise after amplification and quantization can then be expressed by

$$Z_{cq} = \frac{\sqrt{n\tau}\,\rho\, G\, K_i\, L^{\gamma}\, E_T^{\alpha}}{\left[(2-\rho) G^2 A_i L^{\gamma} E_T^{\alpha} + 2G^2 B E_T^{\beta} + 2G^2 C + \frac{2q^2}{12}\right]^{1/2}} \qquad (3.57)$$

where q is the width of a quantization level in volts. Z*cq can be increased at a given target voltage and light level by using a smaller quantization window.

III.4.2.2.7 A SUMMARY EXAMPLE: ACCOMMODATION FOR EDGE DETECTION

The accommodations studied in this chapter have enabled a simple T-test to detect many difficult edges that previously could only be seen by using sophisticated detection algorithms (if at all). Chart 3.4 presents a compilation of the most common problems, encountered in edge detection, with the corresponding accommodations that have proved successful in combatting them.

The combined effectiveness of these remedies will now be demonstrated in an actual case study. The specific problem is to detect the interior edges of a homogeneous, dark cube posed against a light table-top. The table is situated in an ordinary room environment characterized by diffuse, overhead flourescent illumination sources and light beige matte-finished walls.

These conditions produce a moderate directional disparity between the illumination levels incident on the top and side of the cube, but a much smaller variation between the two visible vertical side surfaces. Since the cube is itself homogeneous, it contributes no variations in reflection from face to face. Figure 3.13 shows how this cube appears through the coarse accommodations used for initial acquisition. These include the use of a short lens (1"), wide open clips, and an average sensitivity

Problem	Solution
1. Temporal noise > 1 quantization unit	1. Temporal average
2. Quantizer noise: aliasing (no temporal noise observed)	2. Narrow quantization window
3. No contrast	3. Raise sensitivity, narrow clips, change lens (to seek edge effects)
4. Unsharp edge	4. Focus
5. Saturation limited	5. Color filters
6. Highlight limited	6. Change lens (to narrow field of view), color filter
7. Inconclusive decision	7. Change lens (to get more samples)
8. Spatial noise (dirt, surface gradients, texture)	8. Change lens, spatial average, defocus

Chart 3.4

Overcoming Problems in Edge Detection

with Accommodation

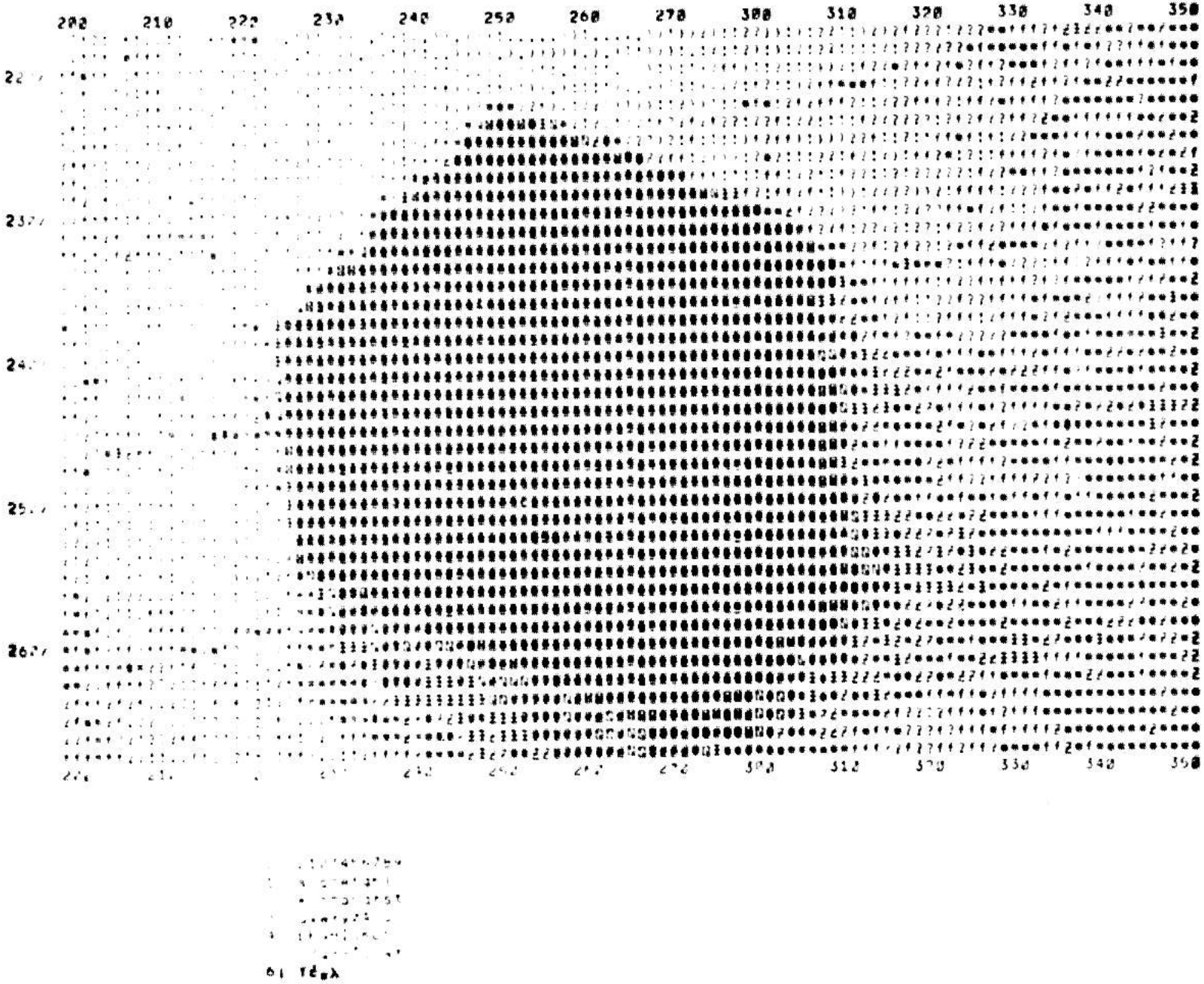


Fig. 3.13 Initial Acquisition of Cube

setting---all intended to maximize the chance of finding objects in an unknown environment. Because of the wide dynamic range in this scene and the low dynamic range of the cube (all dark), the entire body was digitized into the lowest quantization interval.

The most striking problem in Figure 3.13 is the lack of contrast in the interior of the cube. With reference to Chart 3.4, this defect should be attacked by raising sensitivity and reducing the width of the quantization window.

Because of the direct overhead illumination, one expects that the horizontal top of the cube will be the brightest surface. On the basis of this imbalance in top and side lighting, the diagonal boundary separating the top and side faces should be easier to find than the vertical edge between the two sides. Proceeding on this assumption, the sensitivity is raised until the signal from the top face almost saturates. The quantization window is then optimized about the average intensity level between the top and side faces.

Figure 3.14 shows how the cube appears after these accommodations have been made. Both diagonal interior edges are now in plain view (Note that the background has been driven completely into saturation.) The contrast across the vertical interior edge, however, remains inadequate. Since the vertical surfaces are darker and the contrast

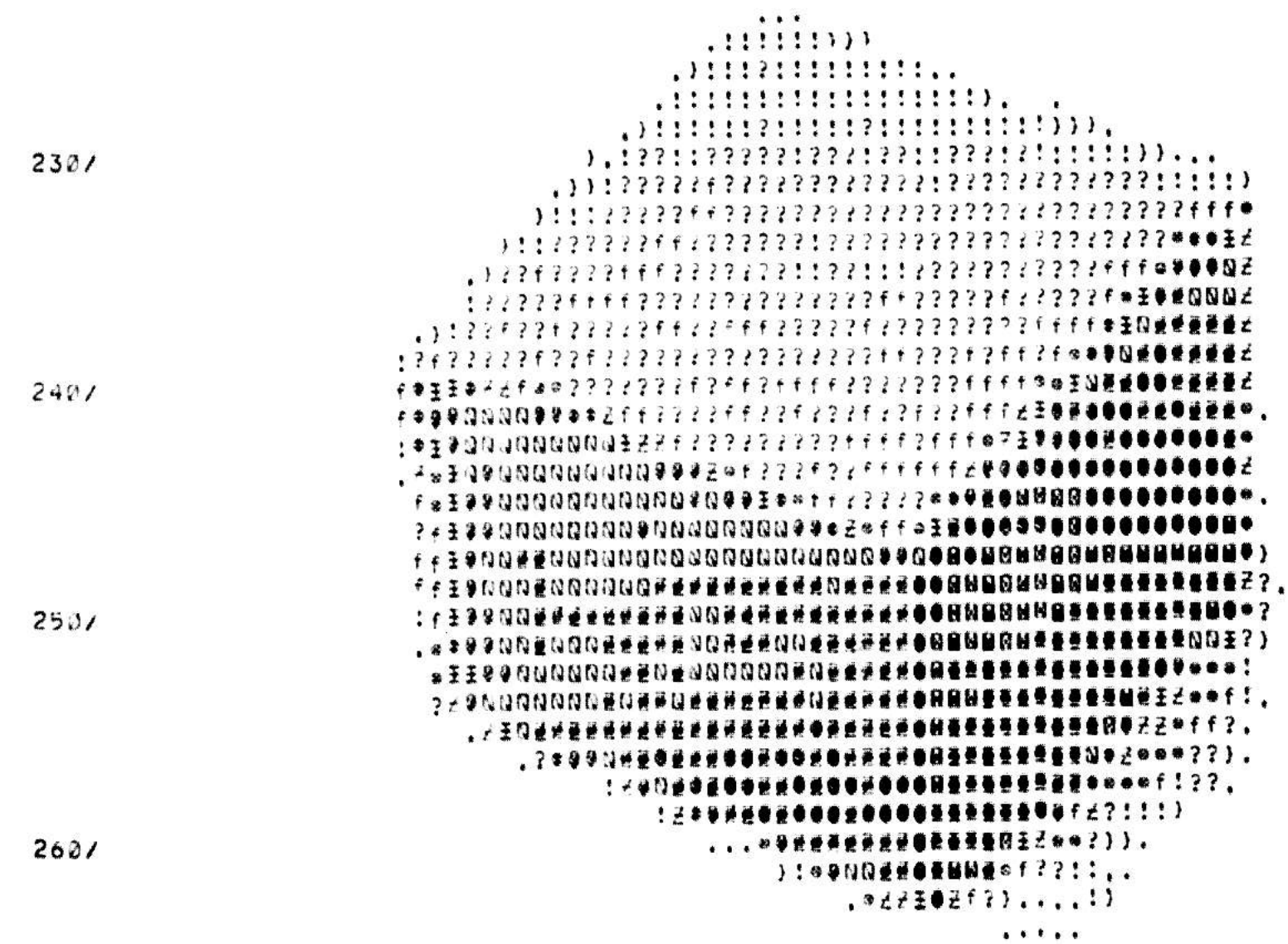


Fig. 3.14 Diagonal Edges Enhanced by Appropriate Intensity Window

between them lower, an even higher sensitivity is needed to detect this edge.

Unfortunately, the relatively high intensity of the background poses the threat of damage to the vidicon if the sensitivity is raised. This situation corresponds to problem 6, "highlight limited", in Chart 3.4. The dynamic range must be somehow compressed in order to extract more detail in the darker part of the scene. The chart suggests that a larger lens be used to narrow the field of view to the immediate vicinity of the edge.

The camera is first centered on the vertical edge. Then the turret is switched to the 3" lens (reducing the original field of view by 1/3). With the threat of damage removed, the accommodations procedures that were successful for the diagonal edges can be repeated.

Figure 3.15 was obtained using a 3" lens, target voltage set to almost saturate the right side of the cube, and a quantization window centered on the range of intensities found in the immediate neighborhood of the vertical edge. Note that the entire image is now filled by the interior region of the cube. (The precise area can be located with respect to the earlier pictures by using the numbered reference axes which are calibrated in raster-sampling units.)

Although this entire region was black in the initial image (Figure 3.13), the sensitivity has since been raised

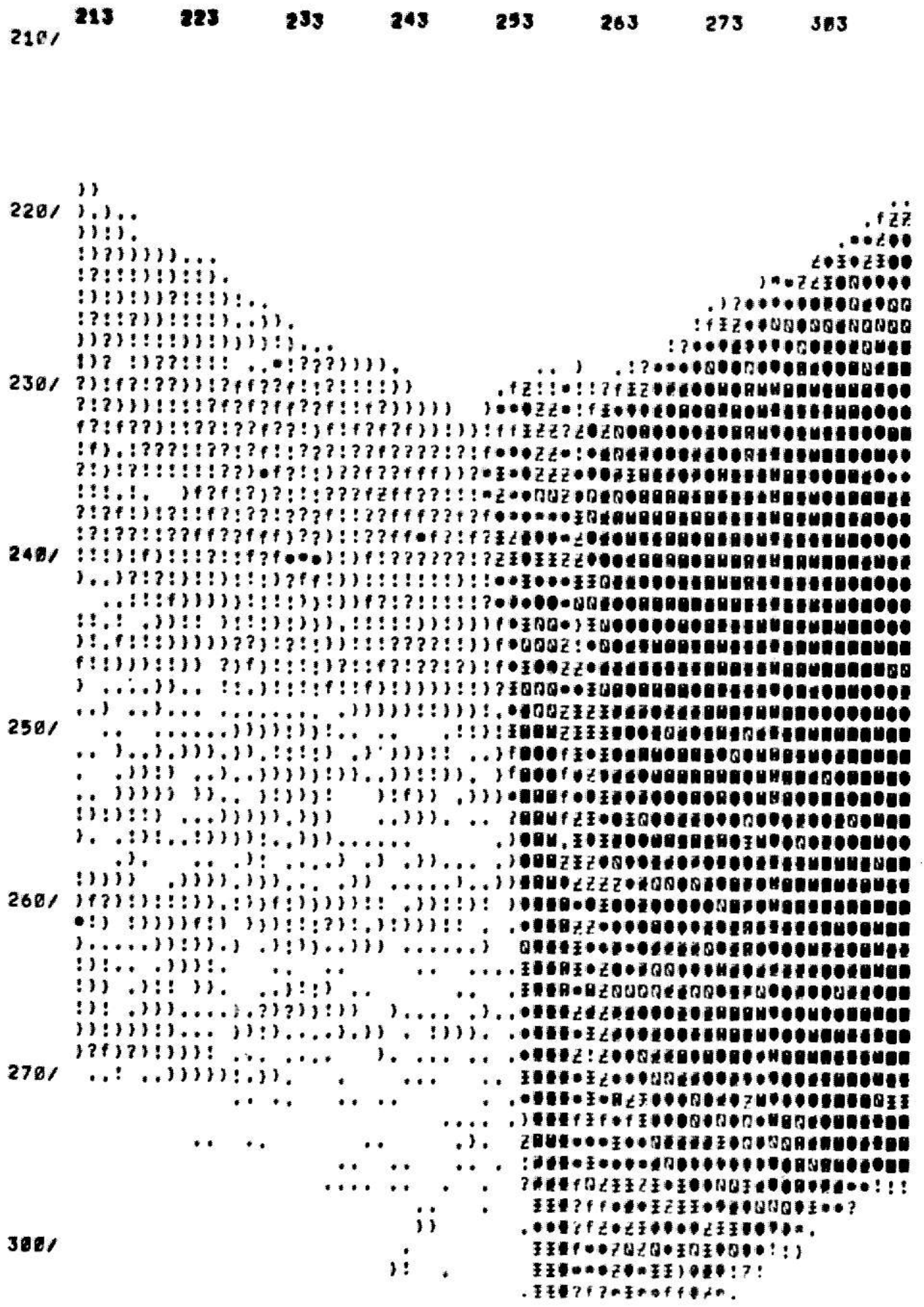


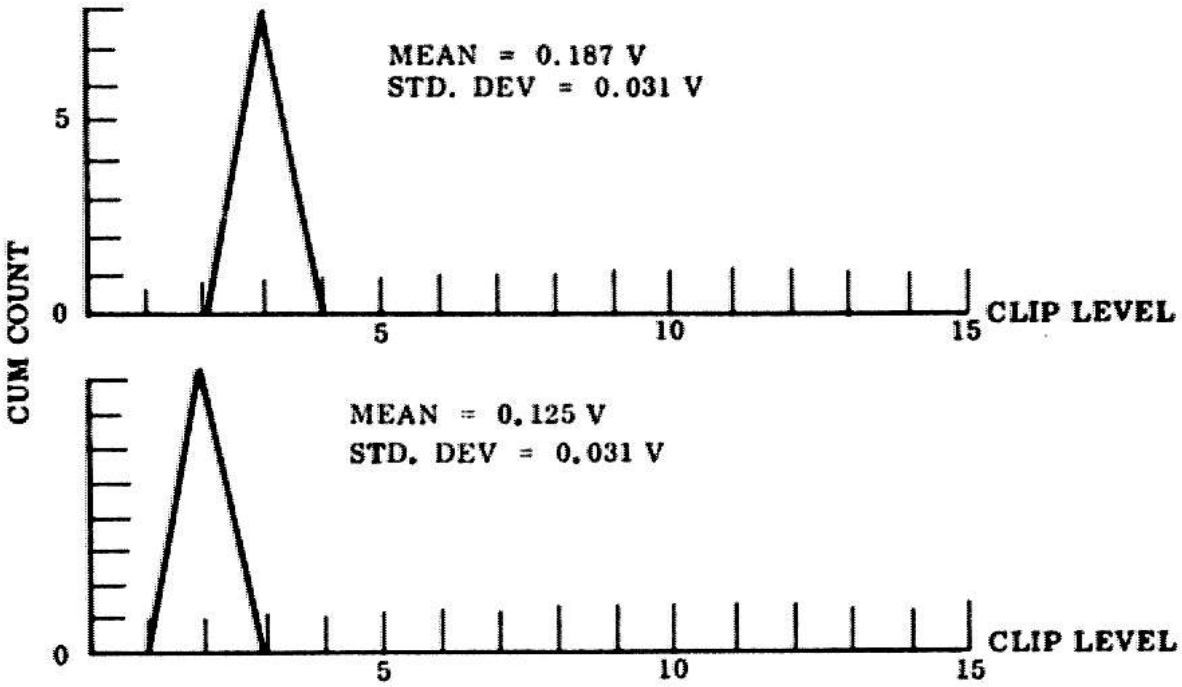
Fig. 3.15 Vertical Edge Viewed with 3" Lens

so high that the relatively brighter top face is completely saturated in the final image. In fact, the top face now actually limits further increase in target voltage.

A Numeric Assessment

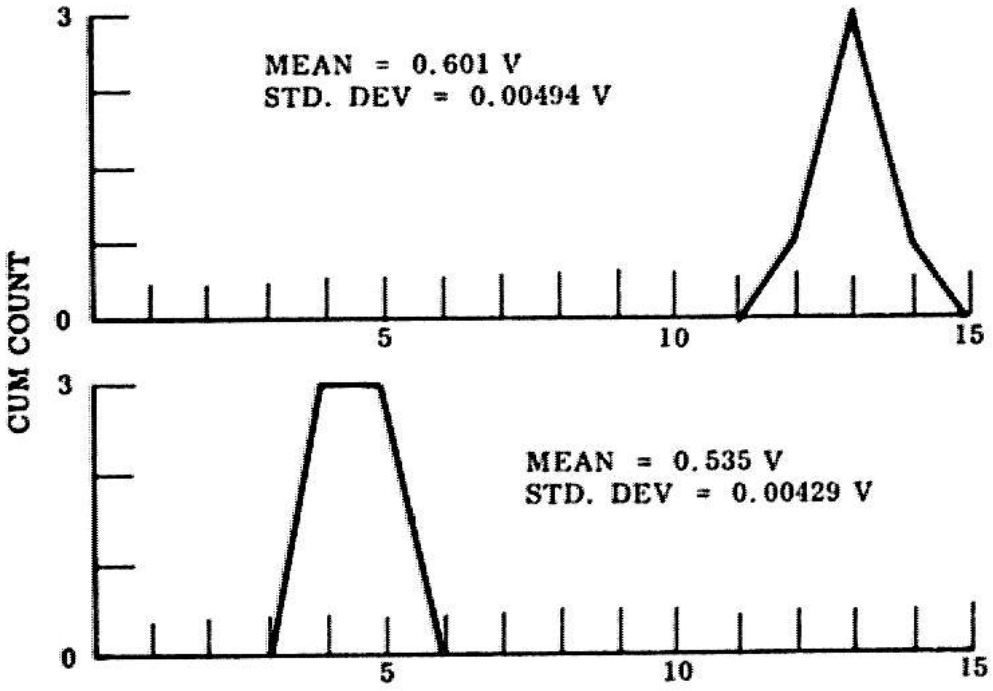
The improvement in the detectability of the vertical edge between images 3.13 and 3.15 is documented in Figures 3.16a,b. These plots show the temporal distribution of intensities observed at 1 raster point on each side of the vertical edge over several successive television frames. The absicissae are calibrated in terms of the 16 quantization levels within the selected window. These are spaced over a 1v. interval in Figure 3.16a and a 1/8v. range in Figure 3.16b, (due to the reduction in quantization window width). The target voltages are 19.94v. and 30.05v respectively. The effectiveness of these accommodations is manifested by the sevenfold increase in the T-score: the corresponding edge confidence has risen from "questionable" to "certain". This score is calculated from the given distributions using Equations 3.9 and 3.10. It expresses the quantized contrast signal/noise given theoretically by Equation 3.57.

In Figure 3.16a, a complication is introduced by the fact that the aliasing requirement is not satisfied. Thus, Equation 3.57 does not strictly apply. A reasonable way to account for the effects of aliasing noise when calculating a



TCLIP = 0, BCLIP = 7, TARGET = 19.94 V, T-SCORE = 1.414

Fig. 3.16a Contrast Signal/Noise Before Accommodation



TCLIP = 3, BCLIP = 3, TARGET = 30.05 V, T-SCORE = 10.159

Fig. 3.16b Contrast Signal/Noise After Accommodation

T-score is to bound the maximum uncertainty. The mean of a signal contained entirely within a single interval can be anywhere within that interval. When samples across an edge fall into adjacent levels (as in Figure 3,16a), they may actually be arbitrarily close to each other, lying just across the quantization threshold, or as far away as two quantization intervals. There is no way to tell. In such cases, we will assume the signal means to be at the mid-point of their respective intervals with an uncertainty of half the width of the interval. The T-score, calculated for the distributions in Figure 3,16a, was based on this assumption. (This bound is invoked whenever the observed distribution of a signal indicates that the aliasing requirement has not been met.)

III,5 CONCLUSION TO CHAPTER 3

In this chapter we have examined some ways in which the primary accommodations affect the performance of our vision system. Chart 3.5 is a summary of the principal roles of each accommodation in the context of three representative perceptual requirements. The appropriate combination of accommodations for any specific application depends, of course, on the particular set of problems that happen to be present. To make effective use of the results summarized in Chart 3,5 the following pre-requisites are also needed:

Accommodations	Edge Enhancement	Level of Detail	Selective Attention
Orientation Pan Tilt	center x,y in field of view exclude irrelevant highlights and detail		
Optical Lens size (x,y,z, texture)	make "edge effects" visible (lines → edges) exclude irrelevant highlights from field of view smooth gradients (reduce spatial noise) allows more samples/area in scene	texture fine detail } vs. { field of view size of object depth of field	
Iris (z,brightness)	s/n (more light)	exclude unwanted textures, objects from field of view	
Color filter (color brightness)	enhance contrast between colors compress dynamic range between bright side of color edge and highlight of scene	enhance contrast of colored texture	enhances or diminishes contrast between selected object colors and background
Focus (depth, texture)	enhances contrast of lines and small features defocus to convert texture to a mean intensity	low-pass filter (attenuate contrast of high frequency texture)	selects depth window selects minimum detail size visible at any contrast
Electrical Sensitivity (brightness)	s/n (limited by highlights and saturation)	bias unwanted texture out of dynamic range	selects dynamic range (intensity window)
Quantizer (brightness)	reduce quantization noise, improve intensity resolution	clip intensity range, eliminating unwanted texture	

Chart 3.5

Accommodation Effects on Typical Perceptual Requirements

1. a definition of the desirable image features for a task (For edge detection, these are concisely stated in terms of the T-score.),

2. a list of problems which represent departures from these desired conditions (as in the left half of Chart 3.4),

3. A corresponding list of accommodations to remedy the problems (the right half of Chart 3.4).

4. diagnostic techniques by which the machine can recognize the existence of these problems in an actual image, and

5. strategies by which the machine, having recognized (or anticipated) a problem, can utilize its knowledge of the task context and the nature of the accommodations to effect an intelligent, cost-effective solution.

In this chapter we have covered the first three requirements to the depth that is practical in a general survey. The last two points have been completely glossed over. Accommodations have only been applied to overcome problems recognized by a human observer. In the remainder of this thesis we shall concentrate on automatic diagnosis and accommodation in the context of specific applications.

CHAPTER IV: EDGE VERIFICATION

IV.1 THE VERIFICATION CONCEPT

The edge verifier is a program designed to find weak edges that a faster, but less sensitive, edge follower has missed. To avoid gross inefficiency, the verifier is used only to confirm the existance of edges whose presence can be reasonably inferred from information already extracted from the scene. For example, when a recognition decision is based on the external contours of an object, the interior edges of that object, as predicted by the machine's internal model, might then be sought to confirm the decision.

Edges may be difficult to see because of

1. low contrast, the result of uniform lighting and homogeneous surface characteristics. An edge follower which tracks local data must discriminate against low contrast edge points to avoid hanging up on noise and minor variations in surface conditions.

2. spotty evidence. An intensity edge may be obscured in places (eg. by dirt). On the other hand, if contrast is low, the most striking evidence of an edge may be the surface defects (eg. paint chips) that tend to collect along the edges of real, planar-faced objects. An edge follower requires continuity of local data and will fail in both cases.

These problems are compounded by the noise inherent in the detection process and by the resolution lost to digitization, spatial sampling, and non-optimal accommodation (eg, unsharp focus).

IV,2 WHAT PEOPLE DO

It is difficult for the computer to detect the presence of the diagonal interior edge in Figure 4.1. Yet at arm's length, the same information usually conveys a clearly distinguishable edge to a human observer. People, of course, are quite familiar with objects, like that depicted in the picture, and, thus, are conditioned to expect this edge. It appears likely that people utilize this expectation in at least three ways:

1. On a local level, noise is combatted by acquiring information from elongated areas, aligned on both sides of the suspected edge. The size of these areas and the integration time vary with the light level and contrast of the edge, affecting a form of accommodation (Buerle [1969]).

2. This local evidence is then accumulated over the length of the contour to build up a global expression of confidence (Hochberg [1964]).

3. The final decision is then biased by the strength of the original expectation (Swets et al. [1961]).

If the expectation is high, the edge may be "seen" even though its presence was only detectable at a few isolated points.

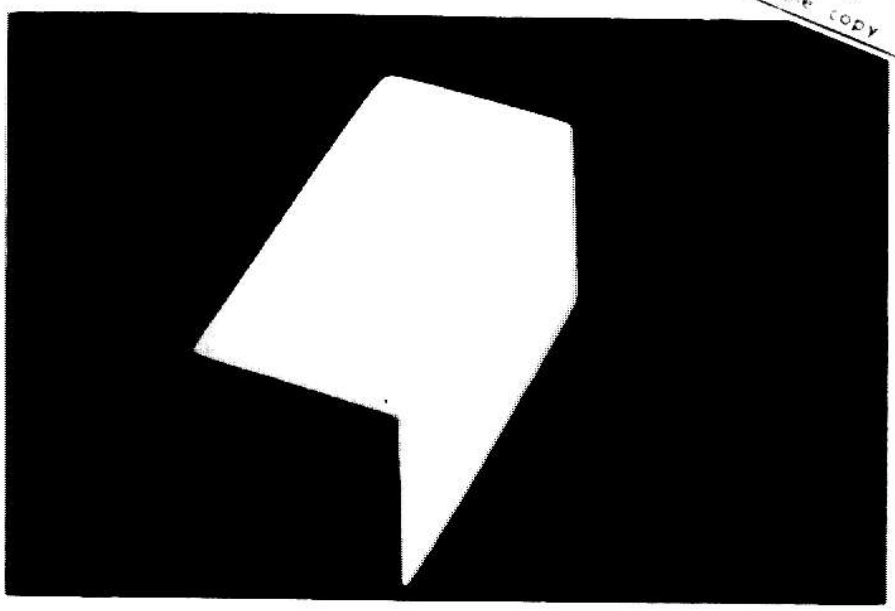
Figure 4.2 provides a convincing demonstration of the importance of global information. The structure of the block has been masked out, leaving only three cross-sectional slices across the interior edge. In none of these slices is there sufficient evidence to suggest that an edge cuts horizontally through them. (This demonstration may be reolicated by the reader, using a cardboard mask at most any point along the edge in Figure 4.1).

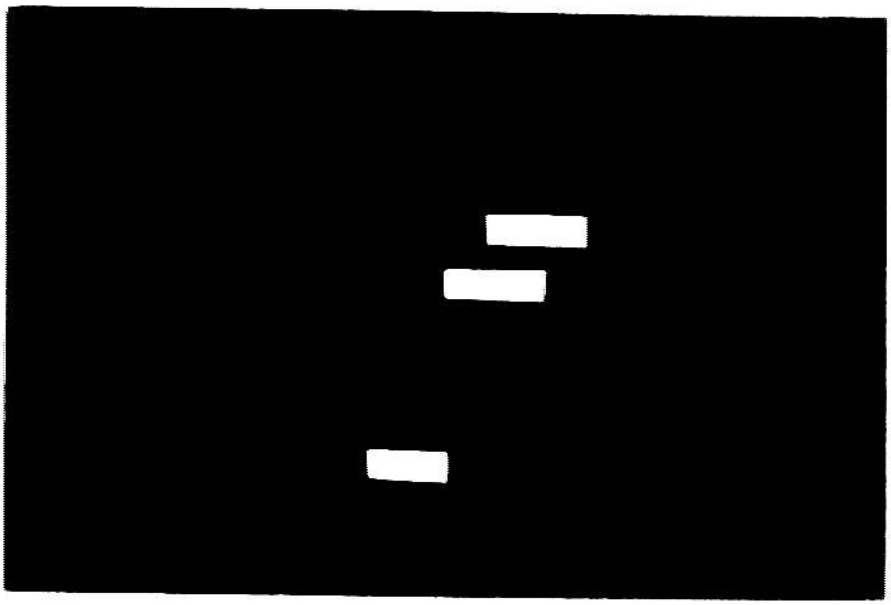
IV.3 VERIFIER TECHNIQUES

The edge verifier succeeds where the edge follower fails by adapting these techniques humans use to capitalize on their global expectations. Specifically:

1. Large operators strategically oriented with respect to the edge improve sensitivity by averaging to extract the underlying noise-free distributions.

2. Accommodation is optimized at each point where these operators are applied to insure the best chance of observing whatever evidence may be present. The camera is tuned to enhance the contrast across the specific boundary where the local edge segment is expected to lie. This procedure illustrates a specific example of how the computer





can use expectation to see what it is looking for.

3. The coincidence of high confidences at various points along an edge is used to augment the statistical significance of local decisions. Global consistency is a powerful filter which allows weak edges to be detected, even when interrupted in places by noise. Strong but isolated surface defects, corresponding to no edge, are rejected as noise. This discrimination is particularly important, since accommodation (technique 2) tends to inject a statistical bias into the local decision of whether an edge is present. (Marginal defects that coincide with the expected edge will usually be detected because of the specificity of the accommodation, applied over that locus.) It is, thus, essential that these biased local decisions be evaluated in the global context of whether they correspond to the feature for which the computer is looking.

The use of large or sophisticated operators and of optimized accommodation is economically practical in verification tasks for three reasons:

1. Only a relatively few points are actually observed.

2. There is a high a priori expectation of success.

3. All effort can be focused on detecting discontinuities in a single expected direction. This approach maximizes the signal/noise that can be realized with a given amount of processing. The edge follower, on

the other hand, must divide its attention among edges at all possible orientations at each place it applies its simple operator.

IV,4 STATISTICAL EDGE DETECTION

VI,4,1 DESCRIPTION OF AN EDGE SIGNAL

We are interested in detecting internal and external surface boundaries of the type normally associated with planar-faced objects. The most common edge type is the simple intensity step treated in Chapter 3. A noiseless profile of this basic function is shown in Figure 4.3. In practice, this ideal signal is commonly corrupted in three ways. Referring to Figure 4.4,:

1. The edge transition is not abrupt. A finite slope is present due to bandwidth limitations in the television system, the discrete nature of spatial sampling, and to the fact that physical edges are never perfect. With a 1" lens, the total transition width for the reasonably sharp edges of plastic blocks at a typical working range of 30" covers about two raster units.

2. Surfaces suffer from illumination gradients due to the directional nature of lighting sources and to variations in reflectivity as the angle of incidence changes along the surface. (An analytic treatment of this effect is

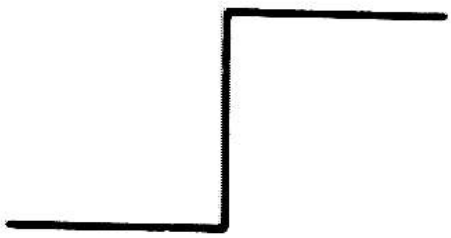


Fig. 4.3 Ideal Intensity Step

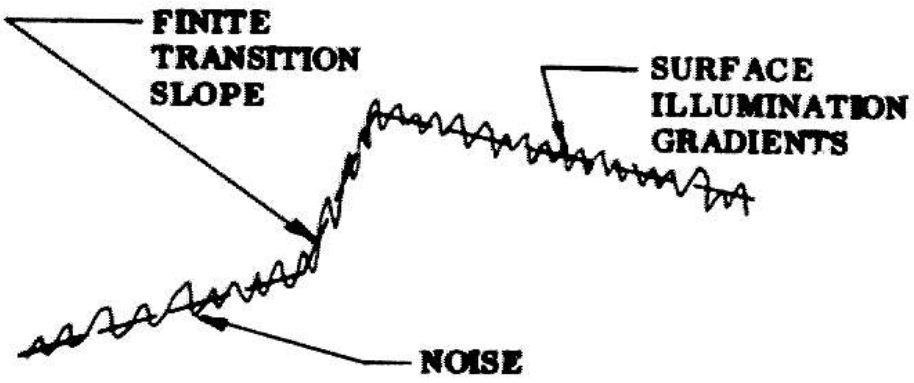


Fig. 4.4 Actual Intensity Edge

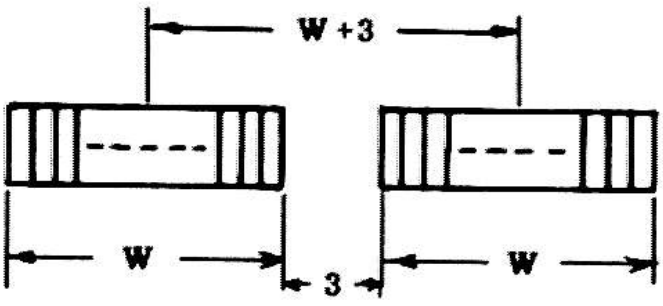


Fig. 4.5 One-Dimensional T-Test Operator

given in Herskowitz [1970].) In most cases, these gradients are a relatively minor problem. The uniform lighting conditions which make the edge difficult to see in the first place also tend to minimize variations in surface lighting. A typical variation is less than .01v. (on a 1v. signal range) in ten raster units, a slope of (10**-3)v./raster unit. (Shadows and reflections will occasionally produce variations up to three times as great.)

3. Random Gaussian detector noise, (see Equation 4.4) is superimposed on these other signal components. For purposes of comparison, the r.m.s. value of this noise is on the order of .01v.

IV.4.2 DETECTION CRITERIA

The statistical T-test discussed at length in Chapter 3, can be generalized to provide a suitable detection criterion for non-ideal edges.

Justification for the Use of T-test

While not the most sophisticated operator that could be applied, the T-test does offer one practical advantage: the effects of the various accommodations on its performance are already understood. Its limitations are known and will be discussed later, with suggestions for more powerful statistical techniques. The emphasis of our present work,

however, is not statistical detection theory but rather to show that, with appropriate accommodation, the simplest statistical test can detect edges that previously had eluded far more elaborate operators.

The T-test serves two specific functions:

1. to discriminate against noisy gradients that do not pass the test of statistical significance and

2. to provide noise reduction by spatial averaging.

The primary purpose of these functions is to guard against errors of commission. Errors of ommission will be attacked by using acommodation to boost the basic signal rather than relying on a more sensitive signal extraction algorithm. Of course, trade-offs between the effort spent on accommodation and on statistical processing must be considered. These trade-offs are largely unanswered questions concerning relative efficiency and reliability. We feel our present emphasis is justified on three counts:

1. The time spent in tuning the camera so that an image is good enough to use with simple operators is apt to be less than the time that would otherwise be required to process a marginal image with more sophisticated algorithms.

2. The goodness of the basic image (as reflected in its signal/noise) inherently determines the reliability of subsequent processing, regardless of sophistication.

3. A considerable body of knowledge is already available concerning the relevant aspects of detection

theory.

IV.4.3 GENERALIZATION OF T-TEST FOR NON-IDEAL EDGES

We first consider the suitability of the specific operator sketched in Figure 4.5, for detecting edge profiles of the type shown in Figure 4.4. This operator is designed to test for the presence of an edge boundary coincident with its three-cell gap. Sample means are calculated from the W intensities observed on each side of the gap. The difference of these averages must be compared with an expected deviation, S•d, to determine the likelihood of an edge.

$$T = \frac{M_1 - M_2}{S_d} \qquad (4.1)$$

(S•d is again the standard deviation of the difference given by Equation 3.9 with N=W.)

Our original use of the T-test to distinguish ideal edges from uniform surfaces was based on the hypothesis that the differences between the two samples, drawn from the same distribution (ie. corresponding to no edge), would be T-distributed with mean zero. Sloping surfaces introduce a systematic error into this assumption: the intensity distributions, sampled at separated points on such a surface, will differ in mean. The appropriate null

hypothesis for use with a surface of slope M is: the difference of two samples drawn from points (W+3) apart will be T-distributed with mean M•(W+3).

A significant T-score is a necessary but not sufficient condition for the existence of a non-ideal edge. The T-score of even the smallest surface slope can be made arbitrarily large by reducing S•d (eg. through temporal averaging). In the last analysis, all that the T-score establishes is the fact that an observed intensity difference is unlikely to have evolved from chance. No judgement is rendered on the significance of the absolute magnitude of the intensity difference.

To distinguish actual edges from sloping surfaces it is thus reasonable to require a minimum absolute contrast. This condition can be imposed on top of the T-test to insure that significant T-scores do, in fact, correspond to significant edges. A convenient threshold for this purpose is twice the height achievable by the maximum allowable surface slope over the effective sampling width, W+3. The complete edge detection criteria can then be expressed as

$$(T > 2.6) \wedge [(M_1 - M_2) > 2 \times 10^{-3} \times (W + 3)] \rightarrow \text{Edge} \qquad (4.2)$$

The auxiliary condition does not radically alter the accommodation rules developed in Chapter 3. It was noted that the T-score could be increased by either enhancing

contrast or reducing noise. Condition 4.2 merely implies that contrast should be the primary concern. If accommodation cannot achieve adequate contrast, there is no need to pursue further noise reduction. If this contrast threshold is realized, then additional effort is warranted to prove that it was not noise-induced.

IV.4.3.1 COMPUTING THE EXPECTED DEVIATION FOR A NON-IDEAL EDGE

S↓d, the standard deviation of the difference in the sample means, is related by Equation 3.9 (with N=W) to the standard deviation of the W intensity samples observed in each half of the operator. For a non-ideal edge, S↓1 and S↓2 will each reflect random detector noise and spatial noise, induced by the uniform slope allowed on each side of the main transition.

The temporal and spatial components are independent and can be considered separately. To define the spatial component, consider a noise-free intensity surface of slope M v./raster unit. The distribution of intensities (about the mean) found in an operator of width W, applied to this surface, is equivalent to that generated by the uniform distribution shown in Figure 4.6. The spatial component of the variance can, thus, be approximated (in closed form) by the variance of a continuous uniform distribution, as

expressed by Equation 4.3,

$$\sigma_g^2 = \frac{2}{M \times W} \int_0^{\frac{M \times W}{2}} x^2 \, dx = \frac{W^2 M^2}{12} \qquad (4.3)$$

The temporal noise of the distribution with the gradients removed is

$$\sigma_N^2 = \frac{1}{\sqrt{r}} \left(G^2 A_f L^{\gamma} E_T^{\alpha} + G^2 B E_T^{\beta} + G^2 C + \frac{q^2}{12} \right) \qquad (4.4)$$

(which is a combination of Equations 2.25, 2.34, and 3.21).

The expected variance of the sample mean of intensities, obtained from an operator of width W, is then

$$\frac{S_1^2}{W} = \sigma^2 = \frac{1}{W} \left(\sigma_N^2 + \sigma_g^2 \right) = \left(\frac{\sigma_N^2}{W} + \frac{WM^2}{12} \right) \qquad (4.5)$$

If comparable noise is assumed on both sides of the edge, Equation 3.9 implies that S•d will be:

$$S_d = \sqrt{2} \left(\frac{\sigma_N^2}{W} + \frac{WM^2}{12} \right)^{1/2} \qquad (4.6)$$

IV,4,3,2 OPTIMIZING THE T-OPERATOR IN THEORY AND IN PRACTICE

The detection sensitivity of Equation 4,2 is limited by its response to a surface of uniform linear gradient, M, in the absence of a real edge. The nature of this limitation is illustrated in Figure 4,7.

The edge signal, h, must be significantly larger than S•d and the systematic bias that would be present in the absence of any discontinuity. A larger operator provides more averaging. Consequently, the random noise will be lower. A large operator, however, implies a wide spacing between operator centers. This spacing results in a large systematic offset from surface slopes which can mask a constant edge signal. For an operator of width, W, the biased T-score from a uniform surface of slope M is

$$T = \frac{M(W+3)}{\sqrt{2}\left(\frac{\sigma_N^2}{W} + \frac{WM^2}{12}\right)^{1/2}} \qquad (4,7)$$

IV,4,3,3 OPTIMIZING W

Since W affects both the operator size and effective spacing in 4,6, a practical compromise is needed. The optimum choice of W is equivalent to specifying the best band pass filter for separating the medium frequency edge

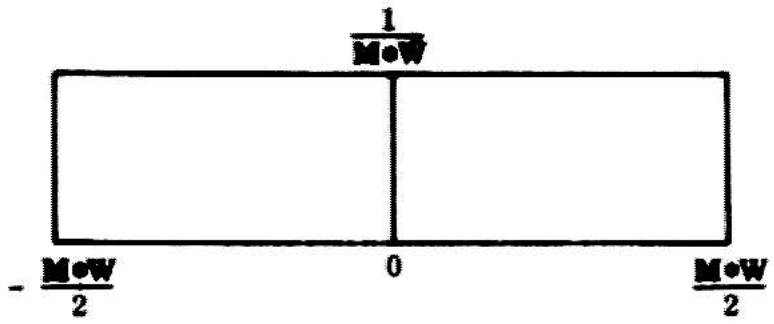


Fig. 4.6 Uniform Distribution Whose Variance is Equivalent to That Induced by a Constant Slope (m) Over an Operator of Width (W)

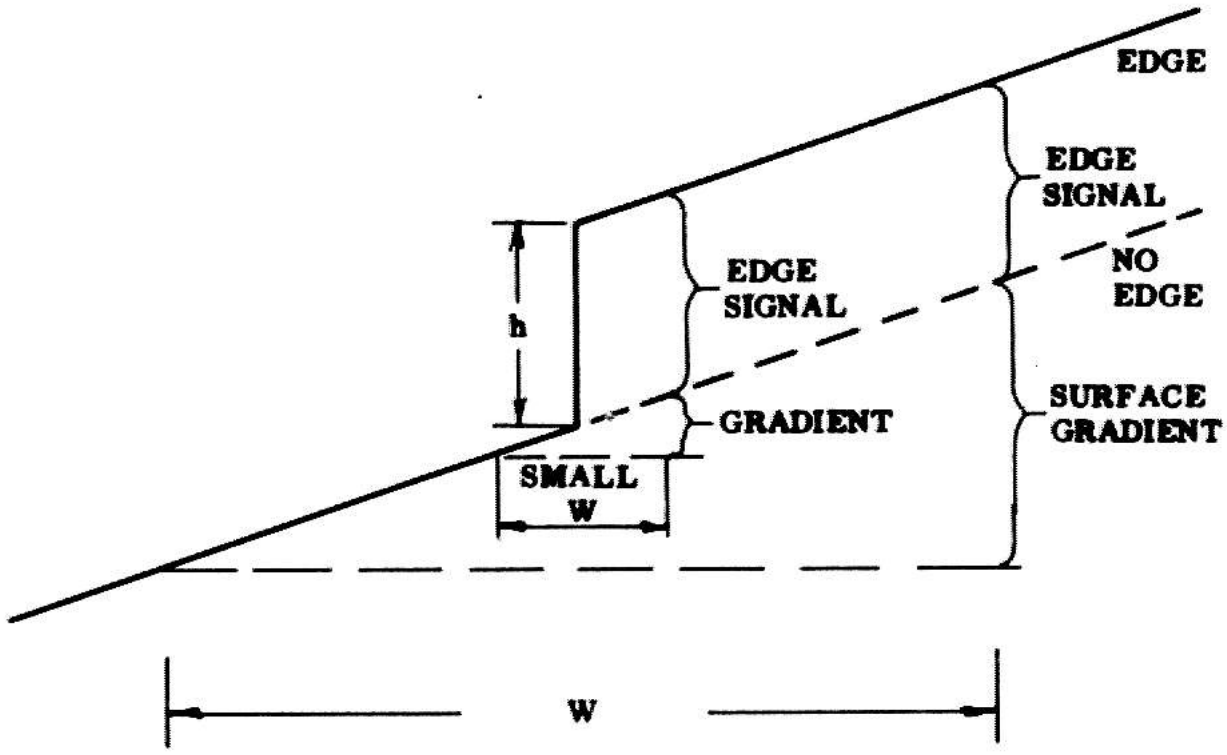


Fig. 4.7 Edge Signal Masked by Surface Gradients

component from the high frequency detector noise and the low frequency spatial noise (gradients). The width of the sampled area determines a high frequency cut-off; a large operator is desired to smooth out temporal noise. The spacing between the operators determines a low frequency cut-off; a small spacing is desired to discriminate against gradual spatial gradients.

While a larger operator will reduce random noise, it will not necessarily lower S/d. This depends on the relative importance of the spatial and temporal components of the noise. In Figure 4.8 Equation 4.6 is plotted as a function of W. For small W the temporal noise is dominant, but, as W increases, this component is smoothed so that the spatial component grows in importance. The overall noise is minimized at W=29.9 when the two components are equal.

Minimizing noise does not, however, maximize sensitivity. An edge discontinuity would need an amplitude of at least 2.33*(10/-3)v. In order to fulfill the contrast required by Equation 4.2. An operator width of 30 would also be impractically wide for all but the largest objects.

IV.4.3.4 DETERMINING A PRACTICAL OPERATOR SIZE

Figure 4.7 illustrates the desirability of confining the operator to the immediate vicinity of the edge transition. If W is kept small, the offset from sloping

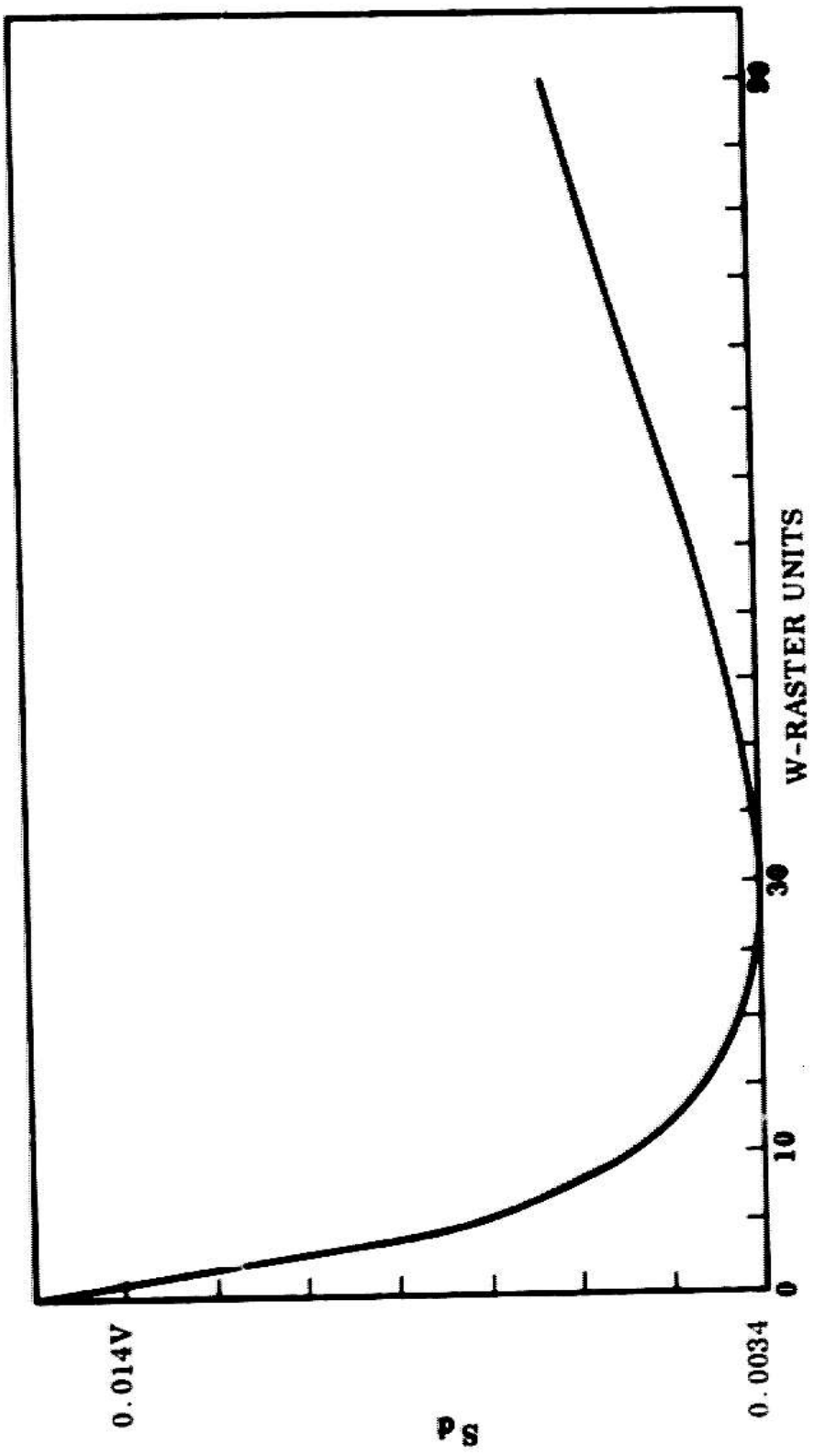


Fig. 4.8 Spatial and Temporal Noise as a Function of Operator Width

surfaces will be less significant relative to a constant edge signal,

For W≤5, the maximum spatial noise component is insignificant relative to the typical temporal noise level of .01v. For W=5, the expected standard deviation of the T-score is thus

$$S_d = \frac{\sqrt{2} \times 0.01}{\sqrt{5}} = 0.00632 \qquad (4.8)$$

If we require that the probability of mistakenly calling a level surface an edge be less than 1%, the minimum detectable edge will have amplitude

$$E_S = 2.6 \times S_d = 0.0169 \text{ volts} \qquad (4.9)$$

This sensitivity is adequate for practical purposes. It corresponds to a contrast that barely exceeds the minimum which our quantizer is capable of resolving. By comparison, the expected signal from the worst case ramp for a center-center operator spacing of 8 will be 8*(10↑-3) (which is 1.4 standard deviations below the minimum acceptable signal). This represents a roughly 8% chance of mistaking a gradual slope for an edge on a single measurement,

If this test is applied ten times over the global extent and a coincidence of eight successful detections is

required to confirm an edge, the possibility that a trend will be called an edge is negligible.

It should be noted that a satisfactory T-score, based on S•d given by 4.8, implies that the contrast constraint (Equation 4.2) is also satisfied. An explicit check of contrast is necessary only when additional averaging is employed to raise the T-score by reducing S•d.

IV.4.3.5 TWO-DIMENSIONAL OPERATORS

Although the width of the T-operator has been constrained, nothing has yet been said about adding a second dimension, parallel to the edge, to create a rectangular operator (Figure 4.9). It is theoretically possible to reduce noise and improve sensitivity by increasing the sample size in this way.

If the surface gradient is resolved into independent horizontal and vertical components, Equation 4.6 can be generalized to two dimensions.

$$S'_d = \sqrt{2}\left(\frac{\sigma_N^2}{WH} + \frac{W M_H^2}{12} + \frac{H M_v^2}{12}\right)^{1/2} \qquad (4.10)$$

W,H are the horizontal and vertical operator dimensions, and M•h, M•v are the corresponding directional surface slopes.

The partials of Equation 4.10 with respect to W and

H can be set equal to zero and solved to yield the optimum shape and size of the operator that minimizes noise for any a priori surface characteristics. The best shape is determined by the combination of horizontal and vertical extent that simultaneously maximizes the number of intensity points (for temporal noise reduction) and minimizes the spatial noise. We find that

$$H_{opt} = \frac{\partial S'_d}{\partial W} = \left(\frac{12\,\sigma_N^2}{W\,M_V^2} \right)^{1/2} \qquad (4.11)$$

and

$$W_{opt} = \frac{\partial S'_d}{\partial W} = \left(\frac{12\,\sigma_N^2}{H\,M_H^2} \right)^{1/2} \qquad (4.12)$$

Substituting W from 4.12 into 4.11 yields

$$H_{opt} = \left[\sqrt{12}\,\sigma_N \left(\frac{M_H}{M_V^2} \right) \right]^{2/3} \qquad (4.13)$$

The horizontal and vertical dimensions of the optimally shaped operator are inversely related to the relative

magnitudes of the horizontal and vertical surface slopes.

Once again, noise minimization cannot be considered out of context. The principal goal is to maximize the limiting sensitivity of the operator to edges. In this respect, a vertical operator has a fundamental advantage; increasing the dimension of the operator, parallel to the edge, does not amplify the signal corresponding to a uniformly sloping surface. By this criterion, the operator should extend exclusively in the vertical direction.

IV.4.2.5.1 PRACTICAL DISADVANTAGES

Four practical considerations dictate the desirability of a small vertical extent (H≤3):

1. Common mode gradients. Figure 4.10 depicts an edge at the boundary of two surfaces with horizontal intensity gradients. The relative significance of a .5 v. discontinuity in a 1 v. signal and a .05 v. step in a .1 v. signal are, of course, identical. This fact would be lost if all of the intensities on each side of the edge were first grouped and averaged before comparison. The in-group variance, contributed by the horizontal gradients, reduces the significance of the T-score, calculated with the grouped data. It is thus preferable to compute an average significance from many T-scores, calculated with individual pairs of intensities along the edge.

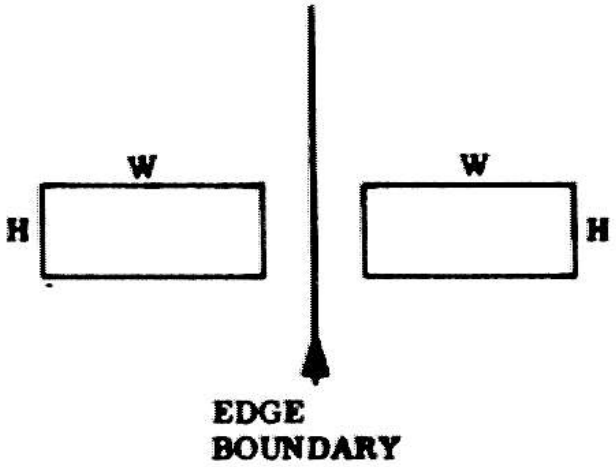


Fig. 4.9 Rectangular Sampling Operator

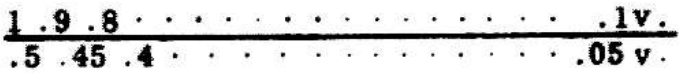


Fig. 4.10 Exaggerated Example of a Common Mode Gradient

a. GRADIENT

b. LINE

c. STEP AND GRADIENT

Fig. 4.11 Other Types of Edges

2. Efficient decision strategy. It is possible to save considerable processing by formulating edge detection as a sequential decision problem. Briefly, the idea is to first look at a small subset of the intensity pairs along an edge. If these pairs indicate an exceptionally strong or non-existent edge, a reasonably confident decision can be made without looking further. Only in marginal cases will more information be needed. Large groupings of data defeat the intent of a sequential strategy.

3. Small anomalies. As Herskovitz [1970] points out, anomalies in otherwise uniform surfaces tend to produce large deviations in intensity over a small area. Such defects can bias the decision of a large linear operator out of all proportion to their size. A preferable approach is to make local edge decisions based on paired intensities at various points along the edge. The global decision is then based on the percentage of local tests that were favorable. The effect of an anomaly on the overall decision is then determined by its extent more than its amplitude. This non-linear technique has demonstrated its superiority in discriminating between occasional surface defects and edges.

A small operator is also at an advantage in detecting valid edges that happen to be characterized by small anomalies, clustered along the discontinuity. These defects tend to be both small and of low intensity. They are consequently lost in large operators but readily

detected in localized pairings. Unlike the isolated surface defect, these "edge effects" are likely to be detected at enough individual points that a globally significant decision can be reached.

4. Optimal local properties. The use of paired data allows individual decisions to be based on the locally strongest property (eg. color, intensity, texture). An edge may be detected most easily by relying on different properties at different points along its extent. The non-linear global evaluation is indifferent to the fact that the local yes-no decisions were based on different features. The significance of a local decision also serves as an immediate criteria by which to achieve locally optimized accommodations.

IV.4.3.5.2 PAIRING VERSUS GROUPING

Statisticians have given much thought to the general question of whether to pair or group data for significance testing (see, for example, J.E.P. Box [1954]). The basic issue is whether statistically significant differences are best detected with fewer, but more reliable, tests (ie. grouping data with larger operators) or by performing more, individually less reliable tests. The consensus is that, when there are valid reasons for pairing, such as those outlined above, it is the preferred alternative.

IV.4.4 ALTERNATIVE OPERATORS

The use of a T-test, applied to intensity averages, has recognized limitations. An average approximates the actual intensity profile (or surface) on each side of a suspected discontinuity, with the best fitting horizontal line (or plane). This approximation clearly breaks down, when the slope of the actual surface becomes appreciable. In these cases, an operator is needed which will normalize out the effects of a uniform gradient.

A second consideration is the existence of other types of edges than those characterized as steps. The profiles of two other common forms are idealized in Figure 4-11a,b. A conventional T-test, applied across either of these discontinuities, would register nothing. Actual edges will often consist of a combination of these basic forms, as, for instance, in Figure 4.11c. In the following section, we discuss several alternative local tests of edge significance which can be used in cases where the simple T-test is not suitable.

IV.4.4.1 REGRESSION

The most direct way to smooth noise on a surface with constant slope is to approximate that surface with the best-fitting plane (in a least square sense). This

approach, known as linear regression, is illustrated for a one-dimensional edge profile in Figure 4.12.

In applying regression analysis, it is assumed that the mean intensity z•m at each point on a line (surface) is linearly related to the independent position variable x (x,y, for a plane). The individual samples of z, observed at each x, will be normally distributed about the corresponding value of z•m with constant variance given by Equation 4.4. The regression line is then the best-fitting straight line that connects the estimated means of these individual distributions (Figure 4.13). Let this line be characterized by

$$\hat{z} = a + bx \qquad (4.14)$$

(z-hat is the best linear approximation to z•m at each x.) a and b are given by the familiar equations of a linear least squares fit (Bennett and Franklin [1954])

$$b = \frac{\sum x_i z_i - \frac{\sum_i x_i \sum_i z_i}{N}}{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{N}} \qquad (4.15a)$$

N = number of abscissa points used in fit

All sums taken over $i = 1 \ldots N$

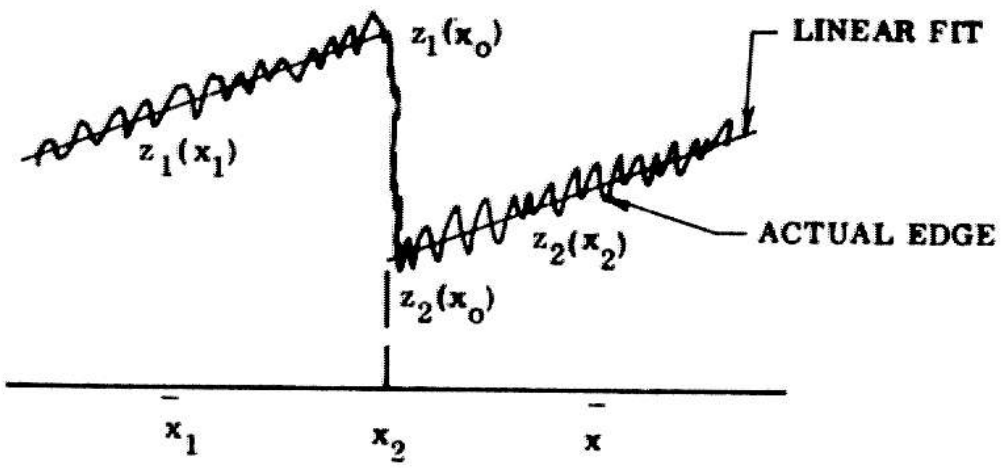


Fig. 4.12 Linear Regression in One Dimension (Applied to the Surfaces on Each Side of an Edge)

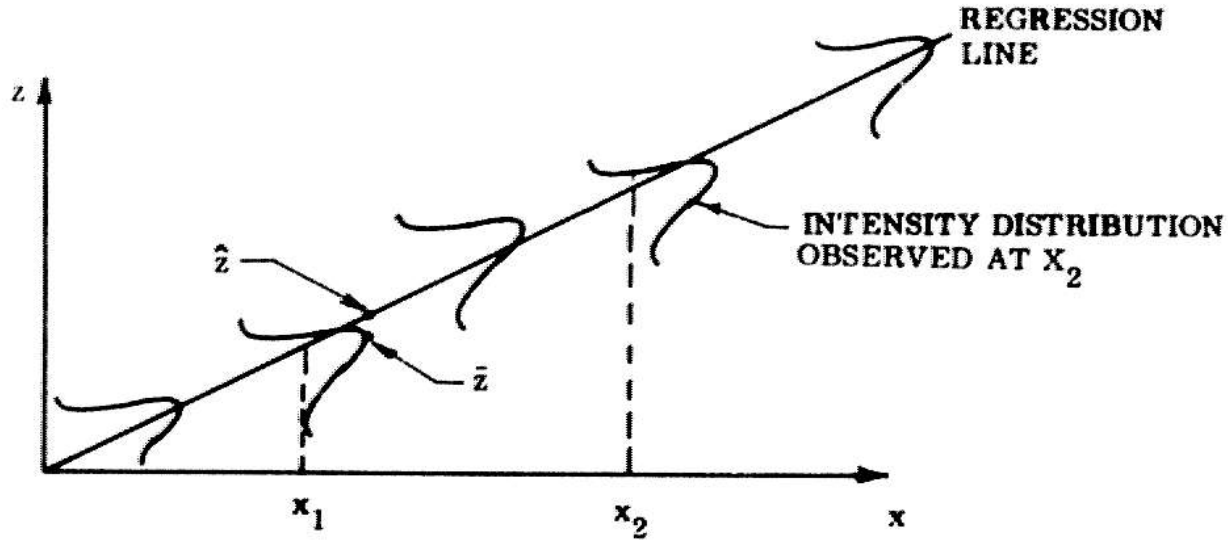


Fig. 4.13 Principle of Regression

$$a = \bar{z} - b\bar{x} = \frac{\sum_i z_i}{N} - b\frac{\sum_i x_i}{N} \qquad (4.15b)$$

a,b are, of course, statistical estimates of the parameters that characterize the actual best-fitting line. In edge detection the fit will usually be based on one sample at each x, in which case

$$\sigma_a^2 = \frac{\sigma_N^2}{N} \qquad (4.16a)$$

$$\sigma_b^2 = \frac{\sigma_N^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \qquad (4.16b)$$

(x-bar is the coordinate at the mid-point of the fitted region.) Since z-hat is a linear function of a and b, its own expected variance at each x can be expressed from Equations 4.14 and 4.16 as

$$\sigma_z^2(x) = \sigma_a^2 + \sigma_b^2 (x - \bar{x})^2 = \sigma_N^2 \left(\frac{1}{N} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \right) \qquad (4.17)$$

Suppose an edge is suspected at coordinate x+o in Figure 4.12. To test this hypothesis, a regression line is fit to the surfaces on both sides of the expected boundary. A buffer region of 2 or 3 points can be left on either side of x+o to absorb any uncertainty in the expectation. The resulting regression coefficients can be interpreted in several ways to establish the significance of an edge. A reasonable test is to extrapolate the line fit on each side to determine the value of z-hat that each predicts at x+o. The significance of any discrepancy can be established in terms of the expected deviations using Equations 4.17 and 3.9

$$T = \frac{\hat{z}_1(x_0) - \hat{z}_2(x_0)}{\sqrt{2}\sigma_N \left[\frac{1}{N} + \frac{x - \bar{x}}{\sum_{i=1}^{N} (x_i - \bar{x})^2}\right]} \qquad (4.18)$$

The advantage of 4.18 over 4.7, for example, is that, when applied to a uniform sloping surface, the expected value of T in 4.18 is zero, Many more points can thus be included in the regression than would be tolerable with a simple average. As a result, the potential limiting sensitivity obtainable with 4.18 is substantially greater than that

possible with 4,7,

Another test is to directly compare the regression coefficients obtained on each side of a presumed edge. The respective values of a,b can be interpreted as the coordinates of two points in a two-dimensional phase space. The distance between these points, normalized by Equations 4,16, is an encompassing measure of the dissimilitude of the corresponding surfaces. This approach can be used to detect gradient type edges (Figure 4.11a) in terms of a significant difference in the b parameter.

Regression analysis can be generalized to two dimensions; the entire surface on each side of the edge can be fit with a plane. On the other hand, this extension is normally not desirable, for the same reasons the original T-test was kept one-dimensional.

IV,4,4,2 SECOND DIFFERENCE

The influence of a constant slope may also be removed by using a one-dimensional discrete approximation to a Laplacian operator (A. Herskowitz [1970]). Figure 4.14 has been adapted from the given reference to illustrate the method. Let L(x) be an intensity profile. D(x) represents the value of a discrete Laplacian taken over a differencing interval 2,

$$D(x) = -2L(x) + L(x + \partial) + L(x - \partial) \qquad (4.19)$$

D is averaged over ∂ points on each side of the suspected edge. The difference of these averages will peak on the boundary of an edge.

$$F(x) = \frac{1}{\partial}\left[\sum_{i=1}^{\partial} D(x + i) - \sum_{i=1}^{\partial} D(x - i)\right] \qquad (4.20)$$

In random noise, D has mean zero and standard deviation of

$$\sigma_D = \left[(2\sigma_N)^2 + \sigma_N^2 + \sigma_N^2\right]^{1/2} = \sqrt{6} \times \sigma_N \qquad (4.21)$$

Thus

$$\sigma_F = \frac{\sqrt{2}\,\partial\,\sqrt{6}\,\sigma_N}{\partial} = 2\sqrt{\frac{3}{\partial}}\,\sigma_N \qquad (4.22)$$

The significance of an edge of height h from Figure 4.14 and Equation 4.22 is

$$T = \frac{h}{\sigma_N}\left(\frac{\partial}{3}\right)^{1/2} \qquad (4.23)$$

Since the expectation of F is zero on a surface of uniform, constant slope, Equation 4.23 suggests that the operator

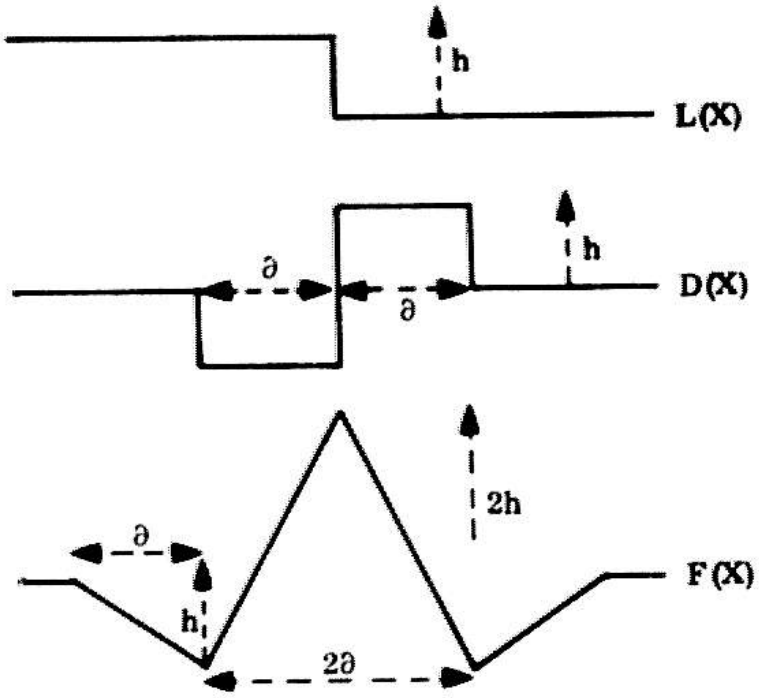


Fig. 4.14 Laplacian Edge Operator

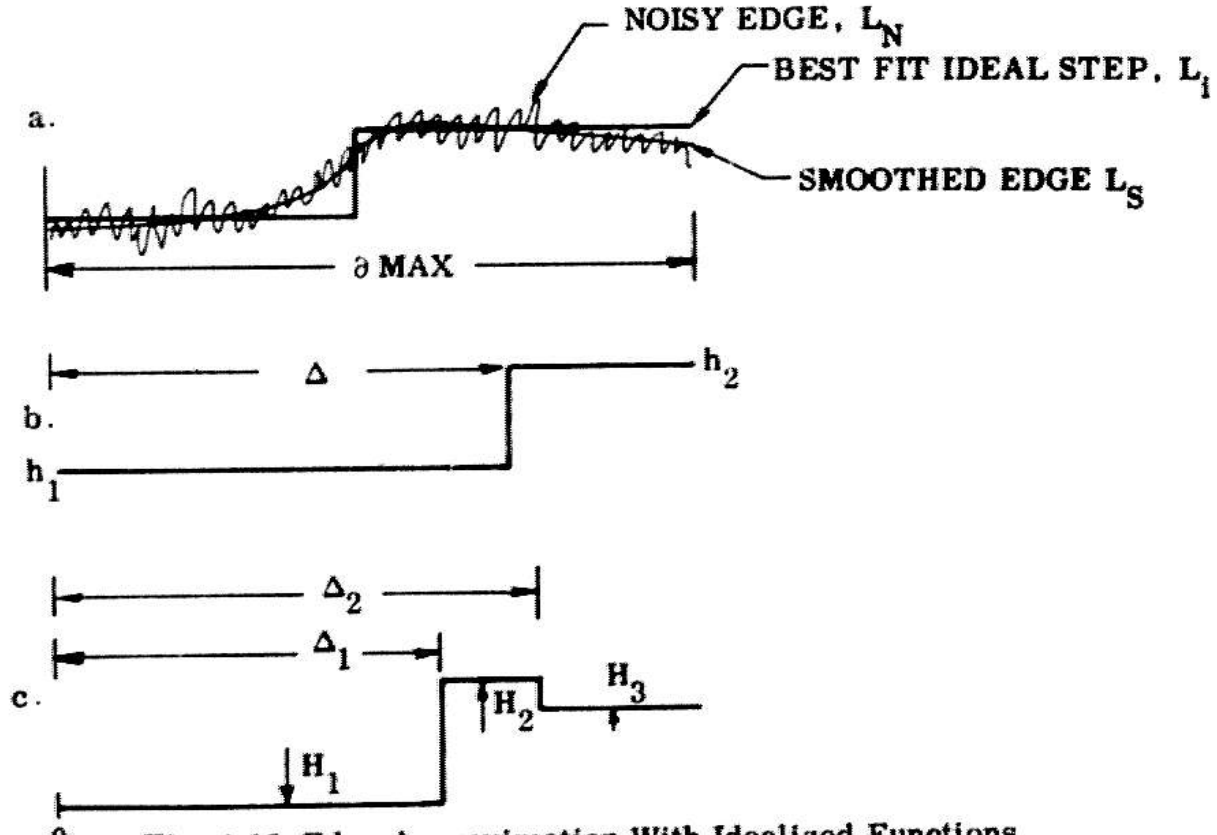


Fig. 4.15 Edge Approximation With Idealized Functions

size, ∂, should be made as large as possible to reduce noise and increase limiting sensitivity. A comparison of Equations 4.22 and 4.17 shows that, for moderate-sized operators, F will have lower noise than a regression estimate.

∂ will be limited in practice by the physical size of an object and the homogeneity of the lighting field. The value of F at a point is determined by the intensities within a range of 2∂ on either side. Using the relation $W = 2\partial$, one can compare the relative signal/noise of F and the T-test for a step of height h. For a given size operator, the T-test has a signal/noise advantage of sqrt (3) over F. However, the T-test is practically limited to a width of 5 whereas $2*\partial$ can in practice be as large as 25. For an ideal step edge, F would then enjoy an advantage of sqrt (5/3) over T. The attractiveness of F grows directly with the maximum allowable surface slope.

IV.4.4.∂ EDGE DETECTION BY APPROXIMATION WITH IDEAL FUNCTIONS

A theoretically appealing approach to edge detection is to approximate an empirically given, noisy intensity distribution with the best-fitting, noise-free intensity step. If the amplitude of the idealized signal is large enough, it is inferred that an edge is actually present in

the original data. Heuckel [1969] developed an effective edge operator based on this principle. A two-dimensional, discrete Fourier analysis is first applied to the intensities found in a small disk (radius ~ 5). The eight low order harmonics are retained to eliminate much of the high frequency, temporal noise (and, inevitably, also some useful information). The smoothed distribution is then approximated with the best ideal intensity step.

An involved mathematical analysis was necessary to efficiently fit all edges that could pass through the disk (at any location and orientation). For verification applications, we can assume that the orientation of the suspected edge is known. The problem is then easier to conceptualize, because fewer degrees of freedom are involved.

In Figure 4.15a the best-fitting ideal step is shown superimposed on a real, noisy edge. The problem is to determine the values of the three independent parameters labeled in the figure. The best fit is defined by that set of parameters minimizing the Hilbert norm between L_I and L_S over the range $\partial < x < \partial + max$ in Figure 4.15a.

$$\epsilon^2(h_1, h_2, \Delta) = \sum_{x=0}^{\Delta} \left[h_1 - L_s(x)\right]^2 + \sum_{x=\Delta}^{\partial_{max}} \left[h_2 - L_s(x)\right]^2 \qquad (4.24)$$

epsilon^2 can be minimized by a straightforward

one-dimensional search over θ; for each value, $1 \leq \theta \leq \theta_{max}-1$, the best horizontal lines are fit over the ranges $0 \leq x \leq \theta$ and $\theta \leq x \leq \theta_{max}$, and epsilon is calculated. (The minimum is unimodal and can be efficiently found.) From Equation 4.15b the best horizontal lines ($b=0$) will lie at the average of the intensity values on each side of $x=\theta$.

The presence of an edge is indicated by the step height, (h_1-h_2) at the θ for which epsilon is smallest and by the goodness of the fit. Theoretically, this approach cannot provide appreciably better sensitivity than a conventional T-test applied to these same average intensities. Its principal advantage is that the fit error is a more selective noise criteria than a blindly calculated standard deviation.

Equation 4.24 can be generalized to handle other edge phenomena besides pure steps. Figure 4.15c, illustrates a possible function with five degrees of freedom, suitable for modeling combinations of step and line-type edges (eg. Figure 4.11c).

Practical experience with Heuckel's operator in the verifier has shown it to be less sensitive in the detection of weak edges than a simple T-test applied to a comparably sized region. It should be added, however, that this operator virtually never committed an error of commission. Because local decisions in a verify task are evaluated for consistancy at a higher level, the Heuckel operator is

thought to be overly discriminating for use in a verifier.

Several additional factors make the Heuckel operator inappropriate to this application:

1. The operator does not presently take advantage of the a priori information that an edge is expected at a particular location and orientation. It thus wastes processing time by considering all possible edge placements. This time could otherwise be spent considering larger samples or temporally averaging the existing intensities to improve sensitivity for the expected orientation.

2. The operator arbitrarily throws away useful information in the pre-filtering stage (see above).

3. The operator is currently sensitive only to step-type edges. This limitation is important, because other edge effects are often crucial in determining whether weak edges will be seen. Heuckel is presently attempting to extend his operator to handle combinations of step and line type discontinuities (see Figures 4.11c, 4.15c).

IV.4.4.4 SOME COMPENSATING ADVANTAGES OF THE T-TEST

We have now considered a number of operators, each of which offers potential advantages over the simple T-test used in the verifier. It is therefore appropriate to reiterate the justification for the choice of the T-test:

1. The T-test appears to offer a very favorable

compromise between effectiveness and simplicity.

2. With small sized operators, the T-test is sensitive to a variety of edge effects, especially lines and anomalies clustered along an edge.

3. No one of the other operators surpasses the effectiveness of the T-test in every respect.

4. Large surface gradients, which pose the principal difficulty, do not appear frequently in practice. When they do, their effects can be largely mitigated by accommodation (see Chapter 3).

The principal reason, however, remains that the emphasis of our work is directed towards demonstrating the advantages of accommodation in a task context. The T-test is particularly appropriate for this purpose because of the results available from the extensive case analysis already conducted in Chapter 3. Furthermore, accommodation has been so successful in enhancing the image of a weak edge that more sophisticated statistical processing was seldom needed.

Optimal accommodation will, of course, improve the performance of any operator. The particular adjustments will necessarily depend on the specific nature of the current test. However, the theory for adjusting sensitivity, clips, and color filter, etc., to optimize contrast signal/noise depends primarily on the characteristics of the scene and the capabilities of the hardware. Furthermore, the actual significance returned by the current operator is

the major criteria used to evaluate the effectiveness of accommodation. This application of performance feedback simplifies operator substitutions.

As an example of this generality, the same accommodation package, designed for the T-test, was used with only a minor modification to optimize the performance of the Heuckel operator when it was incorporated into the verifier.

The basic verification paradigm was designed to accept any local edge operator. The accommodation models, developed in Chapter 2, can be used to rigorously analyze the effects of various camera parameters on each contemplated operator in the manner done for the T-test in Chapter 3 and in more detail in the next section. An ultimate goal is to establish a library of operators, each suited to a particular type of expected and/or empirically evaluated surface characteristic. The choice of the most appropriate one can then be considered as an additional accommodation. This capability has already been realized to a limited extent in the current verifier; the shape of the sampling regions used in the T-test is modified, depending on the class of edge expected.

IV.5 ROLE OF ACCOMMODATION IN VERIFICATION

Accommodation is used in the verifier to maximize

the chance of detecting an edge, when one is actually present. If the significance returned by a local test falls below the required decision threshold, the cause may be that no edge was in fact present, or that the camera was not properly accommodated. Since the verifier was called because an edge was suspected, the latter alternative is a distinct possibility. To minimize the risk of such an error of omission, the accommodation should be optimized for conditions at the specific point where the test was applied.

This strategy reflects an important distinction in the termination criteria used to conclude that an edge segment was or was not present. A significant spread in the means of the intensities, sampled on both sides of the suspected edge, is a clear refutation of the null hypothesis that no edge was present. In the absence of a significant difference, however, one is not equally justified in accepting the null hypothesis, because of the possibility of inadequate accommodation. The termination criterion for deciding against the existence of an edge is thus inherently related to a judgement of the efficacy of further accommodation.

The loop in Figure 4.16 illustrates how a local termination decision is reached in practice. The operator is applied to the suspected edge. If the significance figure it returns exceeds the required threshold, termination with an affirmative response follows

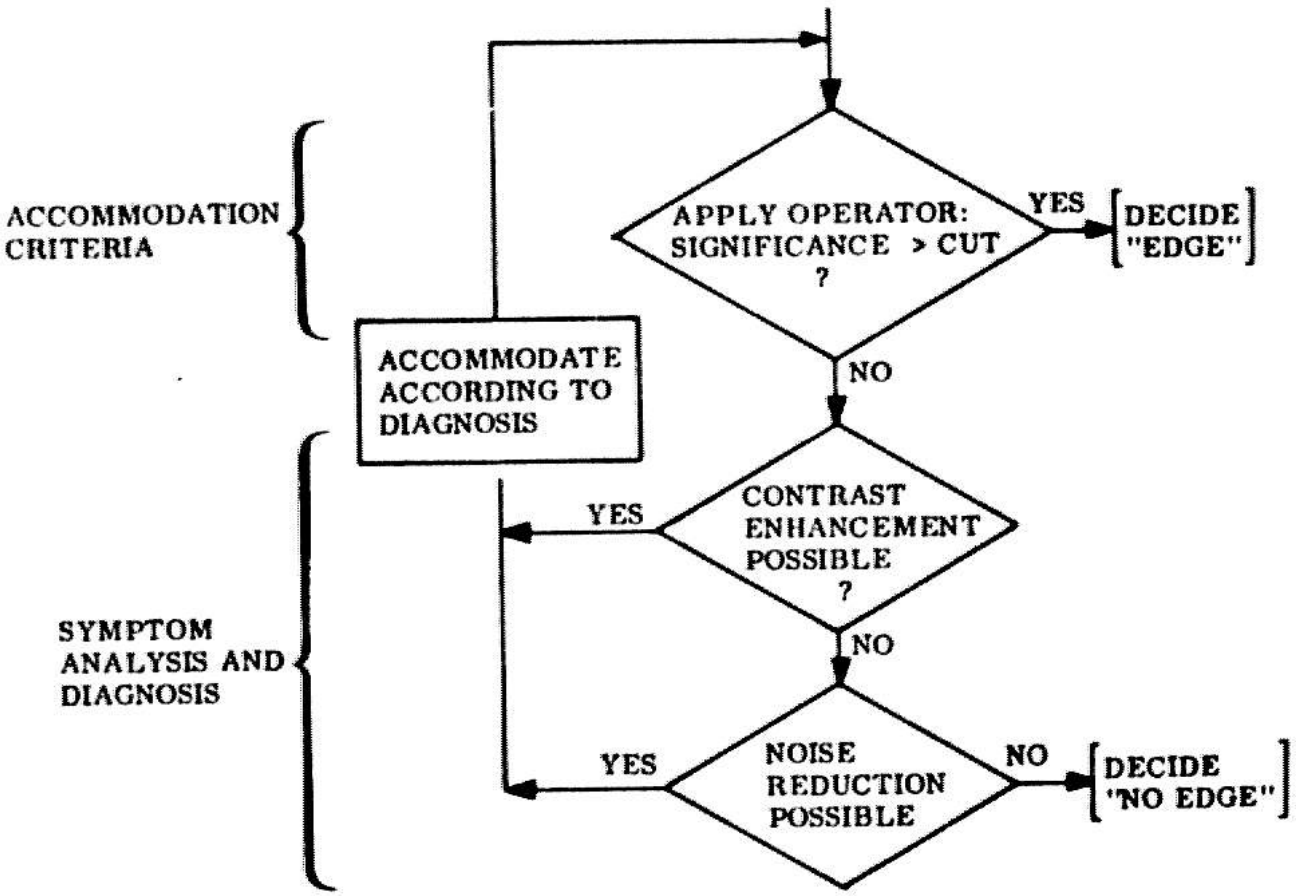


Fig. 4.16 Accommodation Decision Strategy

immediately. If not, the causes for the failure are analyzed to determine if any corrective measures are appropriate. If the problem appears correctable, the indicated accommodation is performed and the loop repeated. Otherwise, a negative response must be issued.

IV,5.1 ROLE OF DIAGNOSIS IN ACCOMMODATION

The effectiveness of this approach to automatic accommodation rests on the ability of the diagnostic procedures to pinpoint the source of low confidence. We have already tabulated remedial effects that the various accommodations have upon the image of an edge. In this section we focus upon the pre-requisite problem of how a machine can determine for itself why an edge cannot be seen. We will describe an accommodation package designed to maximize the local performance of an experimental edge verifier. (The verifier itself is discussed in a subsequent section.)

This package is an attempt to implement the accommodation strategy summarized in Chart 3.4. The implementation is not yet completely automatic. Some of the diagnostic procedures presume system support services that were not available at the time this work was completed. For instance, the ability to determine whether a change of lens would exclude an unwanted highlight from the field of view

requires a metrical calibration of the camera. Sobel is currently perfecting the requisite camera model. For the time being, these aspects of the strategy must be performed manually. However, we will propose suitable ways to automate these functions, based on postulated additions to the system. (Many of the requirements uncovered in this manner are presently being incorporated into the evolving hand-eye system.)

IV.5.1.1 CONSIDERATIONS FOR A DIAGNOSTIC STRATEGY

Diagnosis consists of applying a fixed sequence of predicates to determine from the state of the system and the characteristics of the image the likelihood that a given accommodation would prove effective. Whenever the conditions corresponding to an accommodation are met, the sequence is interrupted. The indicated accommodation is refined and control returns to the evaluation loop. If the cause of a problem cannot be attributed to a specific accommodation, the alternatives must be applied exhaustively.

IV.5.1.1.1 COST EFFECTIVE ORDERING

The diagnostic tests should be applied in an order likely to result in the most cost-effective accommodation,

For example, camera sensitivity and the quantization window must both be appropriately set before any edge can be seen. These accommodations are also relatively cheap to implement. Consequently, it is reasonable to look first for problems associated with these parameters, before more costly and less essential possibilities, like a change of lens, are contemplated.

IV.5.1.1.2 USE OF EXPECTATION

In a system context, specific information will often be available to the verifier that may substantially alter the a priori estimates. The diagnostic procedure can be designed with the flexibility to use a priori and acquired knowledge to influence its priorities. For instance, the verifier may be told by a higher level strategy program that the expected edge belongs to a white block. This fact can be recorded as a constaint on the use of color filters to improve contrast. If, on the other hand, a specific color edge (eg. red-green) was expected, than a color filter (red) should be one of the first accommodations to be tried in the event of insufficient contrast. Similarly, the most effective accommodations at other positions on the edge could affect the order in which action is taken if the same problem is encountered again.

Unfortunately, since no system was available to

provide priorities, the current accommodation package is forced to rely on a fixed, empirically established sequence of tests. It would be grossly inefficient for the diagnostic strategy to ignore the expected colors at a boundary or the expected classification of an edge, were such knowledge on hand. The testing sequence, which we will describe, can be simply modified to reflect specific expectations, that may become available.

IV,5.1,1,3 DEFINITIVE DIAGNOSTICS

We have seen that the default test sequence is ordered by the necessity and costliness of the associated accommodations. This sequence also takes into account the certainty with which the need for a specific accommodation can be established. Thus, another reason why camera sensitivity and quantization are given priority is that the tests which determine the need for these accommodations yield definitive results. (For example, if the intensities on both sides of an edge are clipped out of range low and sensitivity cannot be raised, then the quantization window must be adjusted to a lower range.)

On the other hand, the only way to determine whether another color filter will improve contrast is to try it. Changing a color filter is an expensive operation. Besides the time consumed in mechanically moving the color wheel, it

is also necessary to readjust the sensitivity and clipping levels to match the intensity seen through the new filter. Accordingly, a filter change should not be contemplated until all cheaper attempts to improve contrast have failed.

IV.5.2 CONTROL OF ACCOMMODATION EFFORT

The strategy level of a sophisticated verifier would assign to each local operator a significance cut and a budget allotment. The first quantity represents a projection of how much confidence is needed from each remaining operator to achieve the required global significance of an edge. In our prototype verifier the required significance is always 2.96 (99% confidence). The budget reflects the maximum effort that should be spent in accommodating to attain a significance greater than the assigned cut.

The depth to which the standard sequence of diagnostic tests is pursued will be determined by four factors:

1. the contrast of the edge (The lower the contrast the more accommodation will be needed.),

2. the significance cut (The higher the cut, the more accommodation will be needed for a given edge contrast.),

3. the diagnostic decision (If no more accommodations are deemed suitable, the decision

terminates,), and

4. the budget allotment (When funds are exhausted, the decision terminates.),

IV.5.2.1 COST

In an overall system with conflicting priorities, a budget is necessary to efficiently allocate limited resources such as processing time. The importance of the verifier's decision to the success of a higher strategy program is reflected in the budget it assigns to the verifier. Similarly, the verifier must allocate its assigned funds amongst the individual operators.

The cost constraint can have the practical effect of ruling out expensive accommodations, such as a lens change; the decision may simply not be important enough to expend the time required to turn the turret and refocus. If the most promising accommodation (selected by the diagnostic program) is unaffordable with the remaining capital, testing would continue, as if no decision had yet been reached. Either a cheap enough accommodation will eventually be found or the program will conclude that no appropriate options remain. In the latter case, the normal terminal exit is taken. The cost factor could in fact be used to prune the problems that are considered to that set for which affordable corrections exist.

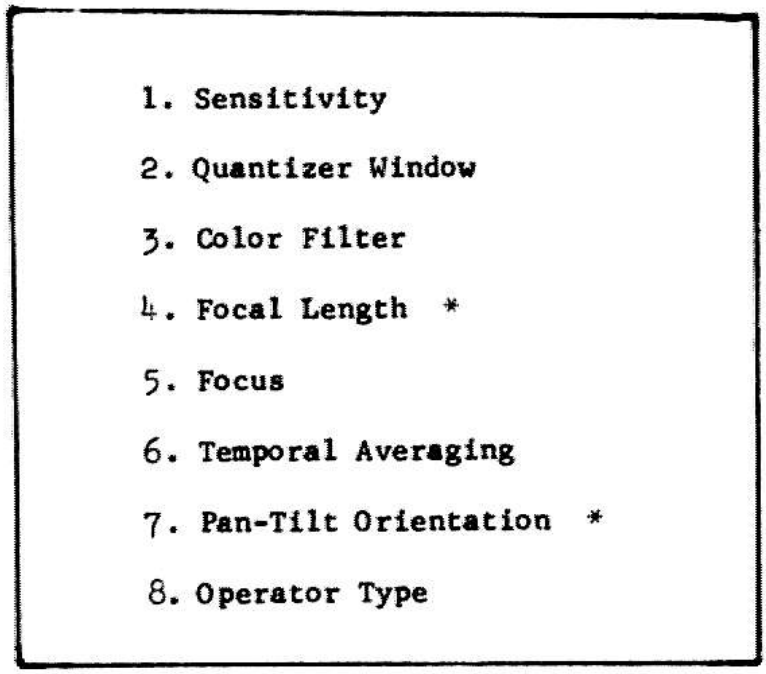
The hand-eye system is still in its infancy. The emphasis of current work is to (reliably) complete a task rather than to complete it within explicit cost constraints. In this vein, we have concentrated on developing the potential of accommodation for verification when cost is not a consideration (ie, every decision is considered crucial to survival!) If the diagnostic program thinks an accommodation has merit, it will be tried without regard to cost.

This approach is admittedly exhaustive. However, it has the merit of demonstrating the full potential of accommodation. (This potential might be tapped, for instance, if some strategy program required hard evidence of a weak, but crucial, edge.)

IV,5.3 ACCOMMODATION IN THE EXPERIMENTAL VERIFIER

Chart 4.1 lists the parameters that are controlled by the accommodation strategy of our prototype verifier. The starred items are manually adjusted in response to commands, issued by the diagnostic program. Only the lens iris is not involved in this strategy. It is always assumed to be as wide as possible in order to maximize signal/noise and minimize depth of field.

Included among the accommodations is the choice of which verification operator to use. Although this parameter is not directly associated with the sensory



*manually implemented

Chart 4.1

Verifier Accommodations

channel, it strongly influences the settings of all of the other parameters, as well as the type of diagnostic criteria employed.

The current verifier utilizes two operators (OP•1,OP•2) illustrated in Figures 4.17b,c (using the coordinate system established in Figure 4.17a). These operators respectively gave the most consistently reliable detection of edge and line discontinuities. The shape and orientation of OP•1 allowed many independent samples of the surface condition on both sides of an edge to be accumulated along a given length. OP•2 was designed to compare anomalies along the actual discontinuity with surface conditions on either side. The selective use of these two operators gave vastly superior performance to what was obtained with either one exclusively or from the 5x5 square operator shown in Figure 4.17d.

IV.5.3.1 CONSTRAINT STATE

The CONSTRAINT STATE (Chart 4.2) is a set of flags which outrightly inhibit further change in corresponding accommodation parameters. In the absence of an explicit cost factor, the constraint set is the sole determiner of whether an indicated accommodation will actually be applied. The constraint state insures that any accommodation that has already been optimized for current conditions will receive

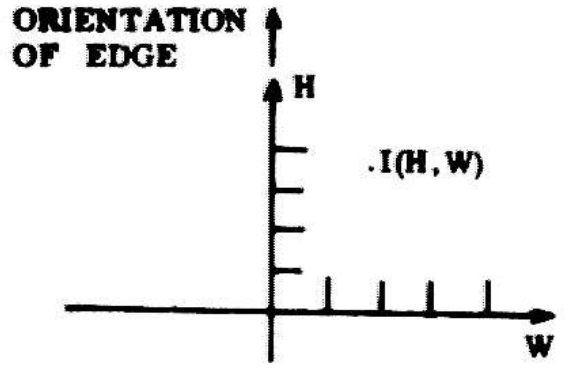
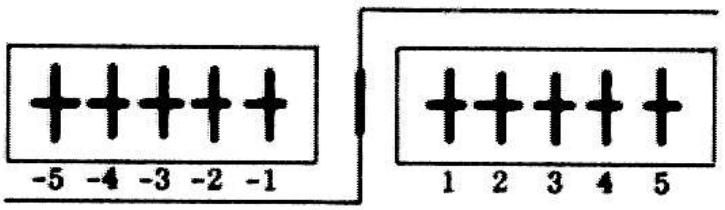
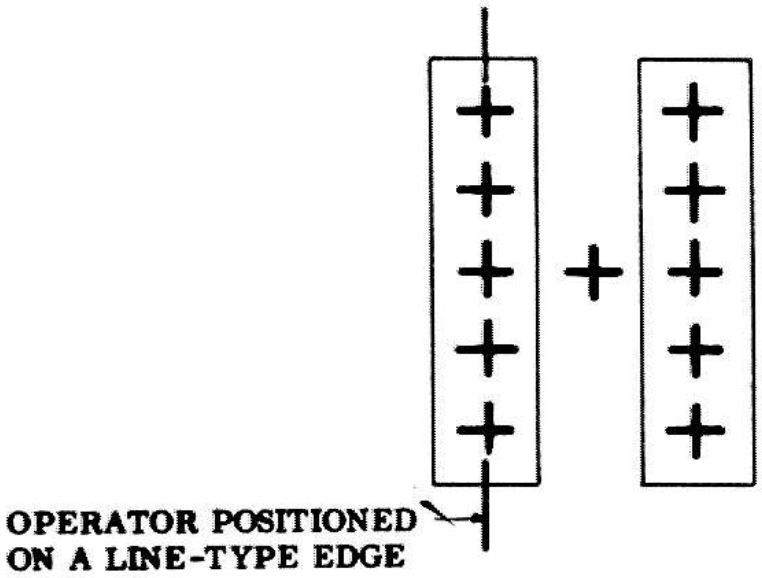


Fig. 4.17a Normalized Operator Coordinates



$$OP_1 = \frac{\frac{1}{5}\left(\sum_{i=1}^{5} I(i,o) - \sum_{i=1}^{5} I(-i,o)\right)}{\frac{1}{\sqrt{5}}\sqrt{\left(\frac{\sum_{i=1}^{5} I^2(i,o)}{4} - \frac{\left(\sum_{i=1}^{5} I(i,o)\right)^2}{5 \cdot 4}\right) + \left(\frac{\sum_{i=1}^{5} I^2(-i,o)}{4} - \frac{\left(\sum_{i=1}^{5} I(-i,o)\right)^2}{5 \cdot 4}\right)}}$$

Fig. 4.17b T-Test for Vertical "Step" Edge



$$OP_2 = \frac{\frac{1}{\sqrt{5}}\left(\sum_{i=-2}^{2} I(1,i) - \sum_{i=-2}^{2} I(-1,i)\right)}{\sqrt{\left(\sum_{i=-2}^{2} \frac{I^2(1,i)}{4} - \frac{\left(\sum_{i=-2}^{2} I(1,i)\right)^2}{5 \cdot 4}\right) + \left(\frac{\sum_{i=-2}^{2} I^2(-1,i)}{4} - \frac{\left(\sum_{i=-2}^{2} I(-1,i)\right)^2}{5 \cdot 4}\right)}}$$

Fig. 4.17c T-Test for Vertical "Line" Edge

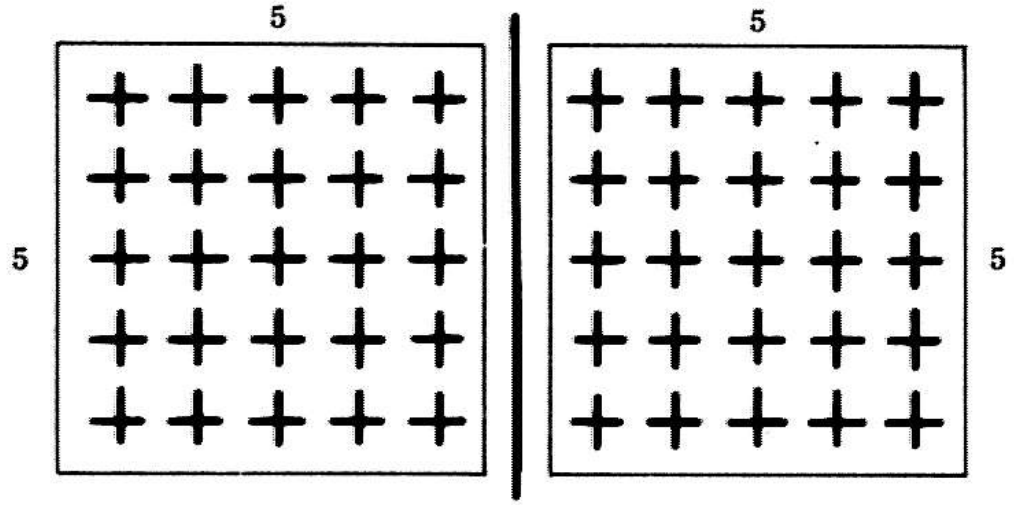


Fig. 4.17d Square T-Test Operator

Accommodation	Constrained / Not Constrained
1. Lens change	
2. Color Filters	
a. red	
b. blue	
c. green	
d. clear	
3. Pan-Tilt	
4. Focus	
5. Sensitivity	
6. Quantization Window	
7. Operator Size	
8. Temporal Average	
9. Utility Flag	

Chart 4.2

Constraint State

no further consideration. For example, it should not be necessary to re-test focus, if the program had just optimized it at a nearby image point. For that matter, it is necessary to avoid re-focusing on a given point, if this was previously attempted but failed to improve edge sharpness.

This use of the constraints allows the sequence of diagnostic tests to be programmed as a very simple decision table. This table is always entered from the top and scanned until a suitable symptom is found. The constraint state insures a NO-MATCH condition on each accommodation already tried to no avail. Thus, on each entry, the diagnosis will progress further through the decision structure, encompassing accommodations that are more costly or of a lower (a priori) expected effectiveness.

The constraint flags also serve important coordination functions. Within the verifier's accommodation strategy, there are internal consistancies that require a common set of interlocks. For example, if the color filter has previously been selected for edge enhancement, it need not be re-considered as a way to subdue an irrelevant highlight. The current filter will already be matched to the bright side of the edge. On a broader scale, an accommodation may be inconsistent with the goals of a global strategy. For example, if another program is also looking at something in the current field of view, it may

not be desirable for the verifier to change a lens at that time. For the present discussion, we will suppress the complicated bookkeeping details required to support these coordination functions.

IV.5.3.2 ORGANIZATION OF DIAGNOSTIC LOOPS

The diagnostic routine is called to determine whether the T-score of an operator, applied at a point on the edge, can be improved by accommodation. The significance of an edge can be enhanced by either improving contrast or reducing noise. The organization of the diagnostic tests, as shown in Figure 4.16, is partitioned along these lines. This division reflects two factors. First, adequate contrast is a necessary condition (by Equation 4.2). Second, with the exception of quantization noise (which, for convenience, is categorized here as a contrast enhancement), the noise accommodations tend to involve higher cost and lower probability of success.

Noise-reducing accommodations either implement or facilitate some form of temporal or spatial averaging. Averaging has a more predictable, but less dramatic, effect on significance than improving the actual signal strength. Thus, although noise tests are definitive, they are applied only after all steps to optimize contrast have been taken.

The diagnostic process is outlined in detail in

Figure 4.18. Under control of the constraint state, tests are applied to the image to determine whether given accommodations are deficient. When a test yields a positive conclusion, an associated series of accommodations, indicated with square brackets, is performed. The diagnostic routine returns "TRUE", signifying that an accommodation has been enacted and that the local operator should be re-applied.

If edge signficance is still inadequate after completing these definitive tests, several additional accommodations are tried exhaustively. With this latter set, it is not possible to anticipate the likelihood of success. If significance is still inadequate, when all avenues have been exhausted, the diagnostic routine returns a "FALSE" decision. The verifier will then concentrate on establishing significance at another position along the edge.

IV.5.3.3 ROLE OF CONSTRAINT SET IN CONTROLLING ACCOMMODATION

The constraint set is managed, in the interests of economy, by heuristic criteria, intended to prevent unnecessary repetition of expensive accommodations. The policy is:

1. All constraints are initially removed so that all accommodations can be tried at least once.

2. Constraints, c#5,c#6, are always reset before an

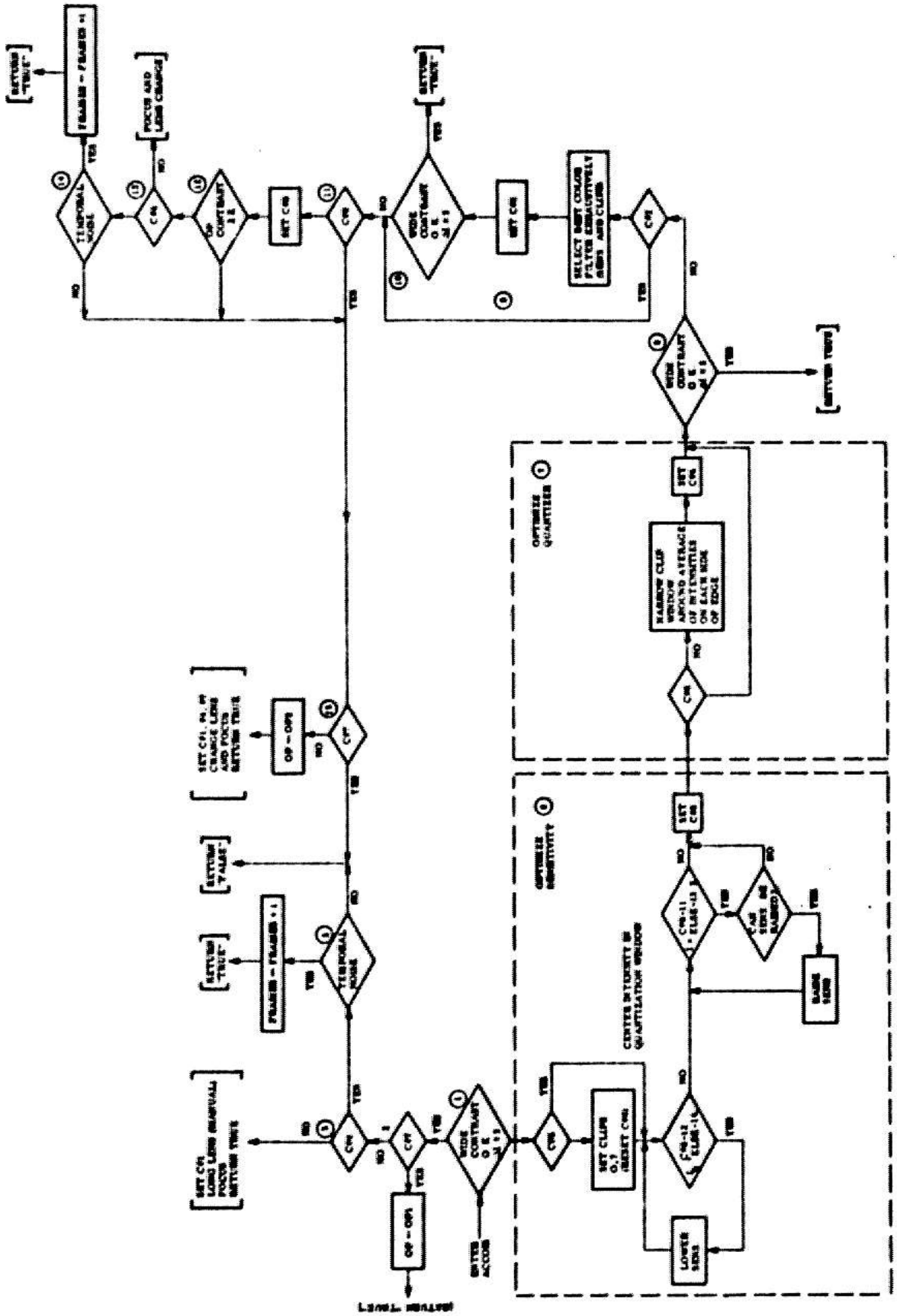


Fig. 4.18 Diagnosis and Accommodation Strategy for Experimental Verifier

operator is applied at a new point along the edge. Appropriate sensitivity and quantization range are "must" accommodations; if either is wrong, edge contrast will be severely compromised. Furthermore, these accommodations depend on intensity level, the characteristic most likely to vary along an edge. Sensitivity and clips are also among the least costly adjustments.

3. Constraints c#1, c#2, are never reset during a run. These restrictions rest on the notion that a long lens is universally beneficial and that in the hand-eye environment color edges are uniform along their extent. Once a long lens has been fitted and color contrast has been optimized, these selections remain in force.

4. Constraint, c#4, is reset when the scan has progressed 75 raster units from when focus was last optimized. While focus is a very expensive accommodation, the conditions when it is needed are clearly indicated (acceptable contrast measured at widely separated points across the edge but no distinct boundary). It is not necessary to re-focus, unless an edge is long enough to extend beyond the depth of field at the point where focus was last adjusted. A practical worst case is provided by an edge of marginal contrast, sloping away from the camera at a median range of 30", viewed with a 3" lens (biggest image, shortest depth of field). The 75 raster unit reset interval is based on this case.

5. Constraint c#3 is never set. There is currently no provision for automatically centering the camera, although the requisite theory is now available (Sobel [1970]). It is assumed that the edge is manually centered in the field of view before the verifier is called and after a change of lens. Additional pan-tilt adjustments will be required only when an exceptionally long edge extends beyond the original field of view. This condition is detected by the routine that accesses the television buffer and will result in a request for manual re-centering to be output on the computer display.

6. Constraint c#8 is set when the number of frames, averaged for each measurement, reaches 32. When the noise level is crucial (because of marginal contrast), the frame count should ordinarily be increased as long as an unacceptable temporal variation in contrast persists. The limit of 32 frames is based on the observed short term stability of the camera electronics. Averaging more frames tends to actually increase the sample variance.

7. c#7 is set when the "step" edge operator (OP•1) is replaced by the "line" edge operator (OP•2) for lack of contrast across the edge. The constraint remains in force, locking in OP•2, until a point on the edge is reached where $OP\bullet_1$ would again be effective. At this time c#7 is reset and OP•1 returned to service.

8. c#9 is used as a special flag to indicate when

all contrast enhancing accommodations have been exhausted for a particular point on the edge. It is reset before moving to the next scan position.

IV.5.3.4 DETAILED DISCUSSION OF AUTOMATIC DIAGNOSIS

The accommodation strategy detailed in Figure 4.18 is considerably less elaborate than might be manually carried out using Chart 3.4. Some simplifications were introduced for practicality.

Several of the noise accommodations in Chart 3.4 were included mainly for conceptual generality. Substantial spatial gradients are, for instance, so rarely encountered that it was not deemed necessary to actually implement a test for this contingency. On the other hand, quantization noise is so basic a problem that it is treated as a matter of course, whenever contrast is low.

Other simplifications were introduced out of necessity. Many of the diagnostics, suggested by Chart 3.4, could not be automated without system support considerably more sophisticated than that currently available. For example, none of the questions regarding the nature of an external highlight can be answered out of context. The machine cannot now decide for itself whether to remove a limiting highlight with a color filter or with a change of lens. Sensitivity in such cases ($I_{HI} < 13.5$ and $E_{dam} < E_{sat}$)

is simply raised as high as the autotarget circuit will allow.

We will suggest ways to overcome these limitations of the present strategy by postulating appropriate features to be added to the system. The discussion that follows is keyed to Figure 4.18.

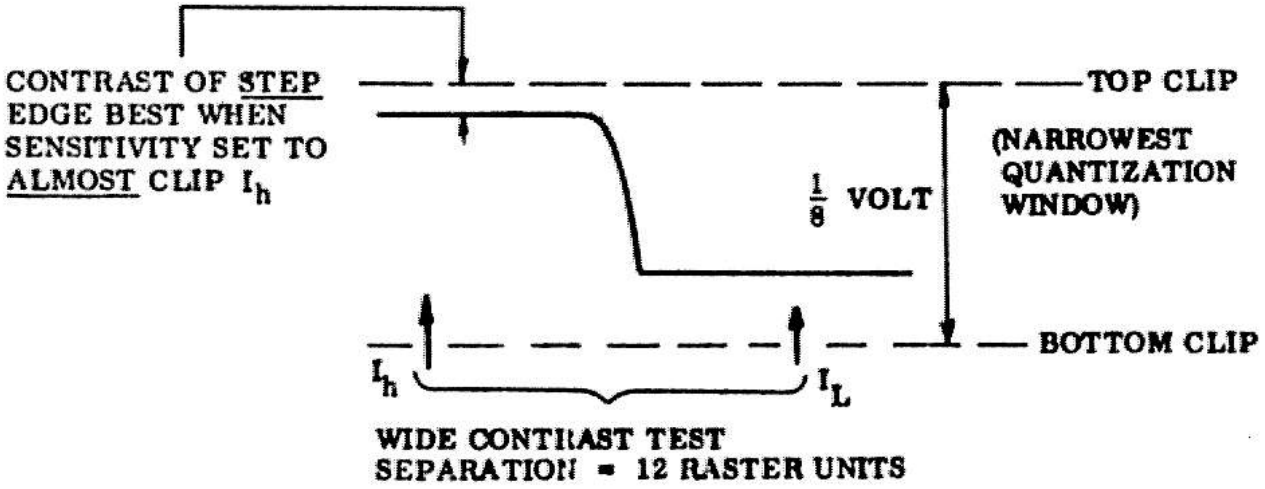
IV.5.3.4.1 HIGHLIGHTED DETAILS OF EXPERIMENTAL DIAGNOSTIC PROCEDURES

[1] The first thing that is checked when significance is low is the WIDE-CONTRAST. Wide-contrast refers to the contrast measured at widely separated points across the edge (see Figure 4.19).

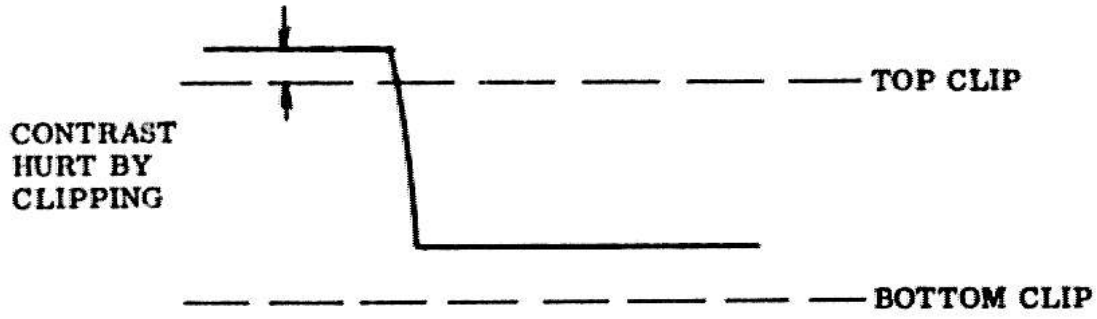
[2] If this contrast is large, the step type operator is reinstated, if it was not currently in use.

[3] High wide-contrast with low edge significance is often an indication of poor focus (see Chapter 6). If focus has not yet been optimized near the current point, it is done now. (Focusing is a time-consuming accommodation. It should not normally be repeated more often than every 50-75 raster points along a suspected edge.) The observed contrast insures that a well-defined focus peak will be found.

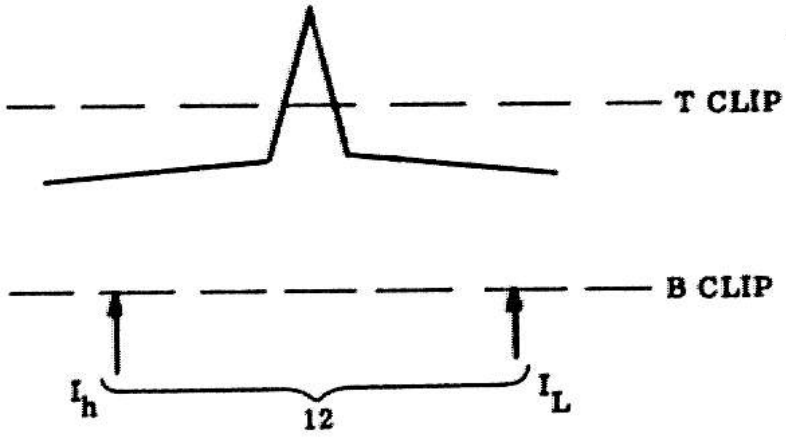
[4] This opportunity is also taken to fit the 3" lens, if this has not already been done. This action is taken, because sharp focus is more easily attained with a



a. CLIPS, SENSITIVITY TO DETECT LOW CONTRAST "STEP" DISCONTINUITY



b. CONTRAST CLIPPING DUE TO DISCRETENESS OF QUANTIZATION RANGES



c. COMPRESSION OF A HIGHLIGHT ANOMALY WHEN ACCOMMODATED FOR A STEP-EDGE

Fig. 4.19 Accommodation Criteria for Experimental Verifier

long lens. Furthermore, the magnification, as we have seen, has generally desirable effects. Since focus is expensive and required for a lens change anyway, it is wise to effect the change at this time. (The lens change is accomplished manually. Since no camera calibration was available, the camera must also be manually re-centered, if necessary, on the current edge point. Moreover, the scan locus must be manually redefined (using the cursor and pot box described later.)

[5] If focus is satisfactory, the remaining alternative is an excessive amount of temporal noise.

A significant level of temporal noise is assumed whenever the contrast component (ie. numerator) of the T-score varies appreciably (eg. by more than 15%) over several measurements. Temporal noise (from the detector) can be reduced either by using larger operators (to obtain a larger sample) or by averaging the intensities, observed over time at each point in the operator, before calculating the T-score. Since a vidicon camera records an entire frame at once but has a storage time overhead of 1/30 second/frame, a spatial average enjoys a considerable efficiency advantage.

Our analysis of optimal operator sizes, however, has shown that the effectiveness with which spatial averaging can be used to overcome temporal noise depends on the steepness of the intensity surface. Whenever this slope

exceeds the nominal rate quoted earlier (which could happen, for example, if other edges are in the vicinity), then conventional time averaging must be used. (The relative importance of spatial and temporal noise can be determined by comparing the standard deviation of intensities sampled over time from a single raster point with the same statistic, obtained with 1 sample taken from each of several adjacent raster points. If the former statistic is significantly less, the noise is primarily spatial. Otherwise, temporal noise dominates.)

The verifier currently employs the following noise reduction strategy. If temporal noise is suspected, the number of television frames averaged for each measurement is increased by 1. If spatial noise (from extraneous edges) is detected, the remedy is to switch to the line-oriented operator.

If these accommodations fail to improve edge confidence, given suitable contrast, the diagnostic sequence currently gives up. However, the program does not yet cope with unwanted textures and isolated surface defects. These phenomena can be resolved by spatial averaging (eg, defocusing or using a larger operator). Also, if systematic surface gradients become a problem, the verifier's repertoire can be expanded to include tests like regression that normalize gradients implicitly.

[6] If contrast is insufficient, the more likely

case, sensitivity must be adjusted. Two cases are distinguished. If the quantization window has not yet been adjusted for this point, the clips are first opened so that sensitivity can be raised to almost saturate the brighter side of the edge (I is set to 14 on a 0-15 scale.). This policy maximizes signal/noise for an assumed step type discontinuity.

Assuming the intensity is not yet saturated, can the sensitivity be physically raised? This question is addressed to the vision system's internal accommodation state which records the physical status of all accommodations. Sensitivity would be inhibited at this level if, for instance, the last time a program tried to raise it, a hardware override from the automatic target protection circuit was detected.

If the system has no objections, sensitivity is raised until saturation is achieved or the auto-target assumes control. In the latter case, the present program abandons sensitivity and proceeds to optimize the quantizer window. A more sophisticated program might, at this point, stop and question why the sensitivity could not be raised. Was the edge the limiting highlight? If not, could the highlight be compressed (relative to the edge) with a color filter, or shielded from view with a larger lens? These questions are not currently posed. Efficient answers would require access to a data structure, summarizing the

characteristics of intensity extremes, previously encountered while scanning the image. This knowledge has not yet been incorporated into the system. Nevertheless, the verifier does eventually overcome conditions that artificially depress signal/noise. This is now done with a rational trial and error sequence that follows those definitive tests the system is capable of making.

If Accom had previously been entered at the current edge position, sensitivity would already have been maximized in this way. Furthermore, the clips would have been set with respect to the optimized intensity levels (see Figure 4.19a). The purpose of a sensitivity adjustment in this case is to center the narrow intensity range of the edge within the previously established quantization window. This action is taken to avoid the situations depicted in Figures 4.19b,c. Figure 4.19b shows the contrast of an edge, compressed by hard clipping, because the quantization range could only be positioned at (eight) discrete levels. (Sensitivity can adjust the signal level relative to this window with about twice the resolution of moving the clips to an adjacent range.)

The other contingency (Figure 4.19c) occurs when a highlight anomaly is compressed, because the quantization range was initially established for a step type discontinuity.

[7] Quantization is set so that the upper and lower

clip levels straddle the average of the intensities, sampled by the wide contrast test at best resolution. (Hard clipping is undesirable for weak edges. However, if either side differs from the average enough to be completely hardclipped by the quantizer, the contrast will almost always be sufficient for the edge to be judged significant.)

[8] If the sensitivity and quantization accommodations successfully restored contrast, the significance should again be tested before more expensive accommodations are contemplated.

[9] At this point all correctable problems that could be definitively identified directly from the image have been resolved. The main justification for attempting the remaining accommodations is that edge significance is low at a point where an edge is expected. The diagnostics, however, are strictly ex post facto; the need for the accommodation is established by its success. The first accommodation to be tried is the color filter. (In a more sophisticated system, specific expectations about a scene would be used with relative cost to determine the priority in which the remaining unconstrained accommodations should be tried). A filter change is a relatively inexpensive solution to a number of problems (Chart 3.4), associated with edge detection. A lens change is also versatile, but by comparison, is more expensive because of the need to re-focus.

If a specific color edge were expected, the filters could be applied in a preferential order based on their effectiveness for enhancing that contrast (see Chart 3.3). Since there is no source of such information, the various filters are simply examined in sequence. Although the order is unimportant, all filters should at least be tried. The spectral composition of the room illumination is often a function of direction of incidence. Thus, no matter what the color of the edge (eg, even if it is white on both sides), a color filter may still have a beneficial effect on contrast.

Sensitivity and clips are optimized for each filter. It is more efficient to perform related adjustments immediately than to wait for the diagnostic program to discover such an obvious need the next time through. The filter and associated accommodations, providing the best contrast are selected.

[10] Another check is then made to see if this contrast is good enough.

[11] Constraint, c#9, is set to indicate that all contrast acommodations have now been tried without success.

[12] At this point it is necessary to invoke the auxiliary contrast constraint (Equation 4.2) to determine whether further effort to improve the operator significance is justified. For this purpose, the contrast is measured with the actual operator currently in use rather than the wide-spacing used in earlier tests. The acceptable

contrast threshold is approximately twice the expected contrast from the largest expected surface slope.

[13] If this narrow contrast is sufficient, the situation is similar to that faced at [3]. The possibility of re-focusing is considered first. The criterion used to evaluate the sharpness of focus is the magnitude of the gradient, integrated over the entire width between the points used to sample wide- contrast. Thus, focus will be optimized for the strongest edge characteristic (that contributing the most gradient) whether it be an intensity step or a line type anomaly.

[14] If focus is already satisfactory, temporal noise must again be suspected.

[15] At this point it is known that the present operator is unsatisfactory. Figure 4.19c. suggests the possibility that adequate contrast could be achieved with "line" operator OP↓2, even though the contrast obtained with OP↓1 was very small. Consequently, if the significance has not yet been evaluated with OP↓2, this is done. If OP↓2 is successful, Accom will not be called again. Consequently, the verifier will continue to use OP↓2, until low significance is again encountered. (Then, only if wide-contrast has improved enough to indicate that the edge characteristic has reverted to a step, will OP↓1 be reinstated.) If OP↓2 was already in use, the diagnosis can only conclude that no edge exists at this point.

IV.5.4 DESIGN CONSIDERATIONS OF EXPERIMENTAL VERIFIER

IV.5.4.1 GOALS

The function of the verifier, concisely stated, is to establish the likelihood that an edge lies approximately on the locus defined by two end points. The verifier is not intended to find the coordinates of the best edge. That job is delegated to an edge follower (Chapter 5).

IV.5.4.2 DECISION CRITERIA

The control level of the verifier has the responsibility of applying local edge predicates over the suspected locus to obtain a global estimate of the likelihood of an edge. The method of combining the individual operator results in a joint measure of edge confidence depends on what assumptions are made regarding the nature and independence of errors.

Complete independence is a bad assumption, because the quality of an edge tends to vary by neighborhood (eg. A smudge might reduce the contrast of many adjacent samples.). To minimize this bias, operators should be applied at random over the edge. Randomization also helps to avoid errors induced by regular sampling of a periodic texture.

Since the exact nature of the local statistical

dependencies is unknown, we are forced to consider heuristically reasonable ways to combine the results of local tests. A simple global average of individual local statistical significances appeared attractive, until experimental results disclosed an unfortunate tendency towards errors of commission. The reason, as previously discussed, was because the absolute confidence, returned when an operator encountered an isolated defect, was high enough to bias the entire cummulative result.

The currently favored statistic, whose effectiveness has been experimentally validated, is to compute the percentage of local operators that exceed the 2.96 significance level (99% likelihood of being an edge). A confident decision is insured by requiring a sample size of at least 5 local operators. A minimum required success rate of 80% will reduce the chance of misclassifying a homogeneous surface because of Gaussian noise to

$$\text{Prob (mis-classification)} = \binom{5}{4} (0.01)^4 (0.99)^1 + (0.01) \qquad (4.25)$$

The 80% threshold was an experimentally determined optimum, a compromise that combined good sensitivity to a variety of low contrast edges with high rejection of various uniformly textured surfaces. The value of 80% is, of course, dependent on two specific assumptions:

1. the particular operator (1 high x 5 wide on

each side of edge) used in the experimental study (see next section), and

2. equal a priori likelihood of an edge or surface. This assumption was adopted for unbiased testing of the system's potential. The verifier, in practice, will return whatever percentage was actually accummulated over the edge. The high level strategy program can then interpret this percentage in the context of its own a priori expectations to arrive at a more informed likelihood estimate of an edge's presence.

IV,5,4,3 GENERALITY OF DESIGN

The verifier was programmed with the ability to apply any operator, sequentially, between any two points on a video image. (This generalized scanning function is complicated by the discreteness of the sampled video raster, when the operator axes (and scan axes) are not aligned with the main horizontal and vertical image axes.) The scan locus can be defined interactively, using a cursor which is positioned on the face of the television monitor by adjusting two potentiometers. The cursor is used to indicate to the computer the image coordinates defining the end points of an edge to be verified. This facility allows the verifier to be tested independently of any other components of a vision system. (The end points can also be

passed as parameters, when the verifier is used with other vision programs.)

An operator is specified by defining a statistical function to be applied to intensities I(H,W), expressed in the normalized coordinates of Figure 4.17a. OP#1 and OP#2 (Figure 4.17b,c) are examples of two possible operators. The program is designed to allow simple substitution of alternative operators and control strategies for purposes of comparison.

IV.5.4.4 TRACKING OPERATOR

Because of spatial quantization, the actual edge will at best approximate the specified locus. The verifier will accept any edge whose end points fall within a maximum tolerance (OFFMAX) of the specified end points (see Figure 4.20).

To allow for edge uncertainty, operators OP#1 and OP#2 must be laterally scanned (perpendicular to the expected edge) at each point where they are applied. This function is performed by the TRACKING OPERATOR, flow charted in Figure 4.21. At the beginning of an edge, the selected T-operator (OP#1 or OP#2) is applied in a progressively widening pattern (Figure 4.22) about the expected center line. As soon as a significance exceeding 2.6 is obtained, the sequence halts and this significance is returned as the

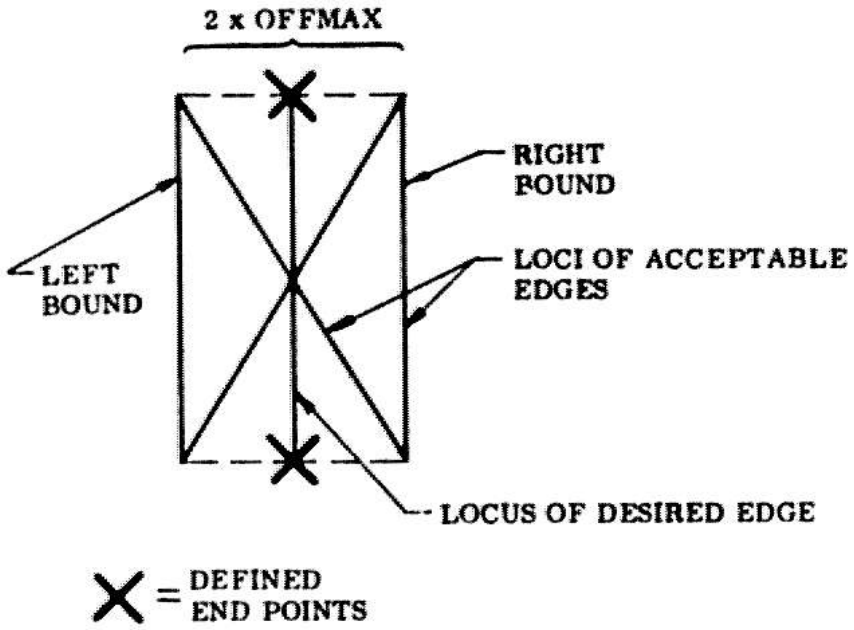


Fig. 4.20 Allowable Edge Uncertainty

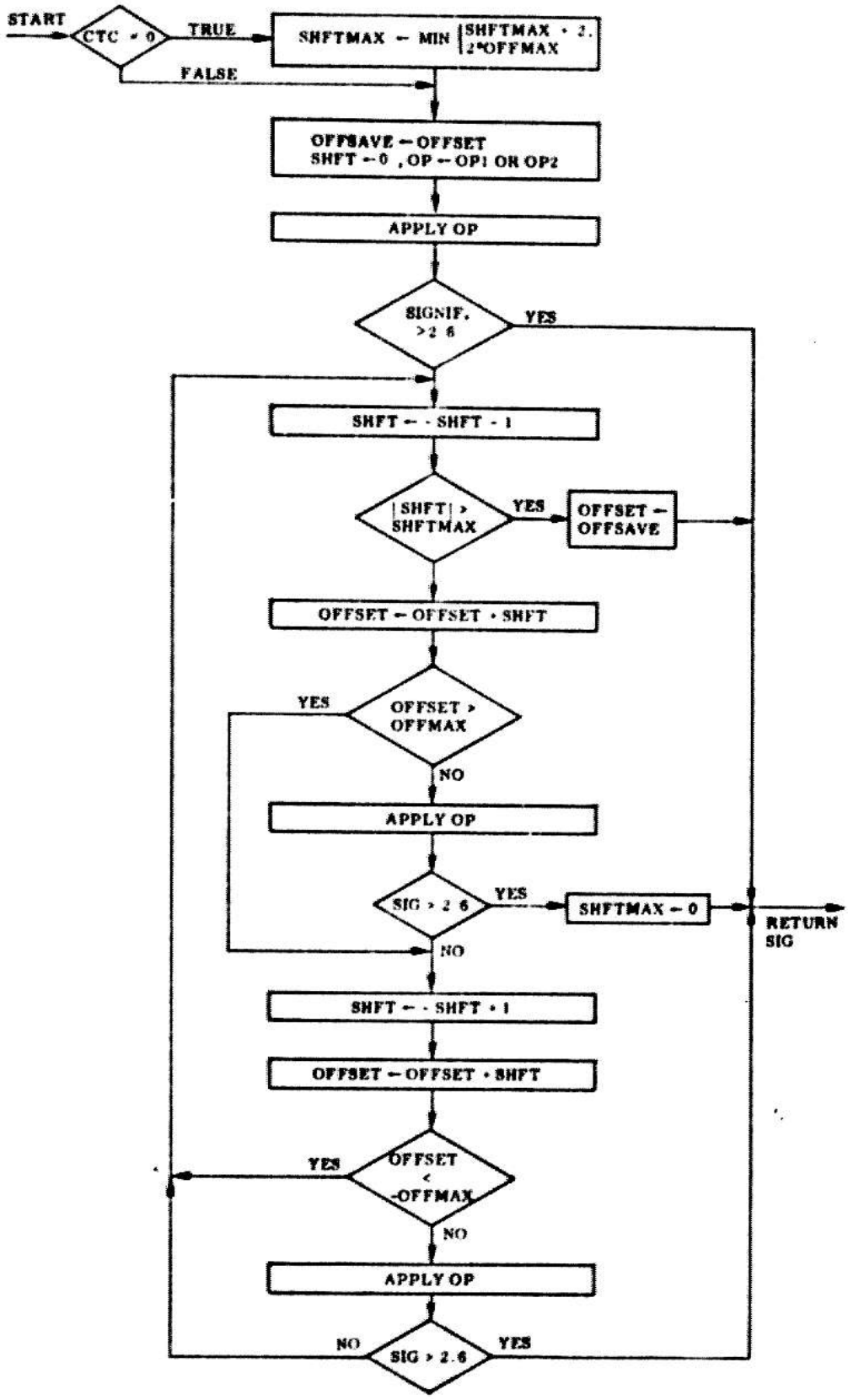


Fig. 4.21 Tracking Operator

value of the tracking operator. (The control level is, of course, oblivious to the composite nature of this operator and processes the significance, as if it were returned from an ordinary static test like OP#1.)

Since an edge is locally continuous, the best center line is now assumed to be parallel to the initally expected edge but offset to pass through the closest observed discontinuity. The operator when next applied, will begin scanning out from an extrapolation of the new center line (ie, the tracking operator). The size of OFFMAX depends on how closely the location of the desired edge is known and on the presence of irrelevant features known to be in the vicinity. Typically, the end points of an edge are known to +/- 4 raster units. If no significant discontinuity is found within this symmetric limit, the operator will again scan laterally out from the original center line the next time it is applied.

The expected deviation of the edge center between successive operators is less than the maximum deviation permitted for initial acquisition. To satisfy the requirement that edges be locally continuous, the maximum internal shift (SHFTMAX) allowed from a previously seen edge point is limited to 2 raster units for each operator length. (The tracked center line must, at all times, remain within +/- OFFMAX of the original center axis.)

(A more global solution to the problem of lateral

scanning and local continuity is suggested by Herskovitz [1970]. At each position along the the edge, the operator is shifted over the entire range, defined by OFFMAX. If the significance at any displacement exceeds the required local significance, the cumulative score is increased. However, the significance and relative position of all points that exceeded the threshold are retained, indexed by position along the center line. If the global percentage indicates an edge, all combinations consisting of one significant point at each index are examined to see if any set define a reasonably straight line. This test will insure that situations, like that shown in Figure 4.23a, will not be passed as an edge. This technique is time-consuming, but has the advantage that operators can be applied randomly over the edge. Some advantages of a randomized scan were mentioned earlier.)

IV.5.5 CONTROL STRATEGY OF THE EXPERIMENTAL VERIFIER

The basic flow of control for the experimental verifier is shown in Figure 4.24.

SNDX is the scan index. It defines a relative position along the directional locus established by the indicated end points. (Chart 4.3 is a complete glossary of the parameters used in Figure 4.24.) The end points themselves correspond to the limits SNDXI (initial) and

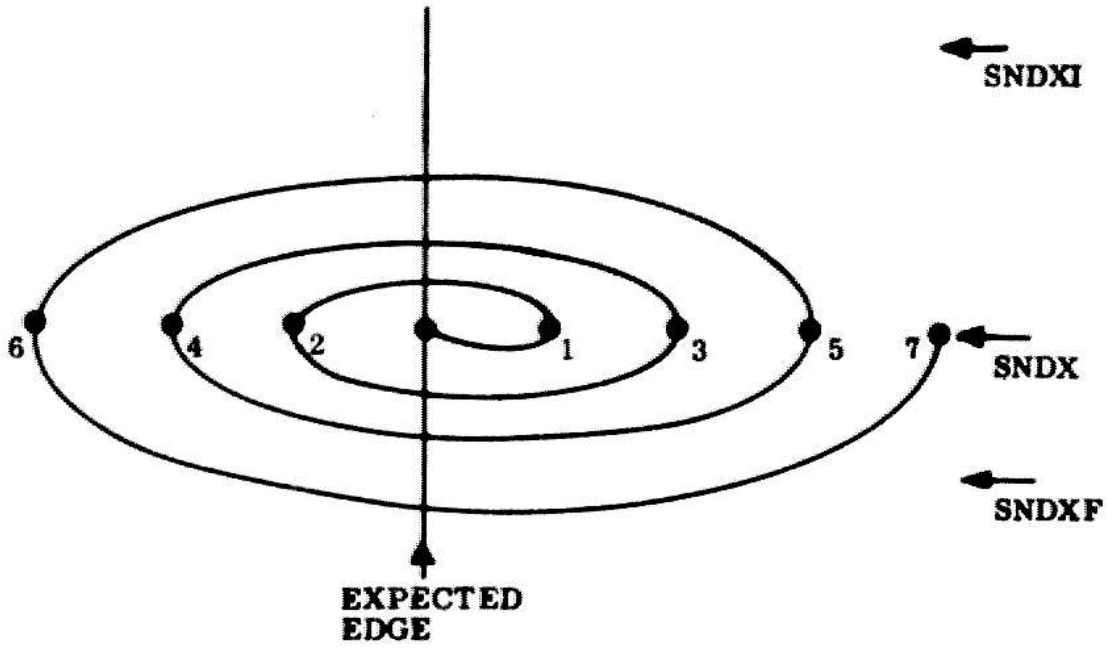


Fig. 4.22 Ordered Sequence to Test for an Edge if Significance is Too Low Where Edge Was Originally Expected

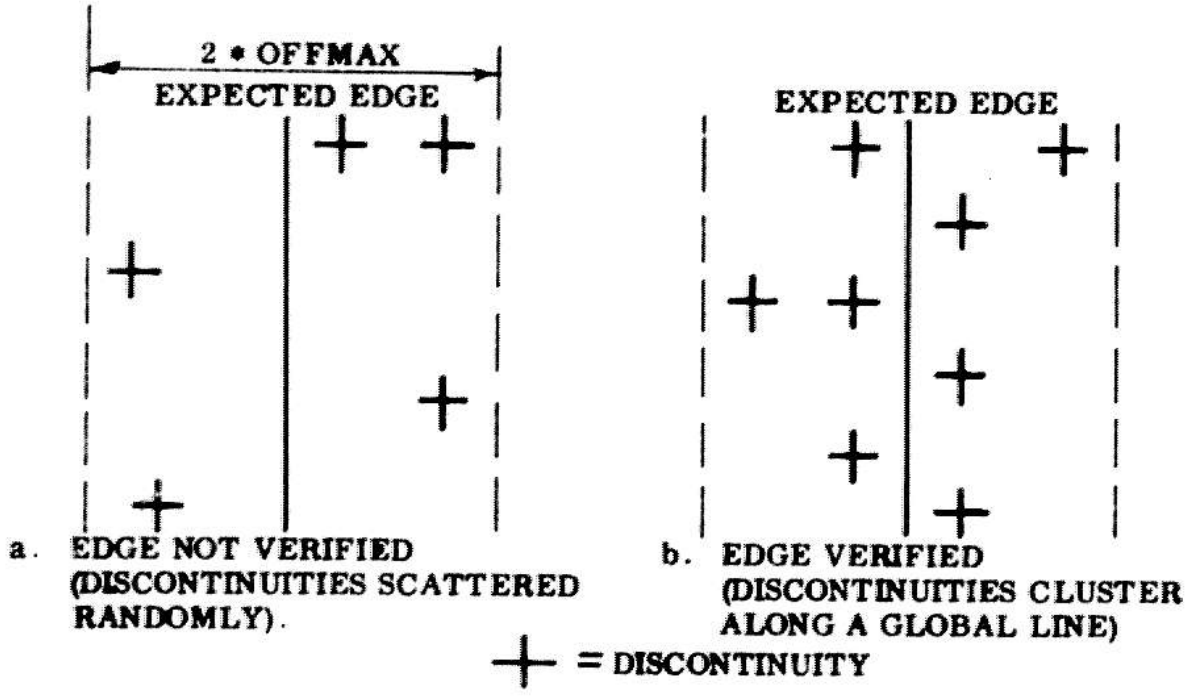


Fig. 4.23 Checking Local Continuity in a Randomized Scan

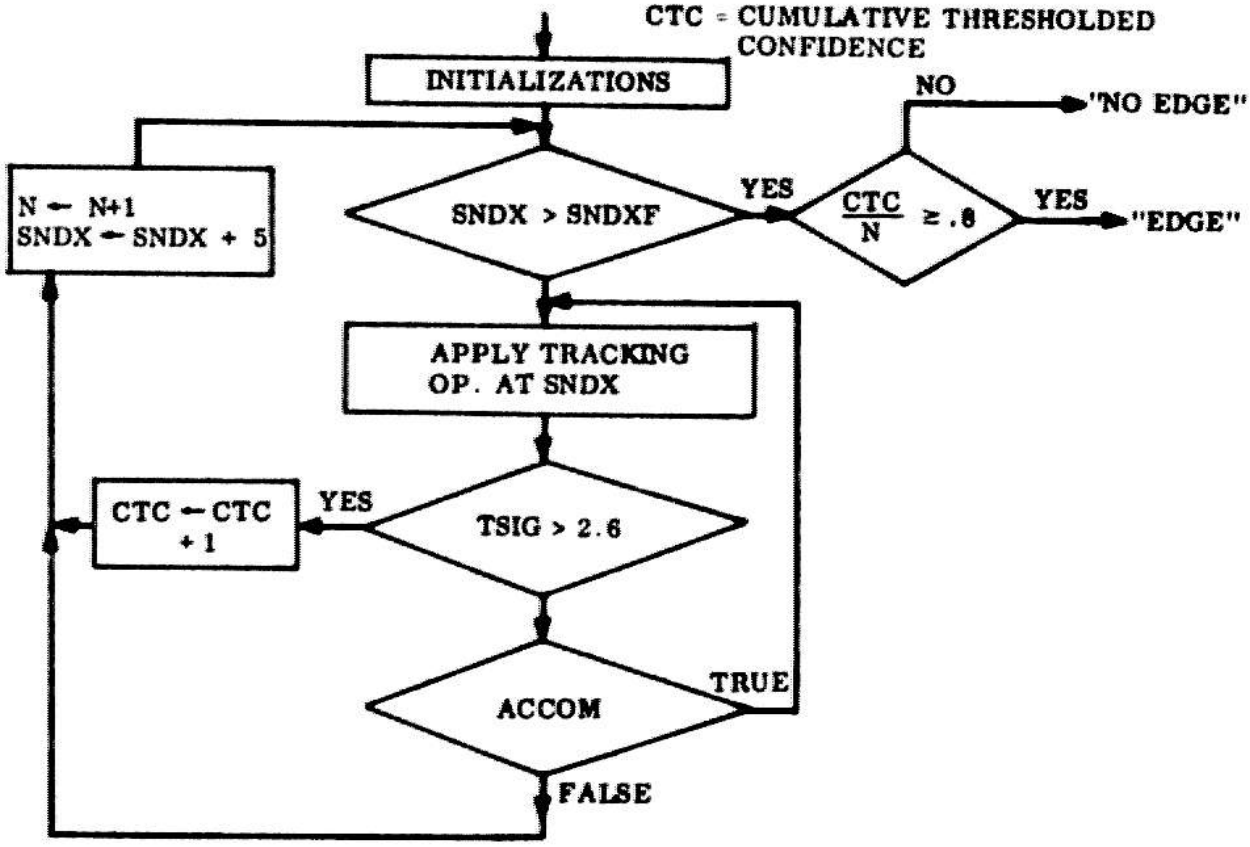


Fig. 4.24 Experimental Verifier: Flow of Control

FRAMES	# of images (used for temporal averaging)
OFFSET	Current deviation of tracking operator from best center line
OFFMAX	Absolute maximum deviation allowed
SHFT	Incremental deviation of center line operator from previous operator
SHFTMAX	Maximum incremental deviation allowed
OP	Operator type; OP1 - 1H * 5W (step edges) OP2 - 1W * 5H (lines)
SNDXI	Initial scan index (along edge)
SNDXF	Final scan index
SNDX	Current scan index
SIG	Current significance of tracking operator
CTC	Cumulative threshold significance count (Number of times sig. > 2.6)
N	Operator count

Chart 4.3

Principal Parameters of Experimental Verifier

SNDXF (final). SNDX specifically indicates the coordinates of the center of the height dimension of an operator. The effective sampling range is constrained by the condition

$$\mathbf{SNDXI} < \mathbf{SNDX} < \mathbf{SNDXF} \qquad (4.26)$$

Operators are applied every 5 raster units (due to the height of the "line" operator) over the entire length of an edge. The number of operators that will be applied along an edge is then given by

$$N_{max} = \frac{\mathbf{SNDXF} - \mathbf{SNDXI} + 1}{5} \qquad (4.27)$$

where N*max must be rounded down to the nearest integer.

IV.5.5.1 HIGHLIGHTED DETAILS OF CONTROL LOOP IN OPERATION

A verifier run begins with the sequence of initializations summarized in Chart 4.4. The main control loop (Figure 4.24) is then entered. The scan index (SNDX) is checked to see if the end of the expected edge has been reached. If not, the next operator is applied. If the score of this operator is significant, CTC is incremented. Otherwise, accommodations are attempted to improve the score. Eventually, adequate significance will be achieved or all appropriate accommodation possibilities will have been

Scan locus	defined by manual interaction or passed as parameters
All constraints	reset
Lens	1" (shortest)
Pan-Tilt	edge centered in field of view (manual)
Color filter	clear
Focus	edge in focus
Sensitivity	.8 * Edam
Clips	0,7
Op	OP1
FRAMES	1
OFFSET	0
OFFMAX	$\pm$ 4
SHFTMAX	OFFMAX
N	0
CTC	0
SNDXI, SNDXF	defined by endpoints of edge
SNDX	SNDXI

Chart 4.4

List of Initializations

exhausted. The main loop variables are updated; N to record the total number of operators applied, SNDX to select the next position along the edge. The run is terminated when SNDX has been incremented beyond the end of the edge, indicated by SNDXF. At that time, the accumulated count of significant local tests (CTC) is divided by the total number of operators applied (N) to determine the final percentage of points along the edge that were judged significant. 80% is needed for a successful decision.

The requirement that 80% of the local tests confirm a discontinuity insures that all sections of the expected edge must contribute to a successful decision. This guarantees that an edge decision will not be reached because of evidence clustered over a limited portion of the desired edge. A shorter edge could, for example, have coincided with a portion of the expected one.

IV.6 FUTURE DIRECTIONS FOR VERIFIER DEVELOPMENT

IV.6.1 IMPLEMENTATIONAL DETAILS

Many of the extensions proposed in this chapter require elaborate global control and data structures. As these projected features become realities, a considerable amount of implementational detail will need to be resolved before the verifier can take full advantage of them. It is

painless to suggest, for example, that a lens change is an effective way to eliminate an undesirable highlight. In practice, an involved series of tests is required just to determine the fact that an external highlight is at fault. Then range estimates of the highlight are needed to decide (with the camera model) whether that highlight will be excluded from the field of view of the new lens. Range estimates are also required for the end points of the edge so that the scan locus can be appropriately redefined for the new lens. Range finding is, in itself, an intricate subject, aspects of which are covered later in this work (see also Sobel [1970] and Falk [1970]). Many similar system considerations have admittedly been brushed over in this presentation.

IV.6.2 COST-EFFECTIVE VERIFICATION

The principal objectives of our current work have been to demonstrate the effectiveness of accommodation. Some elementary heuristic concepts of efficiency, primarily on a local level, have been mentioned. For instance, it is known that a lens change involves more time than changing a color filter and is often less effective when a well-defined color edge exists. However, we have not yet seriously faced the problem of using accommodation in a cost-effective manner. (Indeed, we are probably guilty of using

considerably more accommodation than cost-effectiveness would justify.)

There has not been sufficient experience with the verifier to rigorously quantify the actual cost-effectiveness of the various accommodations. Indeed, we do not yet know even how to begin this task. The characterization necessarily depends on many external factors (ie. scene characteristics, state of global knowledge, parallel tasks, etc.) whose effects are still ill-defined. Since these parameters are conditioned on the entire task experience, it seems unlikely that any meaningful global optimization can be accomplished before we better understand the use of a verifier in the context of a complete vision system.

IV.6.2.1 OUTSTANDING PROBLEMS

We can, however, identify specific areas in which the benefits of more work seem especially promising. The present policy of applying the local operator exhaustively over the length of the edge is very inefficient. This is especially so in cases when the final decision would have been obvious from the strength or weakness of a few local operators, applied at randomly selected points.

The basic theory of sequential testing is well-known (Wald [1947]). Given a required confidence, this theory

defines rigorous termination bounds that are applied after each local test, A decision is then made to accept, reject, or continue testing the hypothesized edge.

An interesting extension of this theory is to include the option of re-accommodating as a fourth decision alternative, Some relevant questions at this level concern how accommodation will alter the other decision boundaries, where and how often should it be invoked, and how far should the diagnosis be pursued each time,

An optimal strategy must, of course, consider the relative cost-effectiveness of alternative accommodations in the context of global knowledge available to the system, The most promising accommodation must then be compared with the merits of examining another point on the edge. A surprisingly effective heuristic may be to simply let each operator accommodate as much as necessary to detect the edge.

An exception to this laissez faire approach might be taken on the grounds that the initial operators would squander the verifier's total budget. Expensive accommodations, like focus, color and lens change, would be tried in an exhaustive effort to achieve the required significance, leaving no funds for subsequent operators. However, this objection ignores the fact that most edges have relatively uniform characteristics along their extent, It is thus probable that the color filter and focus that

give best results at some point on the edge will be applicable at all points. Furthermore, a long lens is universally helpful. If the edge is subtle enough to require these accommodations, it is best that they are done early in order to benefit the most operators. It is unlikely that they will need to be repeated.

An "early" termination rule could also take into account the edge confidence, required by a vision strategy, to justify a global decision in the Bayesian sense. (The higher the a priori expectation of an edge, the less sensory confirmation is required to obtain a given decision confidence.)

In the immediate future, the most pressing requirement is to analytically validate the effectiveness of the diagnostic criteria and related accommodations that have already been implemented. This validation can proceed independently of cost considerations and should lead to more selective diagnostic criteria.

IV.6.3 VERIFYING GENERALIZED PROPERTIES

The verifier concept can be applied in a more general context than the detection of simple intensity edges. An edge can be preprocessed to convert arbitrary surface characteristics, such as texture, color, slope, or depth into numerical values. These values can then be

interpreted by the verifier as though they were intensities. Edges will often be most strongly defined by different properties at different parts of their extent. A verifier would do well to adapt its tests to the locally strongest characteristics. Preprocessing could be thought of as a generalization of operator selection and thus treated as another accommodation option.

IV.6.4 VERIFYING GENERALIZED LINEAR FEATURES

The edge verifier can obviously be generalized to test for features, composed of a combination of edges (such as, corners, formed by two or more planar surfaces, which meet at a vertex). The trivial approach, however, of applying the verifier sequentially to each of the constituent edges is certainly not the best method. For example, the existance of several edges introduces possibilities for "flitting" back and forth between edges as the relative difficulties of the decisions vary. Becker [1972] considers a simplified form of this problem. In our case, the optimal strategy will be complicated by conflicting accommodation requirements which introduce an overhead cost to change edges.

Continuing in this vein, we can generalize a feature verifier into an object verifier or more appropriately a model-driven edge follower. Suppose that the partial

profile shown in Figure 4.25 had been isolated by a cheap edge follower. If the edge follower began scanning from the bottom of the picture, it is reasonable to assume that these edges lie on the table-top. Using the ground plane assumption (Roberts [1963]), the absolute position of the three vertices is then constrained in space.

On the basis of this information a preliminary recognition hypothesis can be made. The corresponding prototype model can be used to predict the location of all remaining edges. The verifier can then confirm the recognition choice by finding the predicted edges. (This technique is probably most useful for rapidly dismissing unwanted shapes when the machine is looking for a specific object.) Similarly, the verifier can confirm that a known object is still where the computer last saw it. Kelly [1970] has suggested still another case in which the "model" function is served by a heavily averaged version of the original image that contains only fuzzy remains of the strongest edges. The profiles of principal objects in a complicated environment can often be easily located, because averaging suppresses all fine detail and minor edges. A tracking verifier, guided by these suspected boundaries, can then be used to extract the actual contours of the desired object. The verifier in each of these cases capitalizes on the known location of expected edges. This information is not utilized by conventional edge followers.

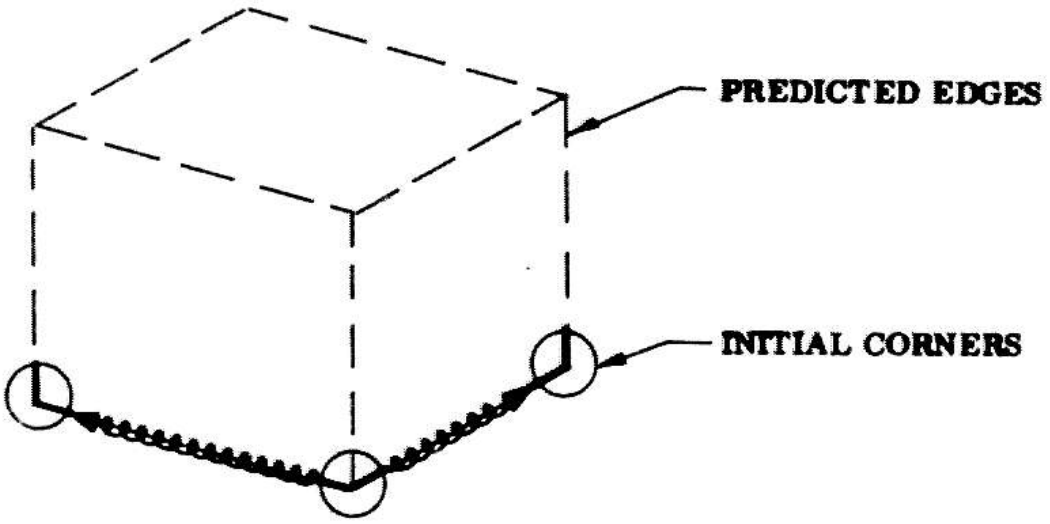


Fig. 4.25 Three Angles Found by Edge Follower Establish an Initial Context to Predict Remaining Edges

CHAPTER V: THE EDGE FOLLOWER

V.1 THE CONCEPT OF EDGE FOLLOWING

Since the current hand-eye system deals with planar-faced objects, straight line intensity edges are the most important features used in scene analysis. The edge follower performs the initial reduction of an image into a crude line-drawing. It first searches the image for a strong discontinuity of intensity (see Figure 5.1). Then it enters a trace mode. The intensity gradient is tracked around a closed contour, which returns to the point of acquisition. This contour provides a basic context for anticipating finer details that can then be sought, for example, with the verifier.

Edge following is based on several simple ideas:

1. Edges are distinguished by locally large intensity gradients. Gradients are approximated on a discrete surface by the difference in intensities at neighboring sample points.

2. Edges are continuous. Successive edge points are found by proceeding a short distance from the present point, perpendicular to the direction of maximum intensity change.

3. Edges of physical objects form closed contours. An edge, once acquired, can be tracked around a complete periphery, returning to the point of acquisition.

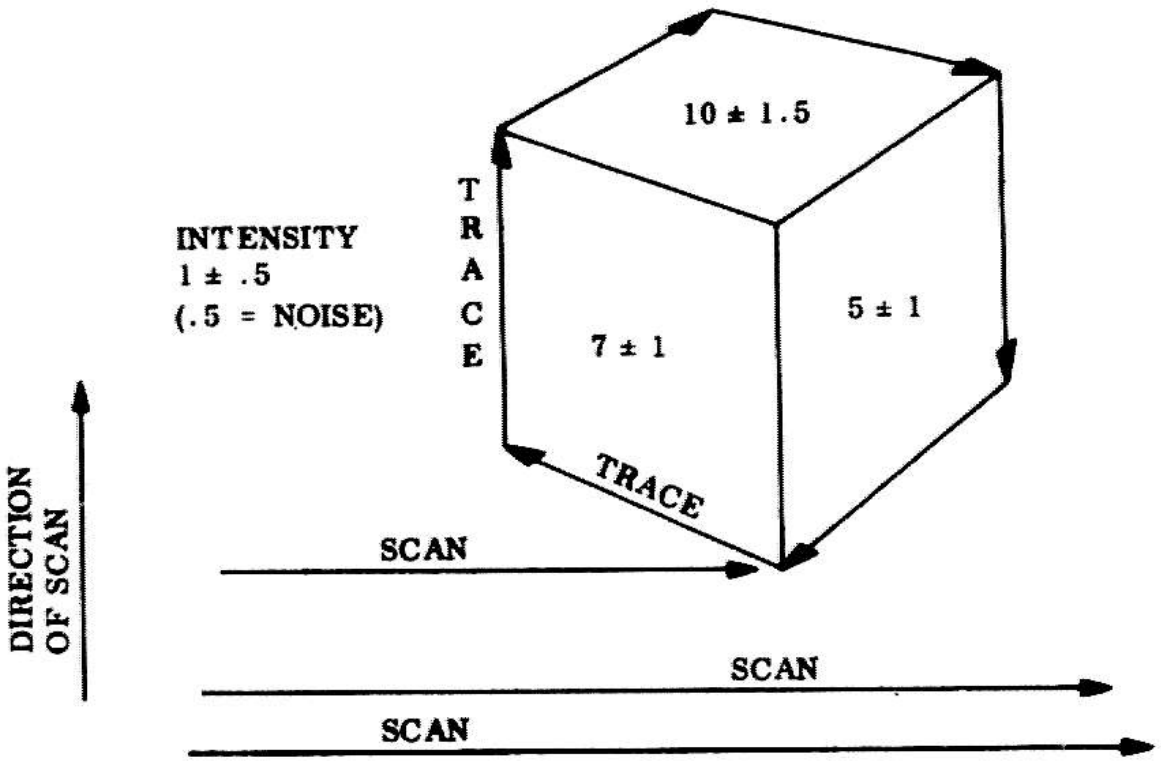


Fig. 5.1 Acquisition and Edge Following

In practice, the mechanics of edge following are preoccupied with the subtle nuances of functions such as:

1. examining the intensity surface to locate the center of an edge,

2. distinguishing the background from the interior of an object,

3. predicting the location of the next edge point at a corner,

4. maintaining data structures that associate edge points with objects and recall whether a point has already been encountered, and

5. fitting lines to the individual edge points (etc.).

The rather elaborate code needed to cope with these details was written principally by Karl Pingle. A detailed description of these aspects will appear in a forthcoming jointly authored paper (Pingle and Tenenbaum [forthcoming]).

In this chapter, we shall discuss the edge follower in the context of an application of accommodation. The treatment will be brief. Understandably, there are many similarities between the accommodation functions that are effective for edge following and those that were described for the verifier. We will describe the differences in philosophy between edge following and verifying and stress the implications that these differences hold with respect to accommodation strategy. The edge follower is a crucial

component of the overall vision system. A significant improvement in the performance of this program has been realized through the use of accommodation.

V.2 DESIGN PHILOSOPHY OF AN ACCOMMODATIVE EDGE FOLLOWER

The edge follower is intended to extract the most significant information from an image without processing it in bulk. Edges are acquired by a coarse raster scan that seeks gross discontinuities in intensity. After an edge is found, attention can be focused on the specific points indicated by the continuity of the edge.

V.2.1 ROLE OF ACCOMMODATION

A simple intensity difference is used to acquire edges during the scanning phase. The accommodation is set to provide low resolution over a wide dynamic range. A wide dynamic range insures that most edges will be contained in the image. Low resolution implies that an edge will most likely be acquired at a point of strength. (If no edges are found, a more sensitive accommodation can be used. However, this is seldom needed. Usually, at least one edge of an object will show up with good contrast.)

As the edge is tracked, it may lose contrast, perhaps by passing through a region of shadow. The

perception of the edge at weaker points is facilitated by the fact that its presence can be specifically anticipated. This motivates the use of sophisticated processing and accommodations to recover the edge, if it is not seen where expected.

V.2.2 ADVANTAGES OF AN ACCOMMODATIVE EDGE FOLLOWER

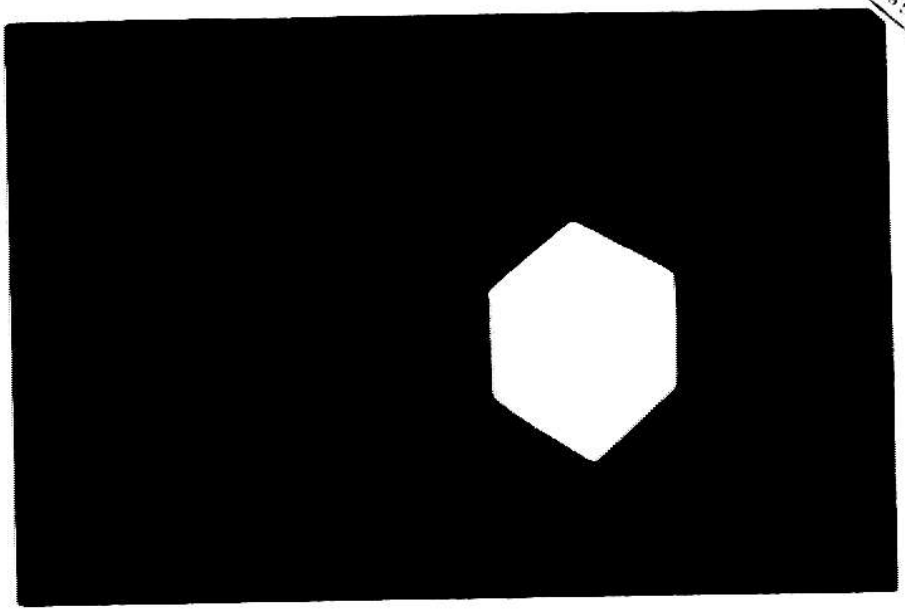
The accommodative edge follower can conveniently adjust its level of effort to suit the difficulty of finding an edge in a particular local region. High sensitivity is reserved for specific contexts where it can be applied economically and where the expectation of a specific edge orientation can significantly reduce the risk of responding to noise.

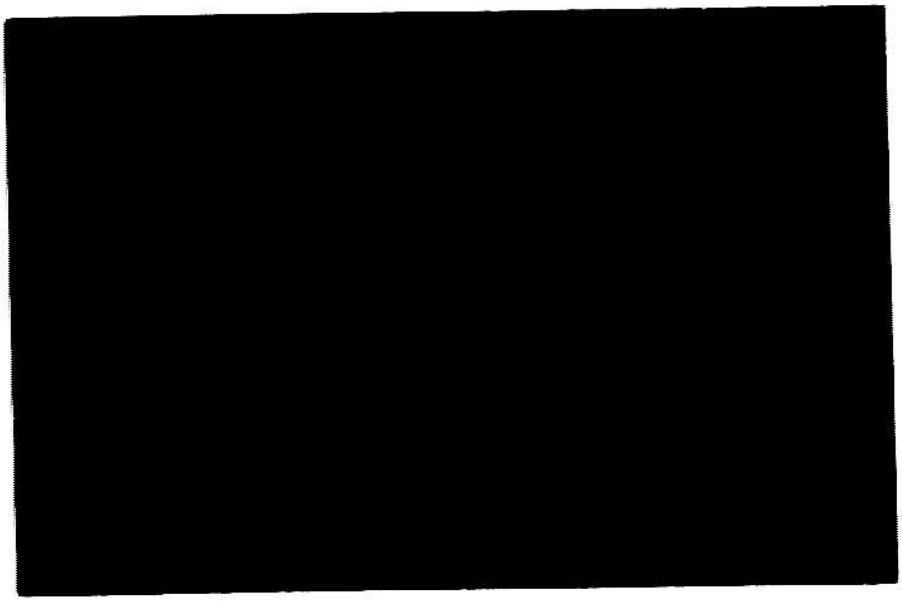
By contrast, a bulk processing approach (ie. Roberts [1963]) must apply a uniform level of effort over the entire image. If the effort level is too low, edges will be missed. If it is too high, the inefficiencies of an already exhaustive approach will be compounded. Furthermore, bulk interpretation of the edge points will be complicated by the presence of many random noise points that would never have been seen, if high sensitivity had been confined to well-defined edges. In practice, any constant level of effort will most likely be either too high or too low at any given image point.

A non-accommodating edge follower is constrained by the same fundamental limitations that handicap a bulk image processor. The situation is in a sense more serious, because the success of an edge follower depends on continuity; once the edge has been lost, at a point of weakness, it cannot be conveniently re-acquired at a later region of strength. Previously, edge followers would fail, because small sections of a contour were slightly noisy or of weak contrast. In most of these cases, the combination of accommodation and a more sophisticated local operator can recover the weak edge element. These functions, selectively applied, allow contours to be efficiently traced under a wide variety of conditions. Furthermore, even noisy edges of consistently low contrast, such as those in the interior of the cube in Figure 5.2a, are often obtained with the accommodative edge follower (see Figure 5.2b).

V.3 ACCOMMODATION FOR ACQUISITION

The edge acquisition schema which we shall now describe is intended for use in the absence of a specific goal and any knowledge of the environment. Since the system has not yet seen anything and has no expectation of what to look for, it must proceed with a thorough search for a region satisfying some low level criteria of interest. A simple and effective criteria of interest is a discontinuity





in the intensity surface.

As we have already mentioned, a thorough search is facilitated by using wide accommodation windows. In addition to a full quantizer range, a short lens is desirable. It provides coarse resolution over a wide field of view, coupled with a large depth of field. A clear color filter is used to obtain unbiased tonal renderings. (If some strategy program were interested in a specific area of the table or a specific range of hues, then the window could be narrowed to reflect these constraints. This accommodation would be done from a higher level before the edge follower was called.)

While scanning for a discontinuity, it is appropriate to record characteristics of the environment that can be used to reduce a future task-oriented search. Currently, we record the highest, lowest, and average intensity in regions corresponding to horizontal strips across the entire image. This information is associated with the specific color filter that was fitted. In this way, a crude color map can be accumulated (over several complete scans). This map can direct attention to specific regions, if a colored object is later needed for a task.

The intensity map is also helpful, when nothing is found with the coarse windows. A higher level strategy can request that the search be pursued in more detail. The clips are then narrowed to resolve the specific range of

intensities now known to exist on each scan line. The key to improved sensitivity is, as always, to optimize the accommodation for more localized contexts. (The lens could also be changed to obtain more spatial resolution. However, difficulty in finding large objects, ie, major dimension ≥1", is almost always the result of inadequate contrast.)

V.3.1 OPERATION OF SCANNER

The operation of the scanner is charted in Figure 5.3. After the accommodation windows have been initialized, the scan proceeds upwards from the lower lefthand corner of the television frame (see Figure 5.1). This direction was chosen in order that nearer objects, which are not likely to be occluded, will be found first. The complete frame is covered by about 30 scan lines. On each line, intensity is sampled at approximately 40 points.

If the intensity at adjacent points differs more than one quantization level, an edge is suspected (This threshold was chosen to avoid detecting quantization contours that result from slowly varying intensity surfaces.).

Before the edge follower is actually called, it is worthwhile to apply inexpensive tests. They insure that the detected intensity difference was not the result of random

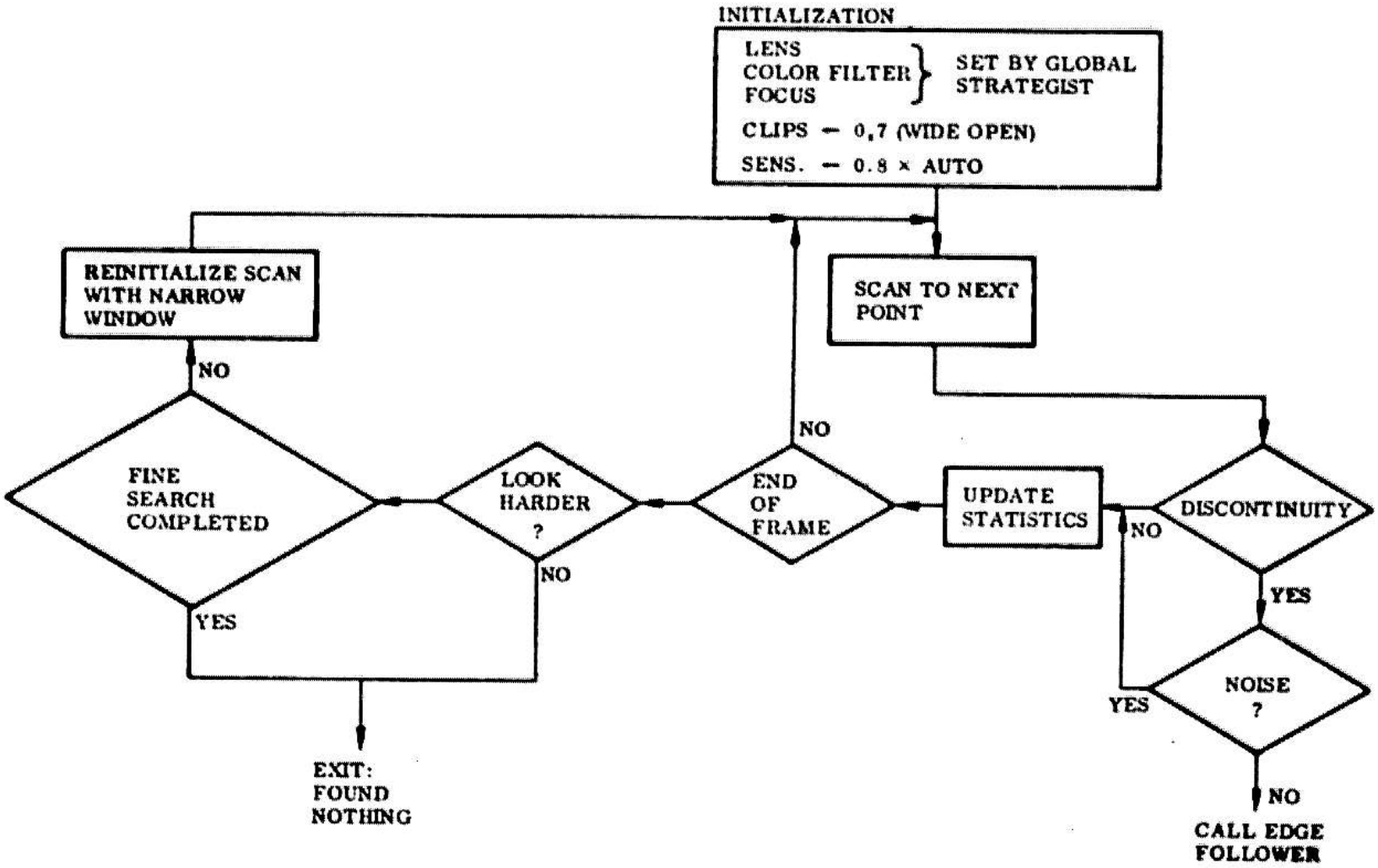


Fig. 5.3 Acquisition Scanner (Simplified)

noise or an isolated anomaly in the surface. An effective validation requirement is that the discontinuity have a sufficient spatial extent. This condition implies that a significant difference also be observed in the average of intensities, obtained from small areas about the original detection points. If this requirement is met, the actual edge is then localized between these two samples and the tracing stage begun.

If nothing has been found at the end of a frame, the scanner will exit unless specifically instructed to look harder. In this case, the frame is re-scanned. The clips, however, are now set to bracket the range of intensities, observed on each scanline during the previous coarse scan.

V.4 ACCOMMODATION IN EDGE TRACING

The accommodation requirements of edge following are theoretically identical to those of edge verification. However, accommodation is used more conservatively in the edge follower. Local continuity is a weaker basis for expecting an edge than the need to establish global consistency. Consequently, the verifier's motivation to exhaust all possible accommodations to see an edge is not present in the edge follower. (Admittedly, the edge follower could utilize more of the global information available to

it, such as the current length of the contour.)

The edge follower also does not have the advantage of a global context in which to evaluate what it finds. It is not equipped with enough selectivity to deal with the full sensitivity that optimum accommodation can provide. For example, when the edge follower was run with the accommodation package, designed for the verifier, it was equally inclined to follow wrinkles in a black cloth covering the table as it was to trace high contrast boundaries of blocks. There was no way to discriminate between these cases at the myopic level at which the edge follower currently functions.

The overall scheme is to use the edge follower to extract medium and strong edges most likely to define the major boundaries of objects. Weaker features can then be anticipated in the global context, established by these major edges and sought using the verifier. Our experience has shown that it is easier to predict the existance of missing edges than to fight the combinatorics, necessary to decide which edges to eliminate, when there were too many.

The accommodation strategy of the edge follower reflects this philosophy. When an edge is lost, a pre-determined amount of effort is committed to recovery. If this effort level is insufficient, the edge follower awaits instructions from a higher level. It may indeed suit the purposes of a strategy program for the trace to proceed with

more costly accommodations in the manner of the verifier. The more usual case is for the edge follower to give up and resume its global scan. If the gradient being tracked was truly part of a major contour, it will most likely be re-acquired at another point. In that case, if the subsequent trace intersects any part of the originally observed edge, the two segments will be re-unified in the data structure.

V.4.1 ACCOMMODATION WHEN AN EDGE IS LOST

The edge follower accommodation strategy is outlined in Figure 5.4. It is, as expected, a simplified and slightly re-ordered subset of the verifier strategy shown in Figure 4.18. The first accommodation that is tried, when the edge is lost, is to apply a more sophisticated edge extraction technique.

V.4.1.1 EDGE OPERATORS

The edge follower makes use of two operators. The principal one, designed by Irwin Sobel, spatially smooths the intensities over a 3x3 region of the raster. It then approximates the gradient with two directional derivatives about the central point of this square region. This operator is illustrated in Figure 5.5. The Sobel operator

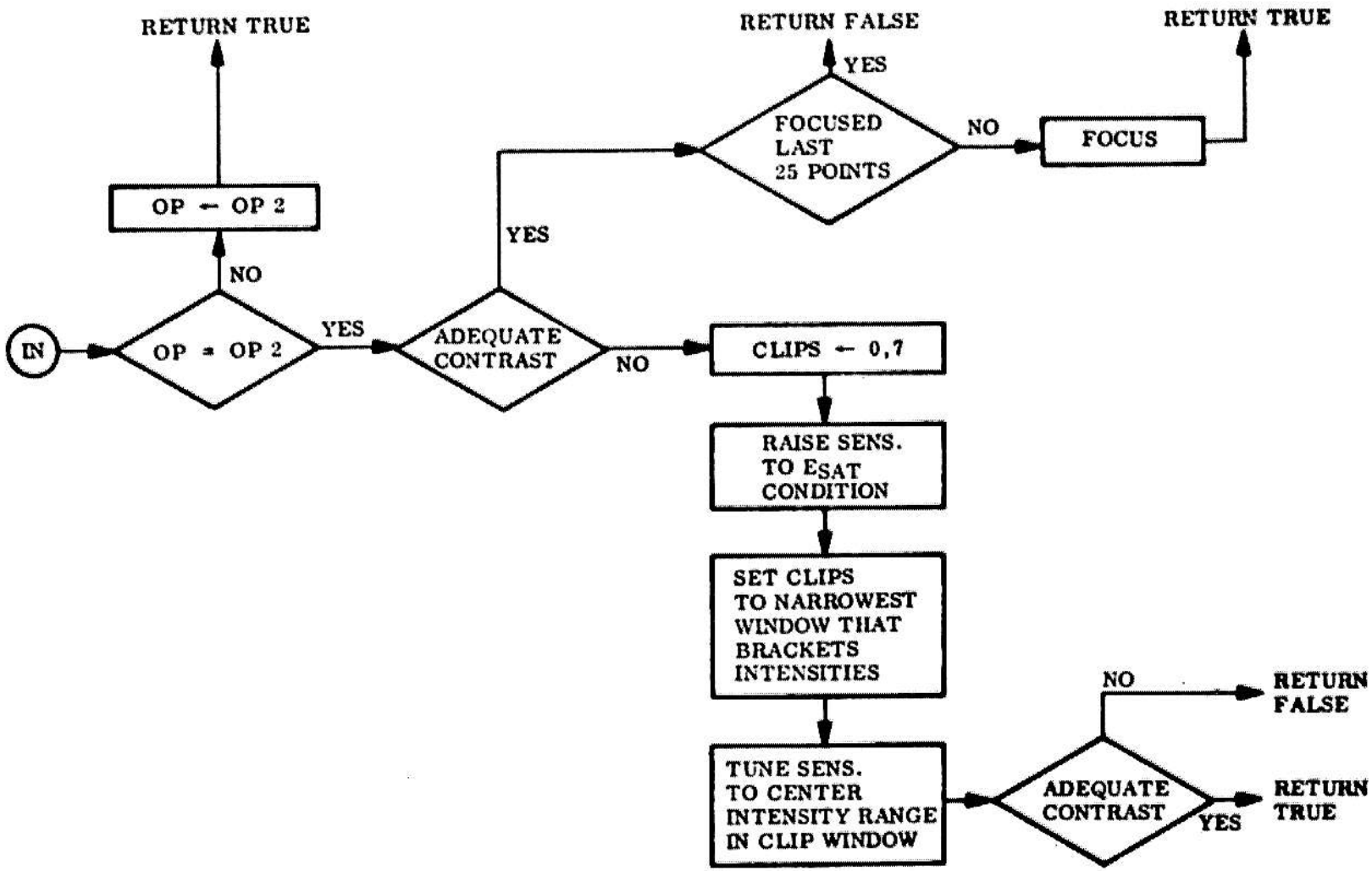
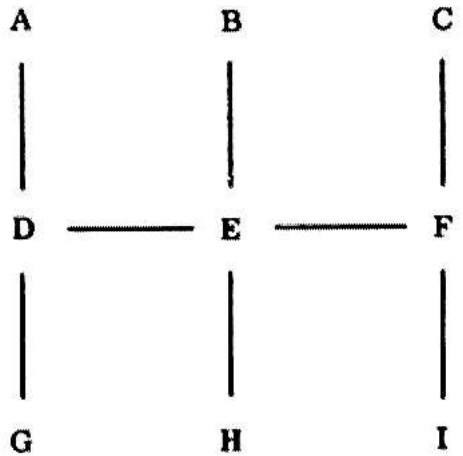


Fig. 5.4 Accommodation for Edge Follower



9 RASTER POINTS

$$\text{GRAD} = i\frac{\partial f}{\partial x} + j\frac{\partial f}{\partial y}$$

$$\frac{\partial f}{\partial x} \approx (C + 2F + I) - (A + 2D + G)$$

$$\frac{\partial f}{\partial y} \approx (A + 2B + C) - (G + 2H + I)$$

Fig. 5.5 3 × 3 High Speed Gradient Operator

is fast and reasonably effective, even in the presence of modest noise.

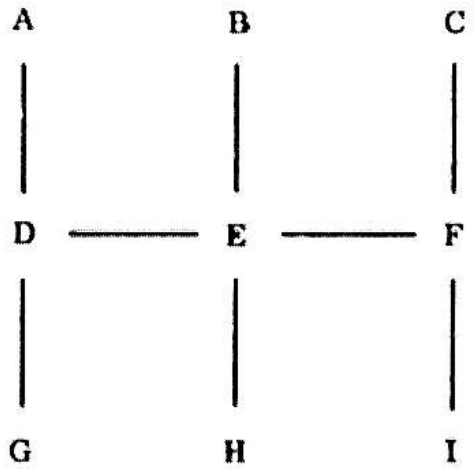
The second operator is the large area test developed by Heuckel [1969] and discussed in Chapter 4. It looks for any edges, lying in a circular neighborhood of up to 177 picture points about where the edge was originally expected. Heuckel's algorithm overcomes two of the most common causes for losing an edge:

1. Weak edges can be obscurred by local noise at the particular point chosen by the edge follower to apply Sobel's operator.

2. Because of noise or an unusual corner, the edge follower can predict the wrong place to look next.

Both of these problems can be combatted by sampling a larger area.

Heuckel's operator is, of course, much slower than the 9 point operator and is used only when the local gradient is inadequate. If the additional processing succeeds in recovering the edge, the trace routine will continue using Hueckel's operator, but only until the edge quality improves enough to switch back to the original 9 point test.



$$\text{GRAD} = i\frac{\partial f}{\partial x} + j\frac{\partial f}{\partial y}$$

$$\frac{\partial f}{\partial x} \approx (C + 2F + I) - (A + 2D + G)$$

$$\frac{\partial f}{\partial y} \approx (A + 2B + C) - (G + 2H + I)$$

9 RASTER POINTS

Fig. 5.5 3 × 3 High Speed Gradient Operator

V.4.1.2 CAMERA ACCOMMODATIONS

If the edge cannot be detected even by Heuckel's operator, the camera accommodations must be checked. The primary requirement is that there be adequate contrast across the suspected boundary. The accommodations used to maximize contrast are analogous to those discussed in the last chapter. The sensitivity is raised to almost saturate the brighter side. Then the quantization window is narrowed about the range of intensities, found in the immediate vicinity of the problem point. Since Heuckel's operator functions best with a faithful characterization of the intersity surface, the narrowest window for which all intensities are linearly encoded should be used. Because the sensitivity can be set to a finer resolution than the clip levels, provision is made to use sensitivity to center the intensity surface within the chosen window.

If the contrast across the boundary is adequate, but no discontinuity sharp enough to qualify as an edge has been found, the problem could be poor focus. Re-focusing will be done, if it has not previously been attempted in the immediate vicinity.

V.4.2 OPERATION OF ACCOMMODATIVE TRACE ALGORITHM

The integration of accommodation into the overall

trace algorithm is illustrated in Figure 5.6. The current operator is applied at a point where a gradient is expected because of edge continuity. If the operator is not successful, the accommodation routine (Figure 5.4) is called.

If the problem is diagnosed to lie in the domain of the available accommodation variables, the indicated adjustment is made and the operator retried. This loop is another manifestation of what has been called "performance feedback". The success of the operator is the ultimate criterion by which accommodation is evaluated. If the operator fails again, the accommodation routine is re-called. This process will terminate when either the operator is successful or the accommodation routine can make no adjustments.

If accommodation cannot recover the edge, the trace returns to the point of initial acquisition and attempts to close the contour by proceeding in the opposite direction. This strategy will succeed, when the edge is exceptionally weak or obscurred at one particular point. It is also helpful in completing a contour containing a very sharp angle (Figure 5.7).

If the operator is successful in finding an edge point where one was expected, the following possibilities must be considered:

1. Does this point complete a closed contour with

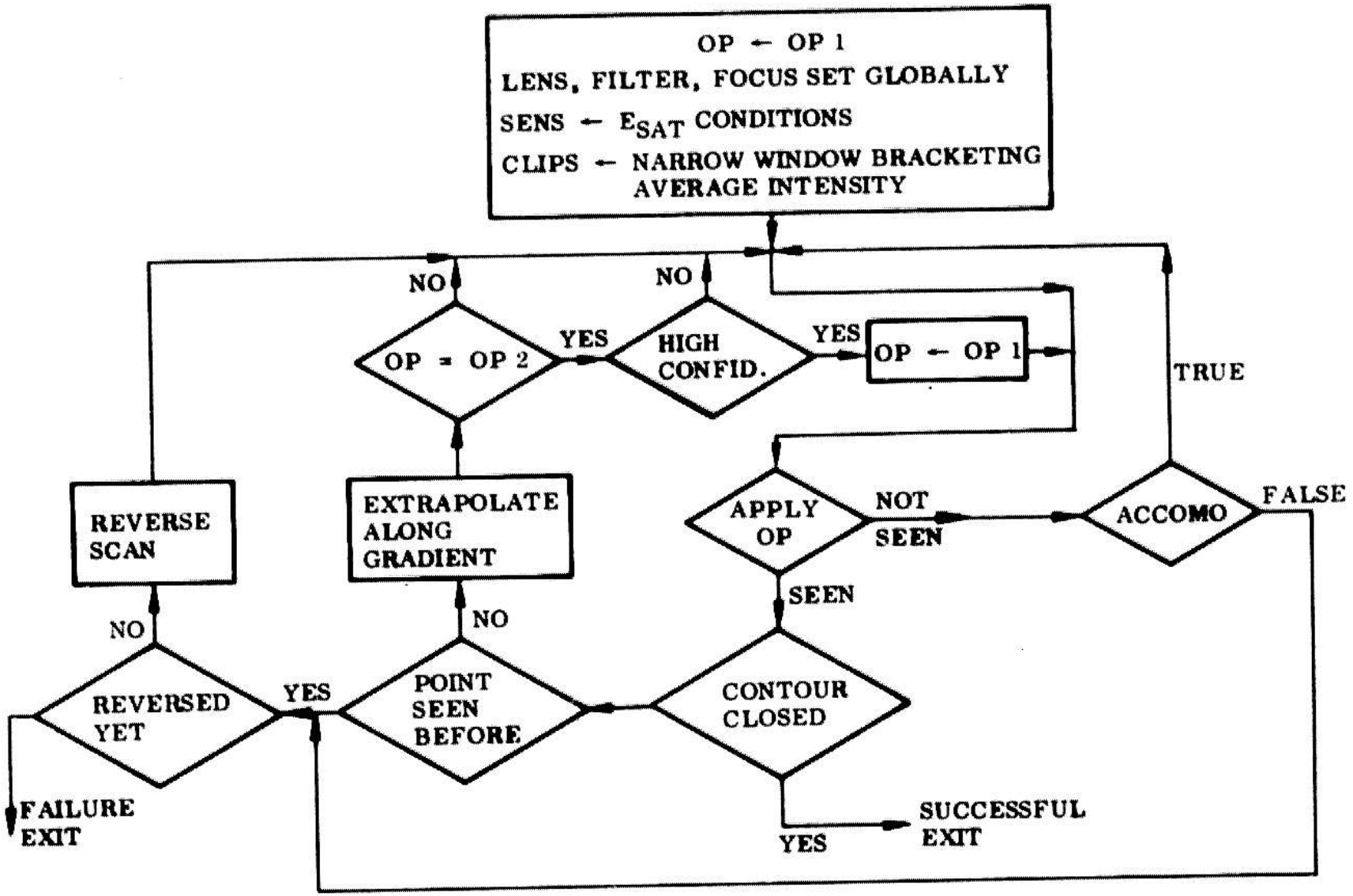


Fig. 5.6 Accommodative Tracing Algorithm (Simplified)

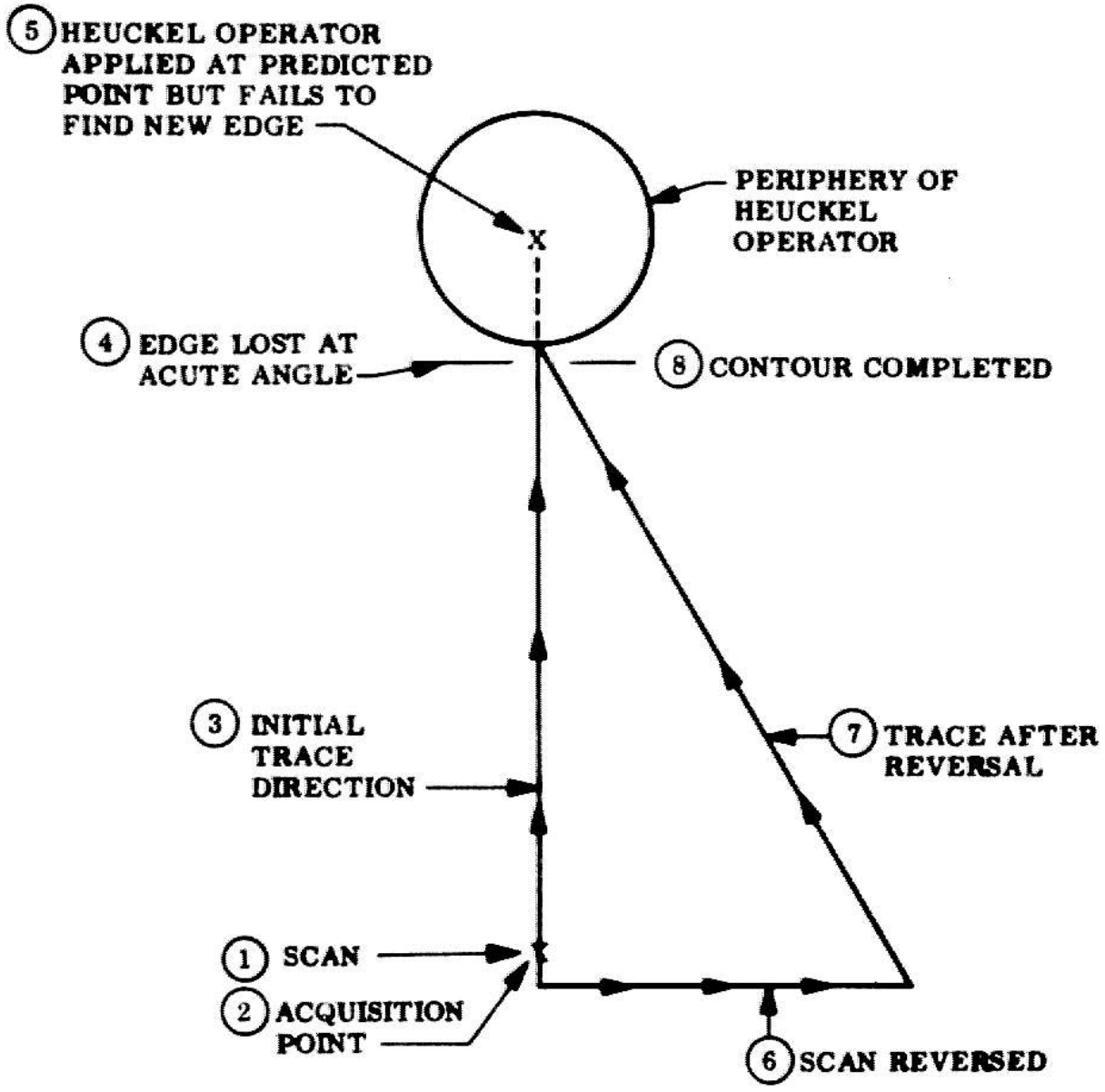


Fig. 5.7 Tracing a Contour Containing a Sharp Angle

respect to the point of initial acquisition? If so, terminate.

2. Has this edge point been seen before? If it has and condition 1 was not applicable, then the edge follower has erroneously looped (Figure 5.8 shows an example of this condition, caused by shadowing.). Recovery is attempted by scanning in the opposite direction from the acquisition point. If the reverse scan was, in fact, already in progress, the trace terminates. The strategy program may then decide that enough of this contour is available to attempt a recognition hypothesis and verification sequence. Otherwise, the global scanning mode is resumed.

3. If neither of the above conditions apply, the edge point is accepted as a valid extension (and, in fact, a confirmation) of the emerging contour. The gradient direction is then used to predict where to look for the next edge point. Finally, if the Heuckel operator had been used, the edge quality (as indicated by a parameter of that operator) is checked to see whether the Sobel operator can be re-instated.

V.5 FUTURE DIRECTIONS

One major shortcoming of the current edge follower is its lack of adequate high level evaluation and control. One can imagine simple criteria for detecting global

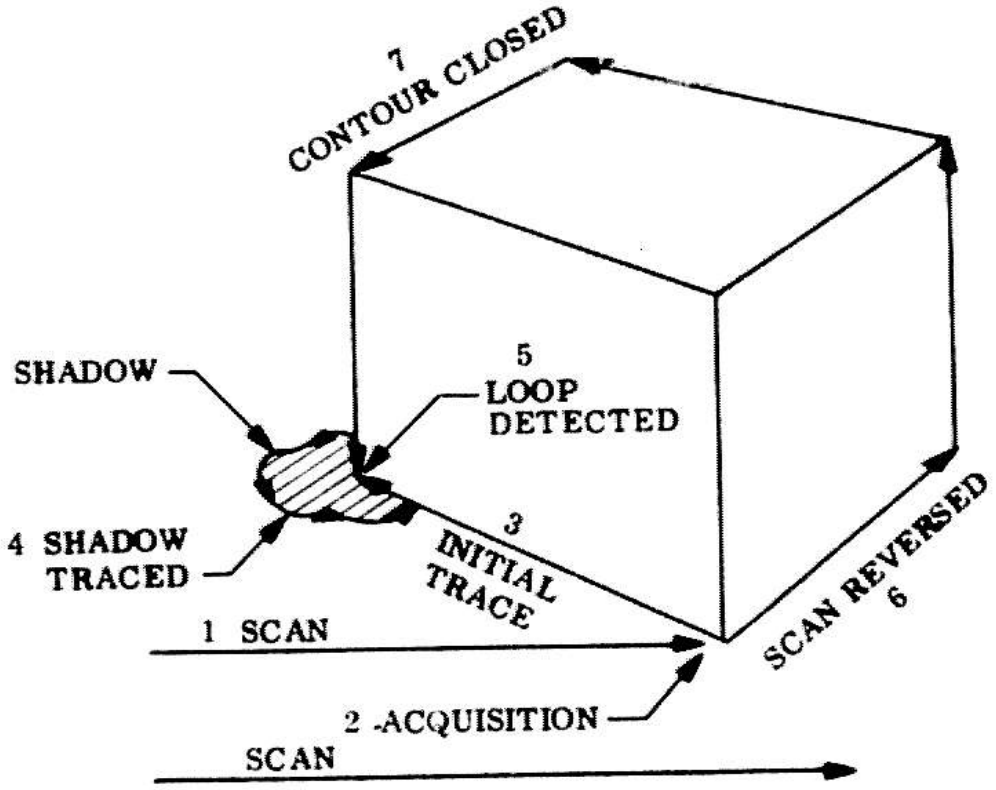


Fig 5.8 Tracing a Contour Containing a Local Ambiguity

indications of poor edge quality. If the edge follower now locks onto a cloth wrinkle or a shadow, it will waste considerable effort attempting to follow a low contrast noisy edge, before it ultimately fails. These situations can be identified much earlier by noting that an edge is particularly wavy and not what would be expected from a planar-faced block. Edge quality could also serve as an effective noise filter, allowing the confident use of more powerful accommodations.

Present plans are to implement this intermediate level of evaluation. We then plan to consider blurring some of the distinctions between the edge follower and verifier. A line predictor can attempt to complete contours when the simple reversal strategem fails at both ends (Grape [1970]).

A predictor can be based entirely on local heuristics. For example, lines can be hypothesized to connect existing vertices. Partial lines can be extended to form new vertices. However, when the system is looking for a specific object, the predictor could also make use of that object's topology. If the partial contour does not correlate with a portion of the desired contour, the global scan can be immediately resumed to find other edges.

CHAPTER VI: FOCUS RANGING

VI.1 INTRODUCTION

Humans infer depth in many ways. They interpret known characteristics of an object and its environment in terms of a very comprehensive model of the visual world (Gregory [1966]). This ability is characteristic of man's reliance on a vast store of experiences to overcome his limited ability to make accurate metrical measurements. A machine, on the other hand, must, at least for the time being, employ complex measurement algorithms to overcome its rather limited store of knowledge. Focus ranging is an example of a technique that can be used to estimate the distance to any resolvable inhomogeneity in the scene (eg. texture, edge, corner, etc.). Identification of the feature is specifically not required. (Animals use focus information also but, apparently, in a corroborative capacity.)

The technique is based on the fact that an object, at an unknown range, will be in best focus at a particular image distance. If the point of best focus can be found, the resulting image distance, y_b, can be used in the lens equation (Equation 2.35) to obtain the corresponding object distance, x_b.

$$x_0 = \frac{y_b f}{y_b - f} \qquad (6.1)$$

To automate this process requires, first, an analytic criterion which the computer can use to evaluate the sharpness of focus in an image (specifically in that part of an image containing the feature whose range is desired). The computer can then focus by varying the image distance until this criterion is maximized. In our system, this variation is accomplished by physically moving the vidicon (ie. image plane) along the principal axis of a stationary lens (see Figure 6.1). The distance between the vidicon and lens is obtained from the voltage of a potentiometer whose wiper is attached to the vidicon (The linear equation between voltage and image distance can be calibrated by focusing at two known ranges.).

In this chapter, we derive a simple model of focusing and use it to develop several focus criteria. The obtainable range uncertainty depends, of course, on the sensitivity of the criteria. The factors that affect this sensitivity will be enumerated and shown to be strongly dependent on the accommodation parameters. A focus ranging system is described. It attempts to minimize range uncertainty with appropriate accommodations.

VI.2 BASIC FOCUSING MODEL

In Chapter 2 it was shown that a point source, out of focus, is imaged as a circular disk (see Figure 2.13).

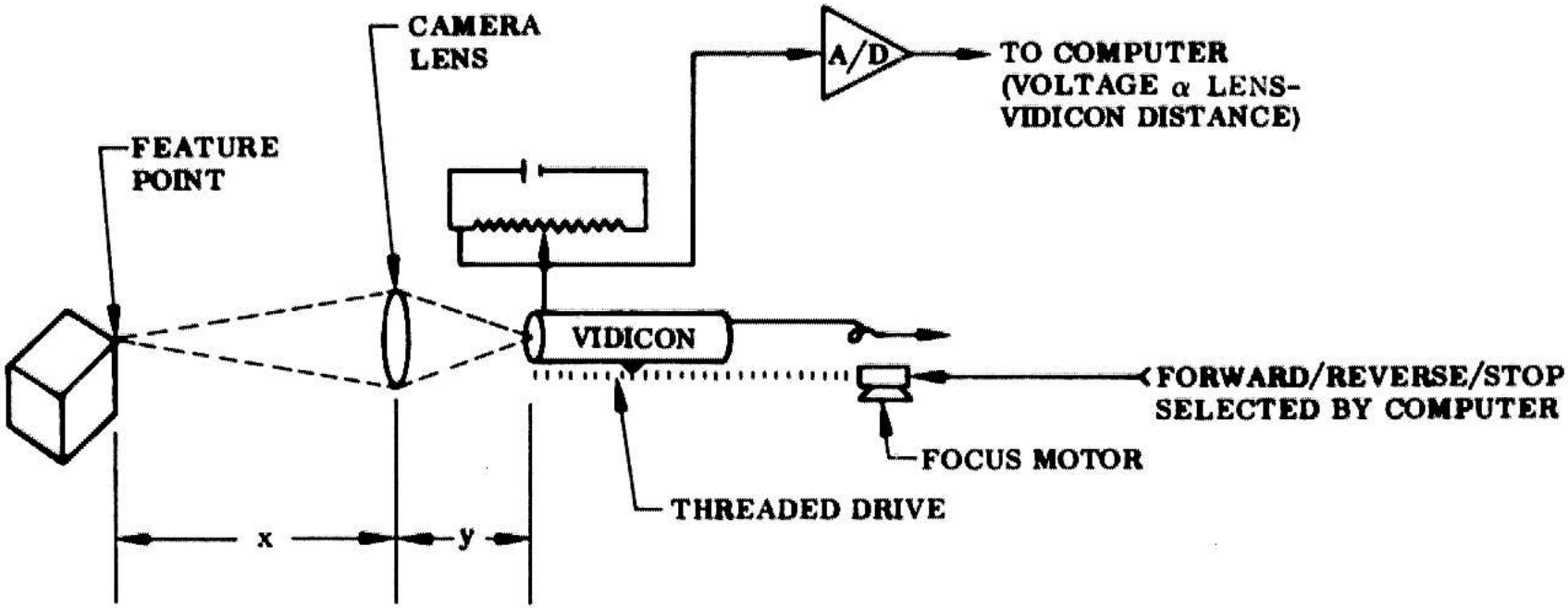


Fig. 6.1 Camera Autofocus Hardware

More precisely, the lens collects a portion of the light energy originally concentrated at a point and distributes it uniformly (assuming negligible diffraction) over a circular area of diameter ∂t. This diameter is a linear function of the distance, ∂y, between the actual image plane and the plane of best focus. This function is related to depth of focus and can be found from Equation 2.57b by making substitutions 6.2 and 6.3 (compare Figures 6.2a and 2.15).

$$\partial y = \frac{D}{2} \tag{6.2}$$

$$\partial t = c \tag{6.3}$$

Solving for ∂t yields

$$\partial t = \frac{d\,\partial y}{f} \tag{6.4}$$

In cross section, the one-dimensional intensity profile of a point source and its defocused image resemble an electrical impulse and the waveform that results from applying it to a low-pass filter (see Figure 6.2b). We pursue the filter analogy to deduce the effects of defocusing on an arbitrary feature. The effect of a linear system on any function can be analyzed by convolving that function with the system's response to an impulse.

A simple lens is a linear system. Its effect on any distribution of light can be obtained by spatially convolving that distribution with the image of a point source. Figure 6.3 shows the effect that defocusing will have on the profile of an ideal brightness edge. The transition of the edge has assumed the width, $2t$, of the pulse with which it was convolved.

(The linear transition produced by this one-dimensional analysis is slightly oversimplified. In two dimensions the edge transition is formed by the superposition of many defocused point sources, extending along the edge boundary. Taking this into account, we can express the actual height of a unit step as a function of position (s) over a transition width $2R$

$$E(s) = \frac{1}{\pi R^2}\left[s(R^2 - s^2)^{1/2} + R^2\left[\sin^{-1}\left(\frac{s}{R}\right) + \frac{\pi}{2}\right]\right], \quad -R < s < R \qquad (6.5)$$

This relation is a close approximation to a linear function except for slight rounding at the edges of the transition. This rounding is not a significant effect in terms of the factors that govern focus ranging. For clarity, the theory in this chapter is thus developed in terms of a one-dimensional analysis.)

The analysis will be developed in terms of edge profiles. Edges are the most important single feature for

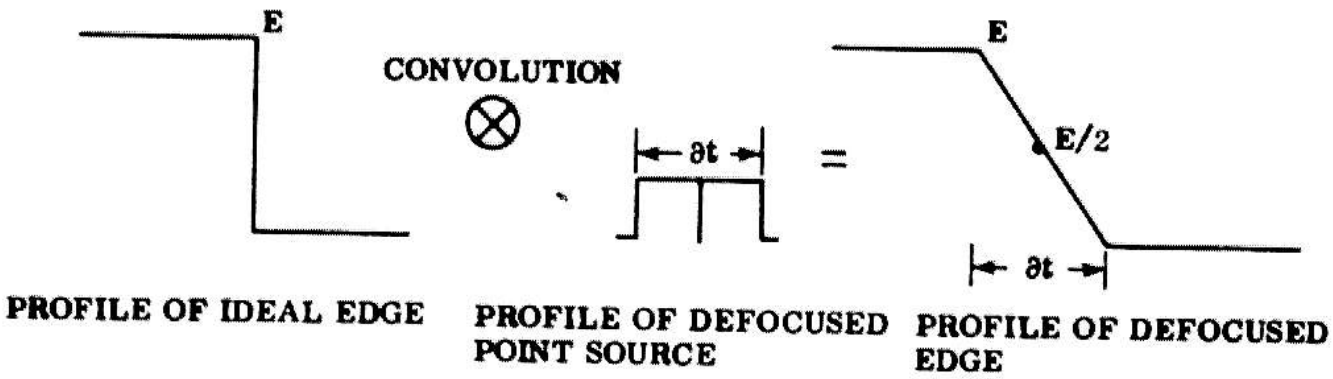


Fig. 6.3 Effect of Defocusing on the Intensity Profile of an Edge

which range information is required. By superposition, the analysis can be extended to any feature (or texture) representable as an array of intensity discontinuities (eg. a checkerboard pattern). We add the temporary restriction that edges lie in a single plane (ie. a single piece of paper, half white and half black), oriented at right angles to the lens axis. This restriction will be removed later in the chapter.

VI.3 CRITERIA OF FOCUS

The goodness of focus is directly related to the transition width of an edge. Figure 6.4 shows the appearance of an edge under varying degrees of defocus. The transition width is, in each case, equal to the diameter of an equivalently defocused point source. Conceptually, the simplest way to evaluate focus is to directly measure this transition. Any of the verifier operators whose value reflects the quality of an edge can serve as a measure of focus quality. Thus, a T-score calculated at points S4I, S4J (in Figure 6.4) would increase monotonically from a to d with improving focus (This result, coincidentally, demonstrates why focus is an important accommodation for verification.).

In applications (like verification) where an edge is expected, these operators constitute the best criteria of

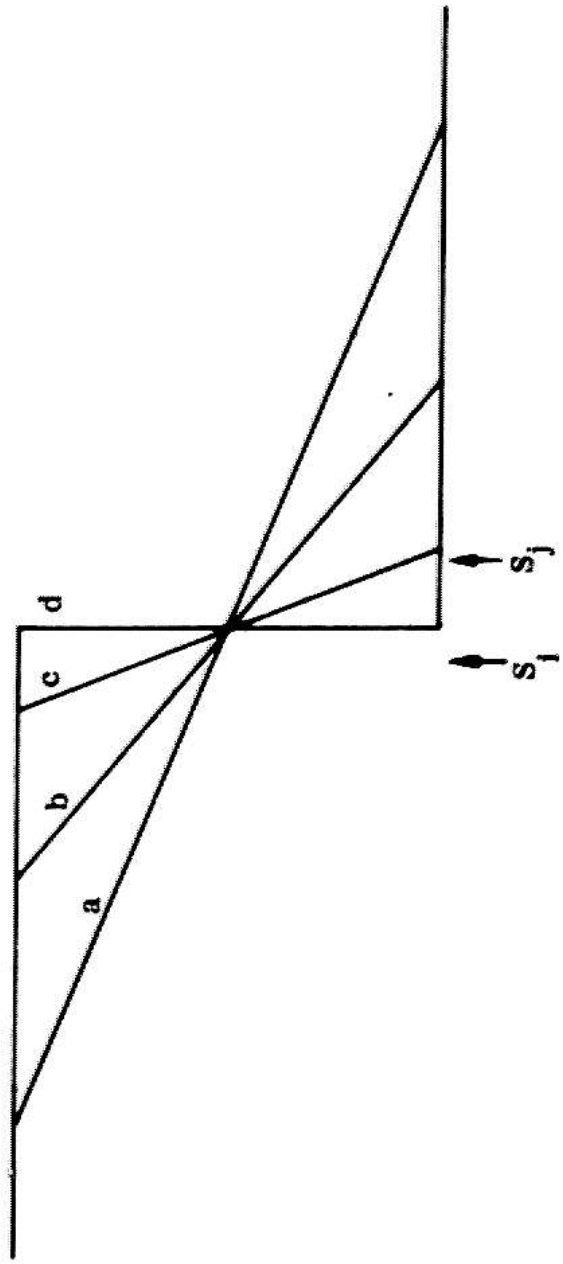


Fig. 6.4 An Ideal Edge Profile Under Varying Degrees of Defocus

focus for the very same reasons that make them sensitive edge detectors. They are difficult to apply, however, when the edge location is not well-known or, more generally, when the individual step profiles are not arranged in an orderly line but rather distributed over a surface to form texture. We next consider examples of analytic and heuristically based criteria which measure the "edge content" of an area.

VI.3.1 FOURIER TEXTURE ANALYSIS

A straightforward analytic approach is to apply a two-dimensional, fast Fourier transform over the region of interest and sum the power in the high frequency terms. This criterion is suggested by the relation between edge sharpness and high frequency content in the corresponding spectrum. Horn [1968] analyzed this criterion and used it as the basis of a focusing program at MIT. We did not adopt this criterion, because the information contained in a complete Fourier analysis is not necessary for focusing. As a result, the FFT algorithm is inefficient in this application. More importantly, the influence of accommodations on focus sensitivity are more naturally described in the space-time domain. The heuristic criterion, to be discussed next, lends itself well to this purpose.

VI,3,2 THRESHOLDED MAGNITUDE OF GRADIENT

To assess the quality of focus in a region of interest, we seek a function whose value is proportional to the local slope of the intensity surface (and thus to high frequency spectral content). The value of such a function, summed over the region, would make an effective focus criterion. A suitable function is the common gradient. A discrete approximation to the analytically defined gradient at position s, is given by

$$G(s) = \frac{I(s+1) - I(s-1)}{2} \qquad (6,6)$$

Two practical problems arise when G(s) is accumulated over a global area:

1. The cumulative gradient over a symmetric portion of any texture pattern (eg. over the range S-1 to S+2 in Figure 6,5a) will always sum to zero. This problem is simply avoided by using |G(s)|.

2. The cumulative magnitude of the gradient across an edge must, by definition, sum to the total height of the edge, regardless of the width of transtion. In Figure 6,5b the gradient over the range S-1 to S+8 is 7 for edges a,b,c. To avoid this situation, we must introduce a non-linearity. |G| could, for example, be squared or thresholded before compiling the sum. A simple threshold is preferred,

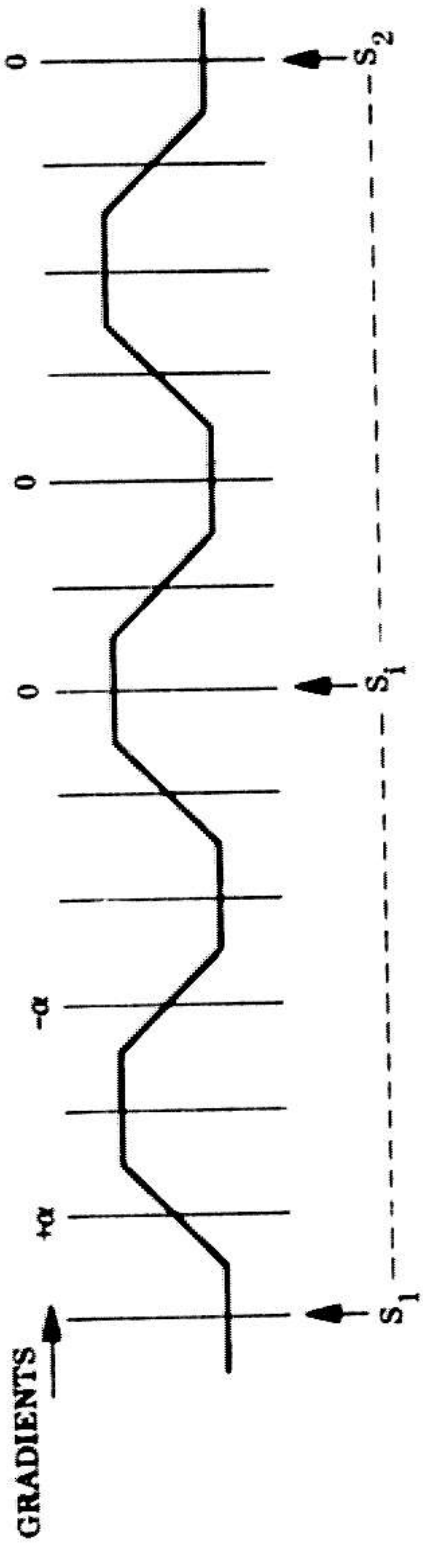
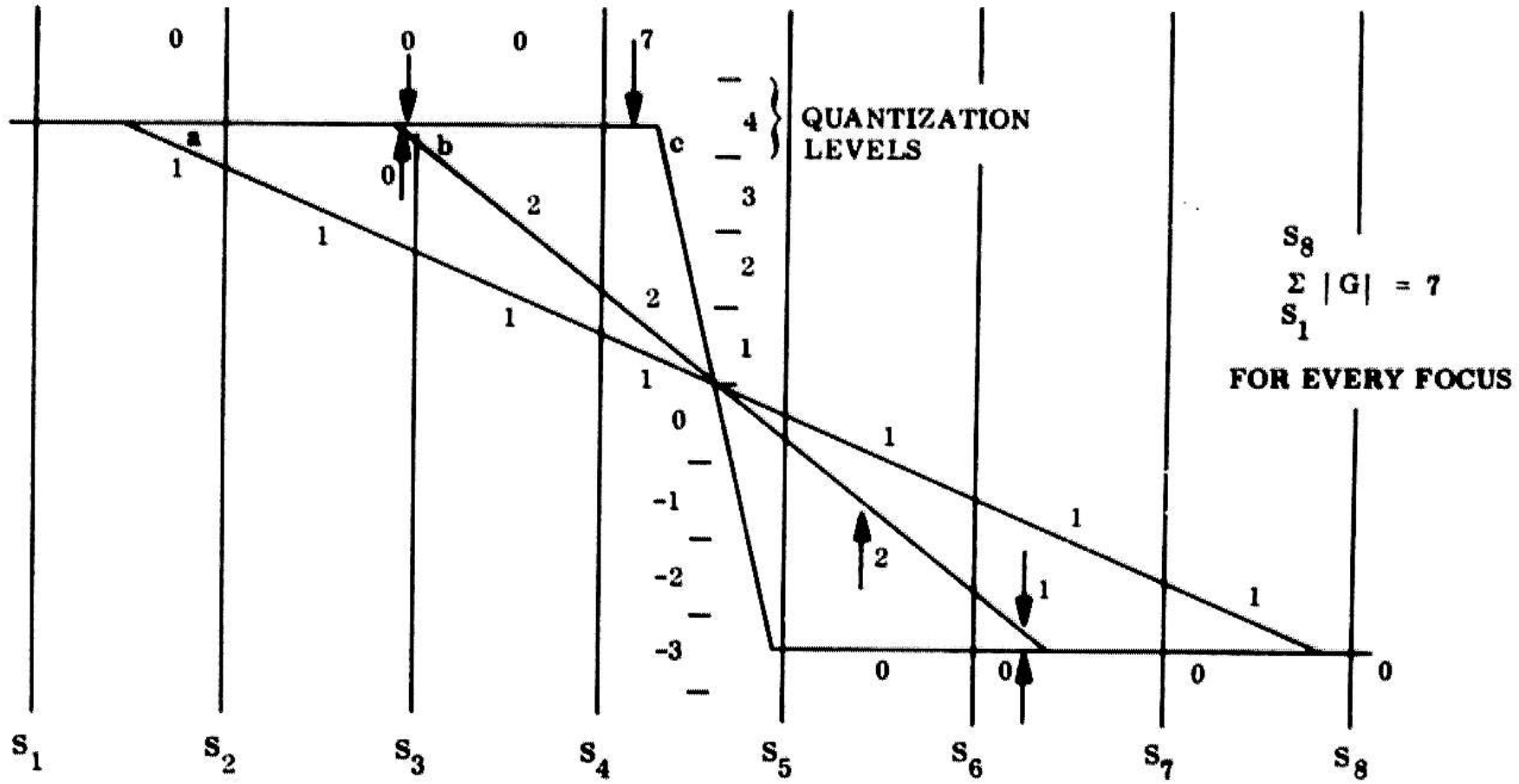


Fig. 6.5a Magnitude of Gradient Must be Used With a Textured Intensity Surface



Note: Numbers indicate local gradient between adjacent samples.

Fig. 6.5b Need to Threshold Magnitude Before Accumulating Gradient

because it is easier to obtain and introduces less noise.

The threshold controls the sharpness of the focus peak and will be treated as an accommodation.

A suitable one-dimensional focus criterion is thus given by

$$C = \sum_{s=S_L}^{S_R} G_T(s) \qquad (6,7)$$

where

$$G_T(s) = |G(s)| \quad \text{for} \quad |G(s)| \geq T$$
$$= 0 \quad \text{for} \quad |G(s)| < T \qquad (6,8)$$

with T=2, C evaluated over edges a,b,c (in Figure 6.5b) will yield scores of 0,6,7 respectively. (Note that the success of thresholding depends on the fact that an edge at any degree of defocus passes through a common fulcrum at the midpoint of each slope. This property follows from the convolution of an ideal step with a symmetric pulse. Were this not the case, C would not be guaranteed to increase monotonically with edge sharpness. The asymmetry of typical non-ideal edges found in practice is seldom large enough to affect monotonicity.)

The gradient operator developed by Sobel (Figure

5,5) is a direct two-dimensional extension of |G(s)|. It provides a directionally unbiased estimate of the magnitude of the spatial gradient. Sobel's operator is used in practice instead of |G(s)|, because edge orientation is often not known.

VI.4 FUNDAMENTAL AND ACCOMMODATABLE DETERMINERS OF RANGE UNCERTAINTY

The position of best focus for a plane surface is uncertain within the interval known as depth of focus. In our system the principal uncertainties result from amplitude and spatial quantization. We will examine how each of these factors limits the improvement in edge slope that can be detected. This analysis will lead to appropriate accommodations for minimizing range uncertainty.

VI.4.1 SPATIAL QUANTIZATION

Only the intensities in the transition region of an edge are affected by changes in focus. The width of this region is constantly narrowed as focus improves. At least one sampling point must intersect the actual transition to detect any change in focus quality. Close spatial sampling is thus essential to obtain a sharply defined focus maximum.

The most interesting case occurs near best focus

(see Figure 6,6). At a certain image sharpness the transition width, ∂t, becomes narrower than the interval, ∂s, between samples. From then on, one sample at most can coincide with the transition region. Assuming that the beginning of the edge transition is randomly placed within a stationary sampling interval, we can express the probability that the edge transition will coincide with a sample point,

$$\text{Prob (detect focus improvement)} = \frac{\partial t}{\partial s} \quad \text{for} \quad (\partial t < \partial s) \qquad (6,9)$$

A probability of ,5 can be taken as a reasonable cutoff below which further improvement in focus should not be expected. From Equation 6,9 we find the transition width corresponding to the threshold probability

$$\partial t|_{(\text{prob} = 0.5)} = \frac{\partial s}{2} \qquad (6,10)$$

VI,4,1,1 RELATION WITH CIRCLE OF CONFUSION

In Chapter 2 we expressed the resolution limit of our system in terms of a circle of confusion. We can now express the limitations of spatial quantization in an equivalent form. A consistent definition of "circle of confusion" for this system is: the image of a point source

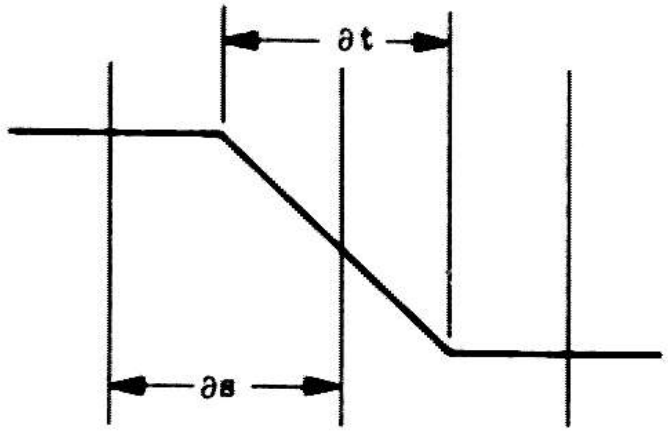


Fig. 6.6 Limiting Spatial Resolution Near Best Focus

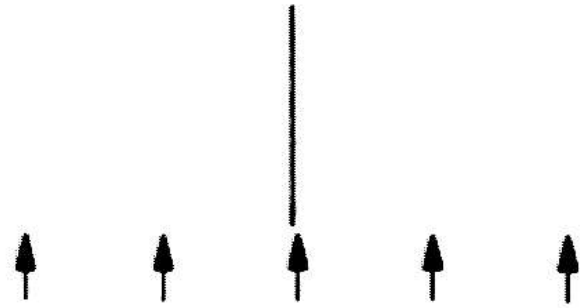


Fig. 6.7a Sharply Focused High Frequency Image

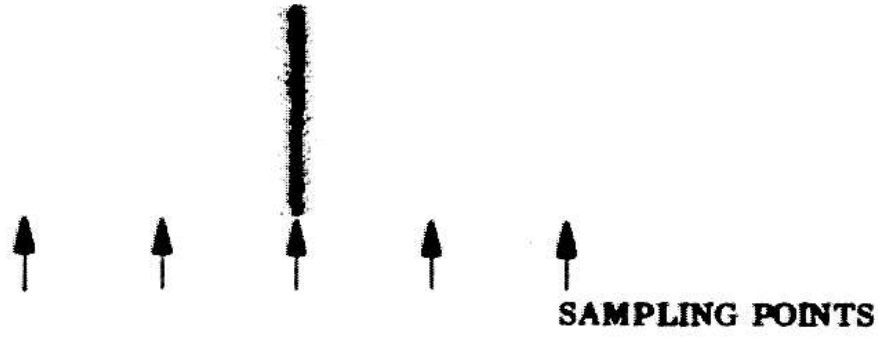


Fig. 6.7b Unfocused Line

defocused to the extent that its diameter, c, equals the minimum edge width given by 6.10.

VI.4.1.2 ACCOMMODATION TO MINIMIZE DEPTH OF FIELD

Depth of field is the conventional measure of uncertainty in object space. From Equation 6.10 and 2.55c,

$$D_{foc} = \frac{x \cdot d \cdot f \cdot \delta s \cdot (x - f)}{d^2 f^2 - \left(\frac{\delta s}{2}\right)^2 (x - f)^2} \approx \frac{x \cdot \delta s \cdot (x - f)}{df} \qquad (6.11)$$

The simplification holds in the usual case when the lens diameter, d, is much greater than the sampling interval, δs. D.foc then grows directly with object distance and inversely with focal length (as determined in Chapter 2). Figure 6.6 provides an interesting way to visualize these results in terms of sampling limitations.

A defocused perfect edge will have a finite transition width and a corresponding probability of intersection with a sampling point. The width represents an attenuation of the frequency content found in an ideal edge. A longer lens will magnify the width of the transition (corresponding to a given frequency content) relative to the stationary sampling interval. This multiplies the probability of intersection for a given degree of unfocus. A similar magnification is realized by moving the object

closer to the lens.

The theoretically minimal range uncertainty at three typical object distances, is tabulated in Chart 6.1 for the 1, 2, and 3" lenses. (The iris is wide open f#=1.4 in each case). To summarize, at any object distance, the smallest range uncertainty is achieved by selecting the longest lens.

To complete the discussion on spatial quantization, it is appropriate to discuss the possibility of aliasing errors due to undersampling. Consider an image with significant spectral power at spatial frequencies much greater than 1/δs. The thin, high contrast, vertical line in Figure 6.7a illustrates this condition. There is virtually no chance that a sample will coincide with the line when the image is highly focused. Consequently, the focus criterion will be zero. However, as the line broadens under progressive defocusing, there is an increasing likelihood of detecting a gradient. The heuristic focus criterion has clearly failed. The problem, however, is due to insufficient sampling. A criterion based on Fourier analysis would also decrease when the image frequencies exceeded the bandwidth of the sampling system. In practice, the limited bandwidth of the video amplifier insures that no image will grossly exceed the sampling capacity of the system.

Range	Lens		
	1"	2"	3"
25"	1.34"	0.322	0.136
35"	2.67	.647	.279
45"	4.45	1.08	.470

Chart 6.1 Theoretically Minimal Depth of Field ($f_{\#} = 1.4$)

VI.4.2 AMPLITUDE QUANTIZATION

The ability to detect changes in focus quality requires not only that a sample point intersect an edge transition, . There must also be sufficient amplitude resolution to detect the small intensity change that indicates a further improvement in edge slope. In Figure 6.8 edge b is barely distinguishable from edge a at the indicated intensity resolution. Any edge whose upper breakpoint fell in the range, $\partial W+1$, would be indistinguishable from edge a, because both would pass through quantization interval 1 at $S+2$. $\partial W+1$ represents the minimal improvement from the focus quality of edge a that can be detected at this intensity resolution. Similarly, $\partial W+2$ establishes the ultimate uncertainty of best focus. The focus criterion will be perfectly flat for all edge transitions between those labeled c and d.

It is clear that these uncertainties can be reduced by increasing the amplitude resolution. We know that better focus will not change the intensities at $S+1$ and $S+3$. It therefore seems appropriate to narrow the quantization window, concentrating all available resolution at the intensity level where the transition intersects a sample point. Edges a and b (from 6.8) have been redrawn in Figure 6.9a. Because of the narrowed intensity window, the focus criterion should now be able to distinguish the difference

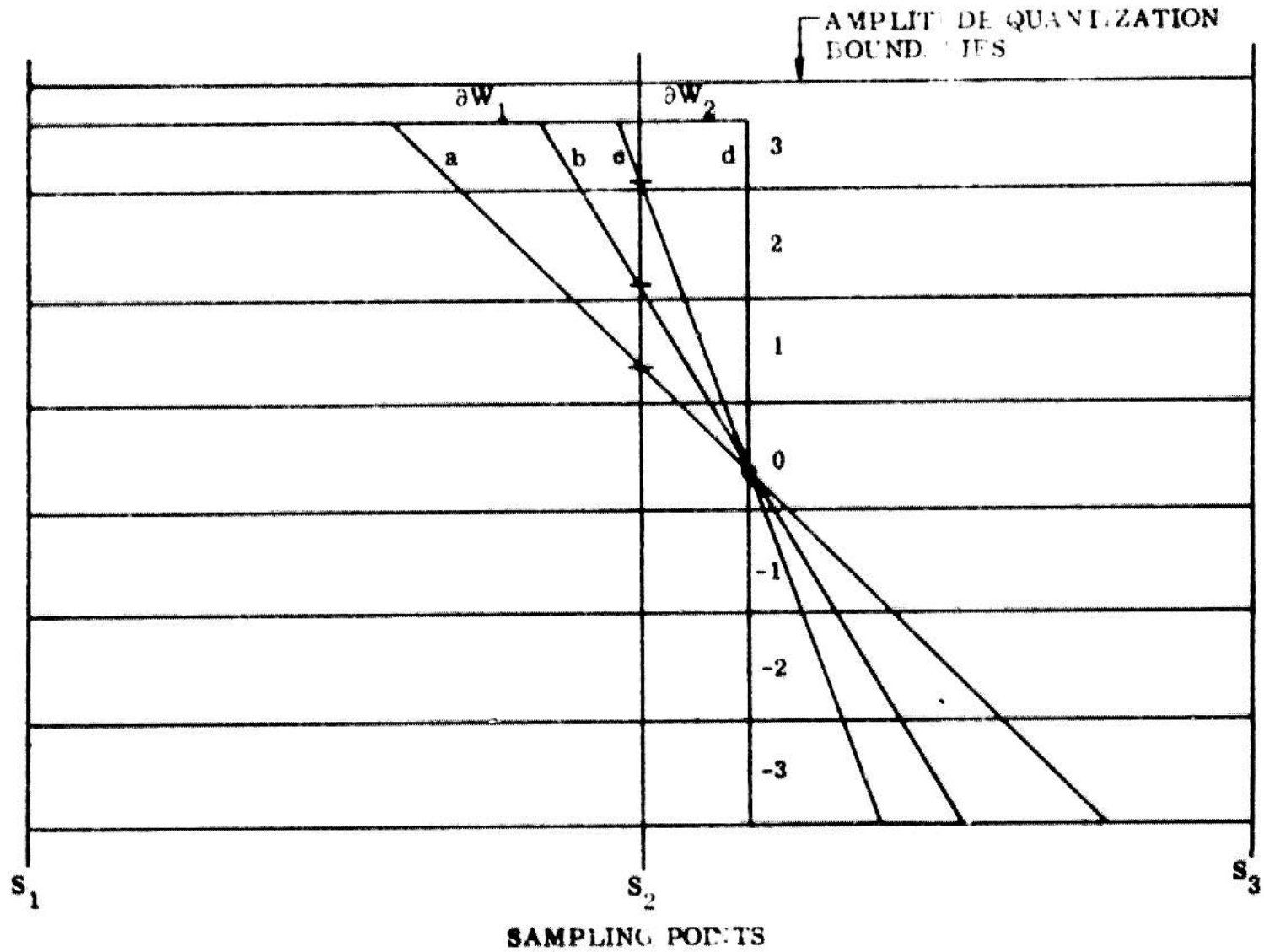


Fig. 6.8 Limitation of Amplitude Quantization

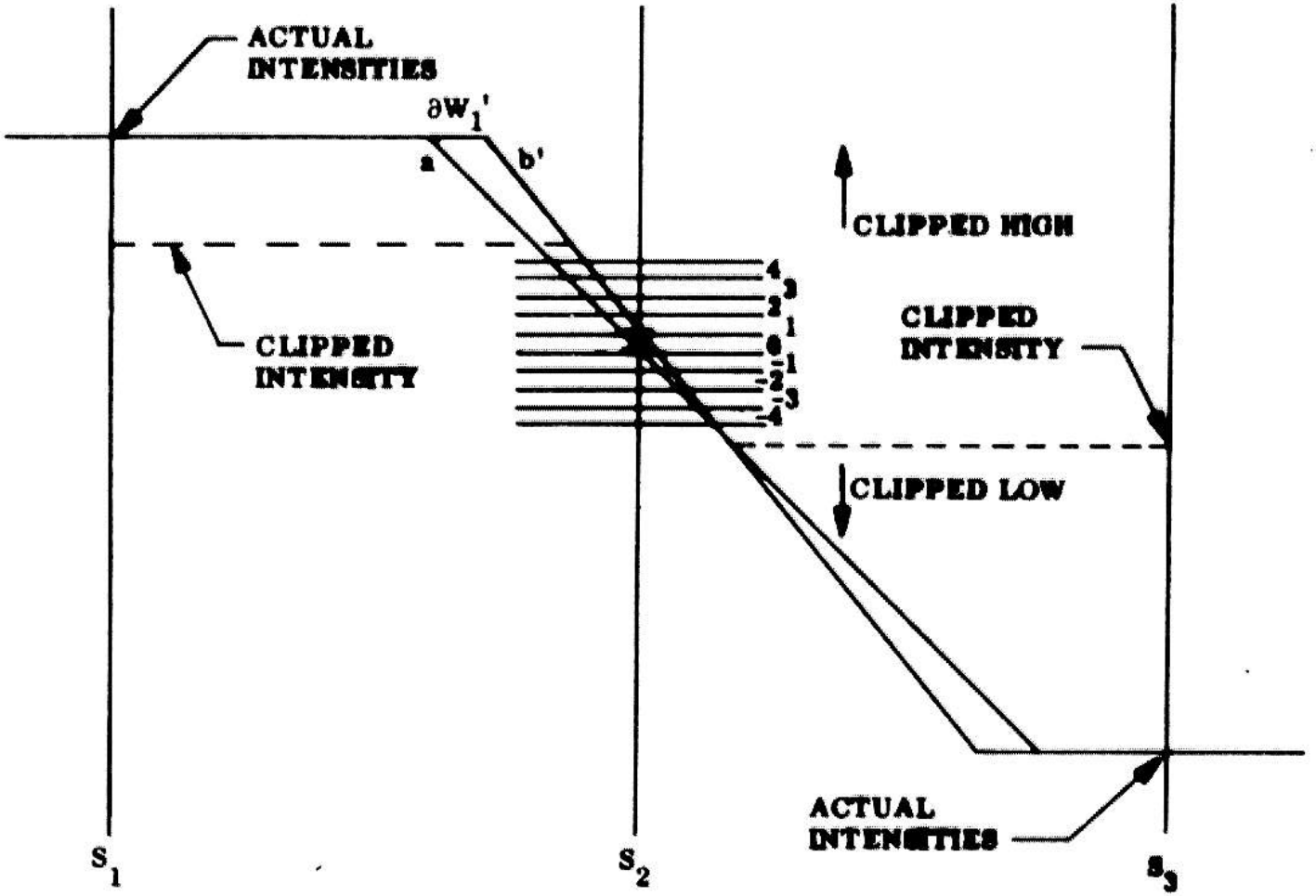


Fig. 6.9a Increasing Focus Sensitivity With More Intensity Resolution

in image sharpness represented by $\partial W'_{i,1} << \alpha W_{i,1}$.

The above sentence contains an important hedge. What the narrowed intensity window does provide is the ability to detect a change in intensity at $S_{i,2}$ between edges a and b'. If it were known that $S_{i,2}$ was sampling the transition of an edge and if the actual intensities at $S_{i,1}$ and $S_{i,3}$ were previously recorded, then the correct value of the focus criterion could indeed be calculated with great precision. In cases in which an edge cannot be assumed it is important that the focus criterion not be blindly applied to the hard clipped intensities at $S_{i,1}$ and $S_{i,3}$. Hard clipping destroys all guarantees that the resulting focus criteria will increase monotonically with focus.

Figure 6.9b shows why hard clipping is bad. A hypothetical, 5 level quantization window was centered to maximize the sensitivity to variations about line b at $S_{i,2}$. (Levels 1 and 5 correspond to hard-clipping.) The limited quantization range has shifted the effective fulcrum; the central pivot condition, required to insure the monotonicity of criteria 6,6, is thus violated. With a threshold of two, the value of the cumulative gradient at $S_{i,1}$, $S_{i,2}$, and $S_{i,3}$ actually decreases (from 8 to 7) as edge be is transformed by better focus into edge c.

These examples demonstrate the importance of an appropriate quantization window for sharp focusing. The narrowest window that does not hard clip any intensities in

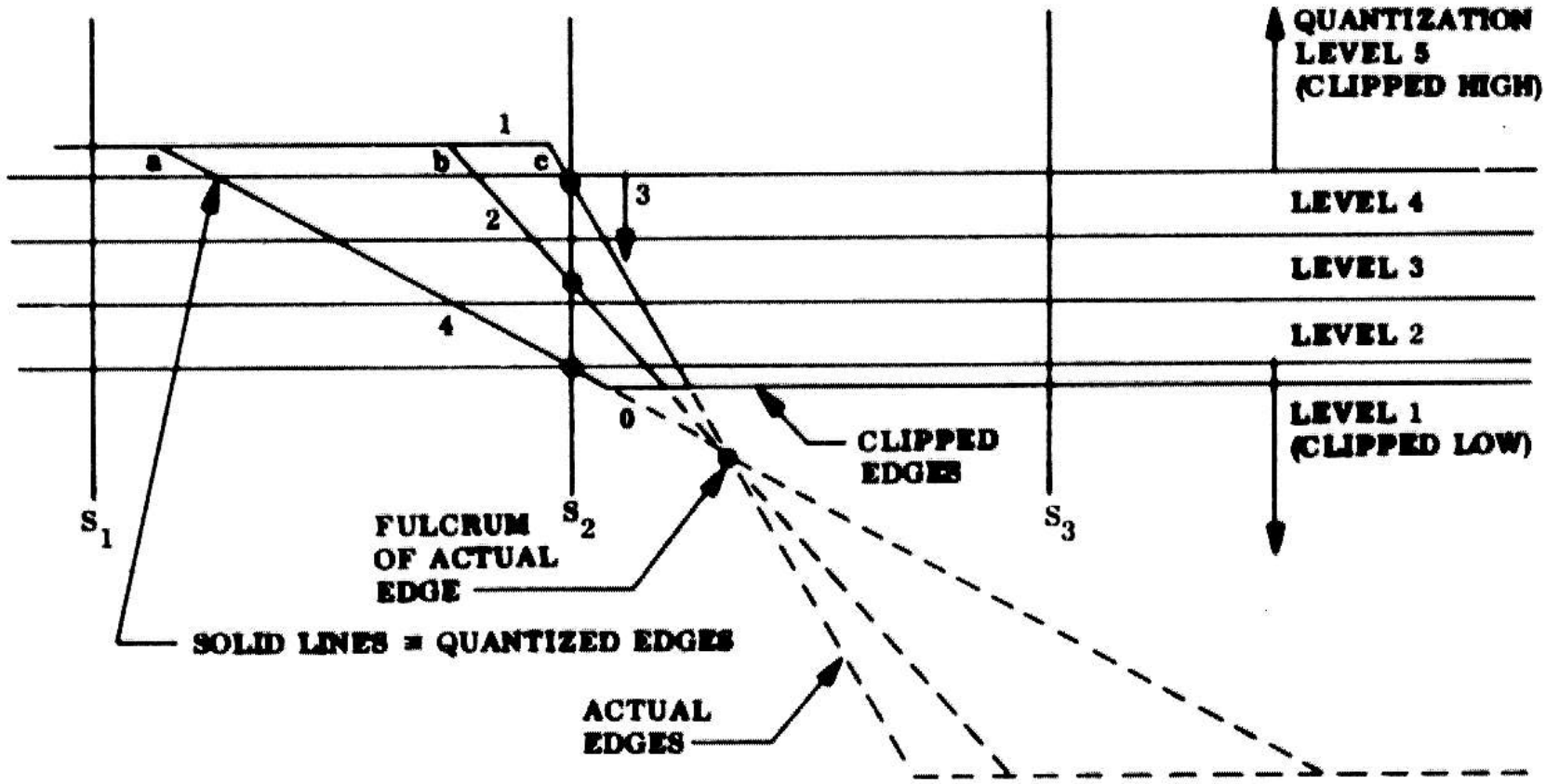


Fig. 6.9b Inappropriateness of Hard Clipping for Focus Evaluation

the region of interest is the most generally applicable accommodation (see Figure 6.10). This compromise maximizes the resolution that can be attained without risking local maxima of the focus criteria.

In applications requiring the utmost accuracy, the effective resolution can be increased using the quantizer sub-ranging scheme mentioned in Chapter 2. This method attains 6.5 bits of resolution for intensities over the full dynamic range but at a considerable overhead in processing time. Figure 6.11 is a comparison of the relative sharpness of the focus criterion attainable with representative quantizer windows.

VI.5 EFFECT OF SCENE CONTENT ON SHARPNESS OF FOCUS

Thus far, we have studied the inherent system characteristics that limit attainable range accuracy and have proposed accommodations to optimize performance subject to these constraints. The characteristics of the scene pose another constraint over which there is no control. However, it is useful to know, at least in general terms, how these characteristics affect performance.

VI.5.1 OBJECT CONTRAST

It is easier to focus on areas containing high

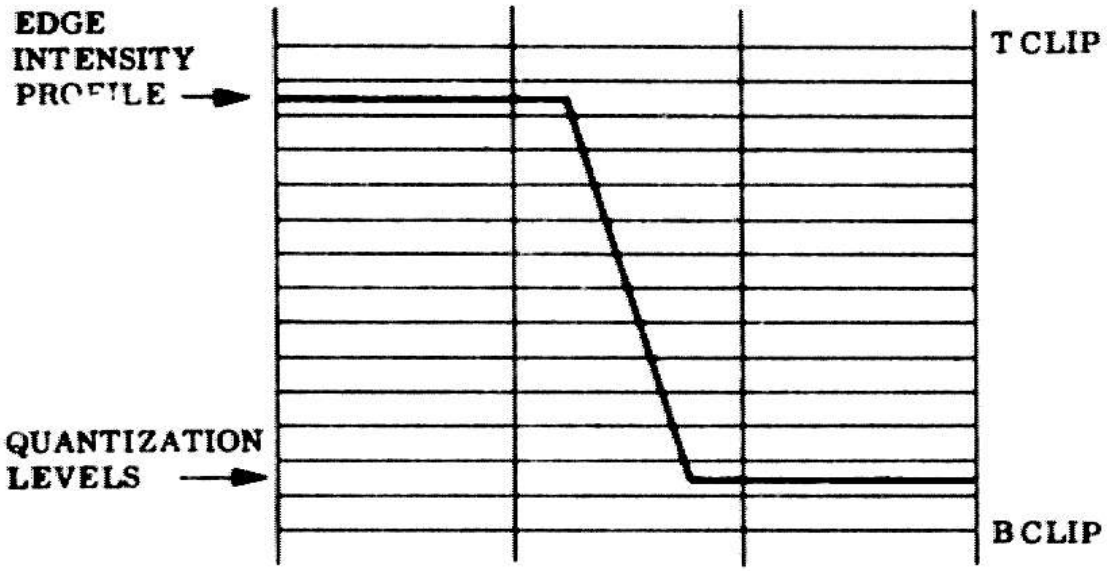


Fig. 6.10 Compromise Quantization Window for General Purpose Focusing

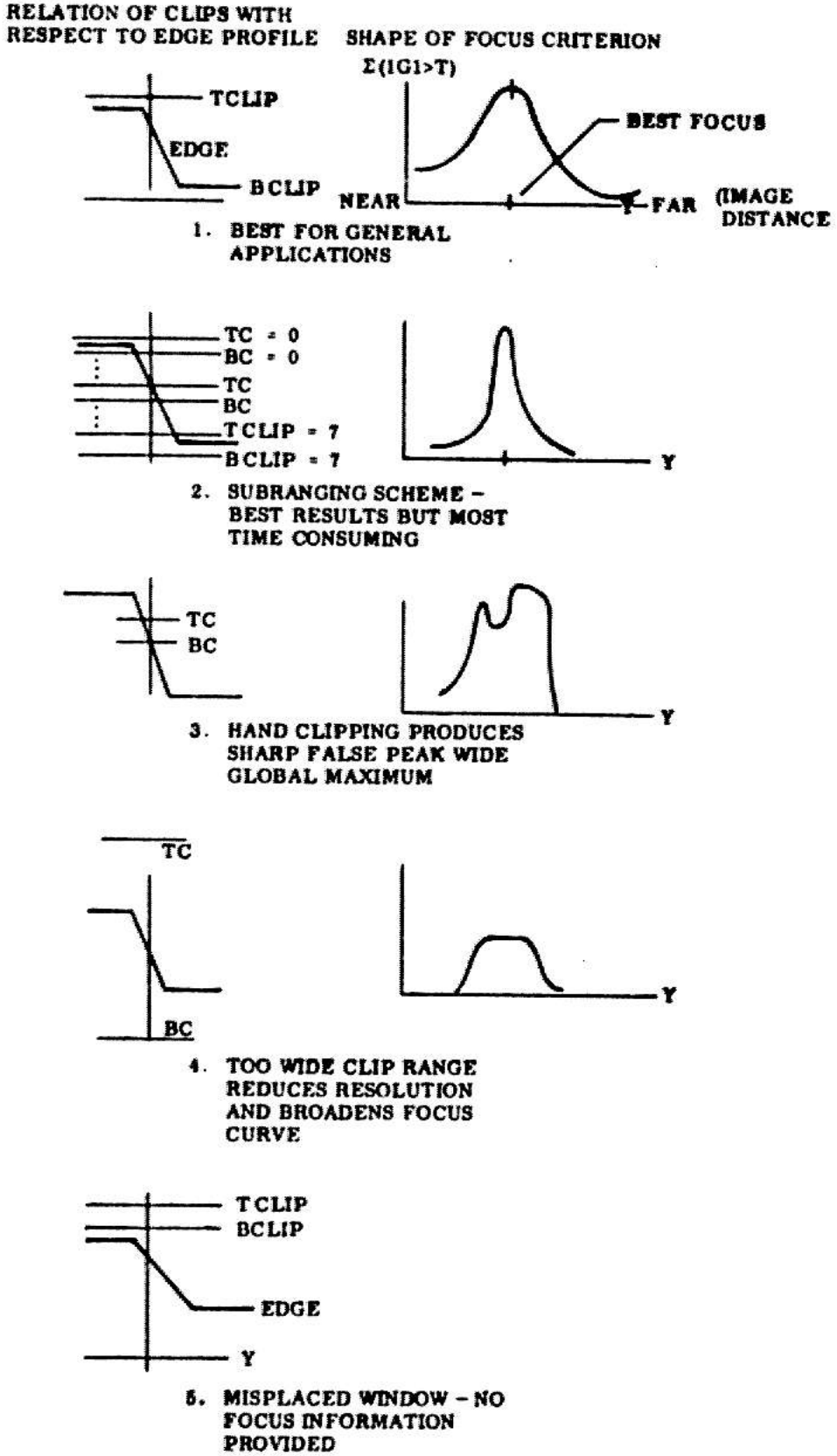


Fig. 6.11 Relative Sharpness of Focus Criteria for Various Quantizer Accommodation

contrast. Figure 6.12 shows two step edges (E↓1 and E↓2) at the same amount of defocus, indicated by δt. To improve focus requires the ability to detect changes in intensity at S↓2. δt varies linearly with changing image distance. For any decrease in δt, $\delta I_{\downarrow 2}$ will decrease (E↓2/E↓1) times as much as $\delta I_{\downarrow 1}$.

For a given quantization resolution, δt need change on the average (E↓1/E↓2) less to detect a minimal intensity change at S↓2 with the high contrast edge. This advantage does not apply when using clipping window $\#_1$ (see Figure 6.11), since the quantization fineness also decreases inversely with the height of the edge. The increased sensitivity, however, would be realized by the sub-ranging scheme.

A gradient is a noise sensitive operation. High contrast provides desirable signal/noise advantages. Consequently, all of the accommodations used to enhance contrast for verification (eg. high sensitivity, color filter,etc.) are also appropriate to employ in focusing.

VI.5.2 OBJECT TEXTURE

A uniform surface provides no information regarding the quality of focus. At the other extreme, a surface covered with high frequency, highly contrasting texture would yield sharply peaked focus optima. (Distributed

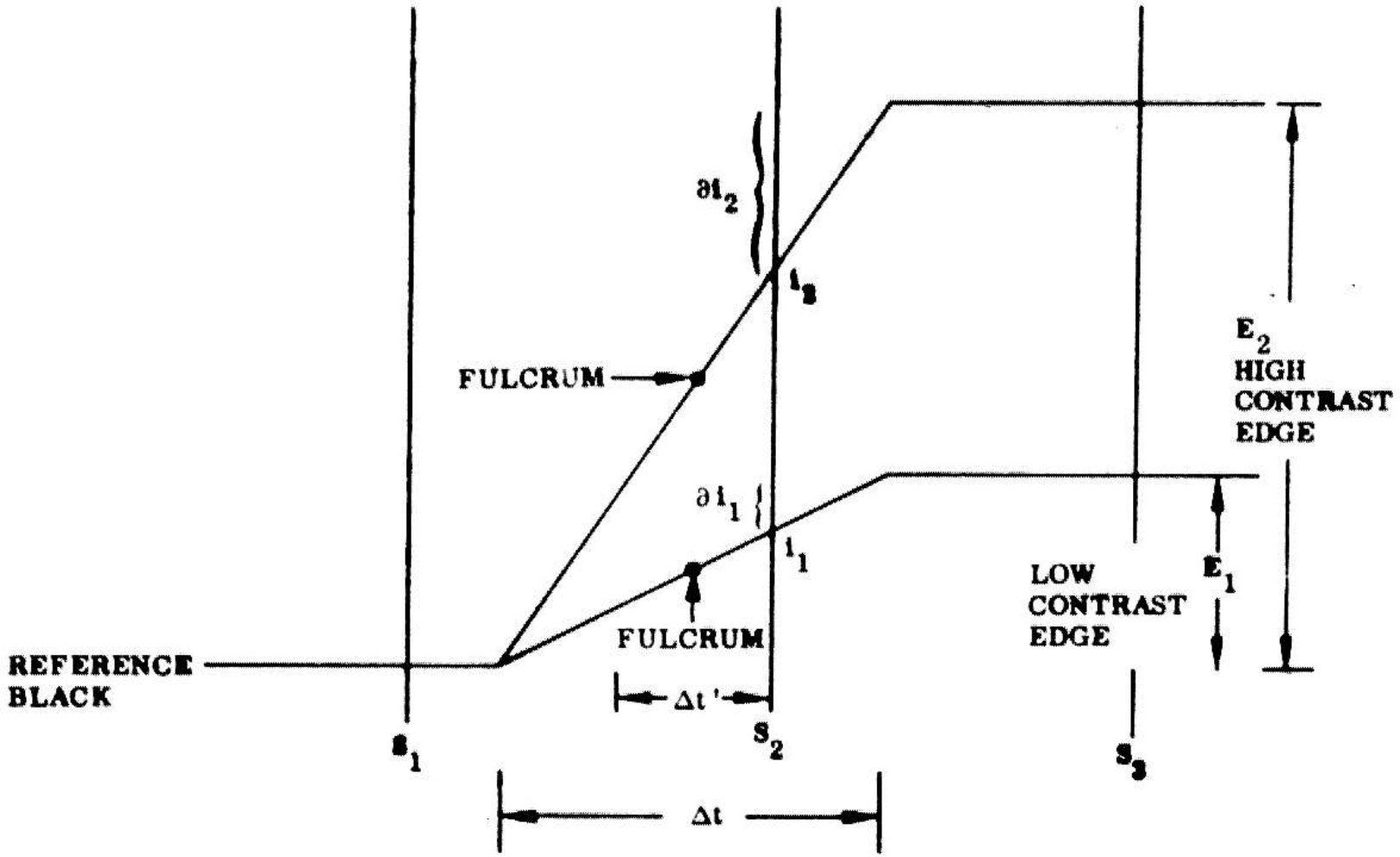


Fig. 6.12 Effect of Object Contrast

texture increases the likelihood that samples of a given density will intersect transition regions.) This influence of object texture can be quantified in the frequency domain.

Defocusing is a low-pass filter. The frequency response of this filter can be determined from the Fourier transform of its impulse response. In one dimension, the transform of a unit amplitude pulse of width δt (symetrically placed about the origin) is

$$\frac{2 \sin\left(\omega \frac{\delta t}{2}\right)}{\omega} \qquad (6.12)$$

(Goodman [1969] generalizes this analysis to two dimensions, using the Fourier transform of a disk of radius δt.)

The effective passband of this function can be approximated by the frequency range

$$0 \leq \omega \leq \frac{2\pi}{\delta t} \qquad (6.13)$$

As focus improves, δt narrows, widening the passband. More high frequency texture components will be passed by the optical low pass filter, boosting the focus criterion. If the object is composed primarily of low frequency textures, like a cloud, it is clear that when the filter passband ($2\pi/\delta t$) exceeds the highest frequency at which significant textural energy is found, finer focusing will not improve

the image.

The maximum frequency response of the optical system is set by the diameter of the limiting circle of confusion (Equation 6.10).

$$\omega_{max} = \frac{2\pi}{\frac{\partial s}{2}} = \frac{4\pi}{\partial s} \qquad (6.14)$$

We conclude that objects whose principal spectral energy lies below this frequency will have broad focus peaks (and, correspondingly, more depth uncertainty) than would be expected from Equation 6.11. If all significant object frequencies lie in the range $0<\omega<\omega_{mo}$, a more accurate bound on focus uncertainty can be established.

$$D_{foc} = \frac{2 x d f \frac{\partial s}{2} (x - f)}{d^2 f^2 - \left(\frac{\partial s}{2}\right)^2 (x - f)^2} \quad \text{for} \quad \omega_{mo} \geq \omega_{max} \qquad (6.15)$$

$$= \frac{2 x d f \left(\frac{2\pi}{\omega_{mo}}\right) (x - f)}{d^2 f^2 - \left(\frac{2\pi}{\omega_{mo}}\right)^2 (x - f)^2} \qquad \omega_{mo} < \omega_{max}$$

VI.5.3 OBJECT DEPTH

For simplicity, we have thus far restricted attention to planar, textured surfaces lying at a

well-defined depth. We next consider where the focus criterion will peak when textures lie at several depths within the field of view (ie. an edge that extends away from the lens). Our previous results can be generalized to treat this case, using superposition.

At every focus position the imaged diameter of a point source at each depth in object space is well-defined. To extend the previous analysis, simply partition the field of view by planes perpendicular to the lens axis at regular intervals of depth. The focus criterion is evaluated for all scene components, using the pulse width appropriate to the nearest depth plane. After accumulating the thresholded magnitude of the gradients associated with each plane, a sum of these sums is taken over all depths to define the goodness of focus at this image distance.

This criterion will tend to maximize the overall detail in an image. However, close details and high contrasts are weighted heavily. The focus peak, of course, will be wider than if all detail were at a single range.

VI.6 FOCUS RANGING PROGRAM

In this section an automatic focusing program is described. It utilizes the accommodation considerations discussed in the chapter to obtain the most well-defined focus peak. The best focus is used to estimate the range to

the field over which the focus criterion was accumulated.

The focus program is called with the following parameters:

1. h,v: the horizontal and vertical raster coordinates of a feature whose range is desired. (These indices and the coordinates of the lens center define a ray in space that intersects the feature.)

2. x[est]: estimated range (based on support hypothesis, geometric inference, etc.).

3. dx[est]: maximum uncertainty of x[est]: x[est] and dx[est] are used to bound the initial search for a focus maximum. They will be refined by the program as the interval thought to contain the best focus is narrowed down.

4.dx[req.]: required range accuracy. This is the termination criterion that is applied to dx[est].

5. cost: a measure of the maximum effort that should be exerted in trying to attain dx[req.]. Like the verifier, the current version of the focus program does not test this condition. It is included in anticipation of the time when a cost effective system strategy and accurate cost estimates of the various accommodations are available.

6. region description: This parameter is actually intended as a pointer to a list of attributes describing what may be known about the feature to be focused upon. This knowledge would be helpful, for example, in selecting a focus criterion particularly suited to the feature (eg. An

edge operator could be tailored to the contour of an edge.). It could also indicate the desirability of special accommodations, like a filter change, if a color edge were specified. The focus program currently accepts only the horizontal and vertical dimensions of the region (centered at h,v) over which the focus criterion is to be maximized.

VI,6.1 NEED FOR BOOTSTRAPPING

dx[req,] will usually be less than 1/2". At typical working ranges this accuracy requires a longer lens than the 1" (wide angle) unit commonly used for general surveillance. The length of the lens needed to attain the specified range uncertainty can be found by solving 6.11 at x=x[est]. However, it is generally not possible to switch directly to a longer lens before performing a coarse search for the focus peak.

The problem is that, while a long lens is needed to get accurate range, fairly accurate range is needed to change lenses. More precisely, a sufficiently good depth estimate is needed in order to locate and perhaps re-center the desired feature in the new field of view (Gill [forthcoming]). The longer the lens, the narrower the field of view, the better the depth estimate that is required. Fortunately, the attainable range uncertainty also decreases with longer lenses. This situation suggests that a

bootstrapping approach might be used to establish the focus optimum.

Starting with the short lens, a coarse search for the focus maximum is conducted. This search involves evaluating the focus criterion at perhaps 10 image distances covering the initial range uncertainty dx[est]. The focus peak found on this initial trial is used to narrow the focus uncertainty enough for the feature to be re-centered in the field of view of the next longest lens. Proceeding in this way, 10 more samples are taken over the remaining uncertainty interval. The refined peak will then permit a change to an even longer lens, if that is needed to achieve dx[req.]. A flow chart of this process is given in Figure 6.13.

Bootstrapping enables the computer to overcome its lack of global comprehension about what it is focusing on. A person, by comparison, relies strongly on this asset to manually re-center the camera after changing a lens. It enables him, in conjunction with his human adeptness for real time servoing, to track the outline of a feature on the television monitor, until the feature is centered. A human is thus able to position the camera without a precise range estimate and even when the new lens is moderately unfocused.

(A computer equipped with a continuously variable focal length (zoom) lens might be able to keep the feature centered by tracking it as focal length was gradually

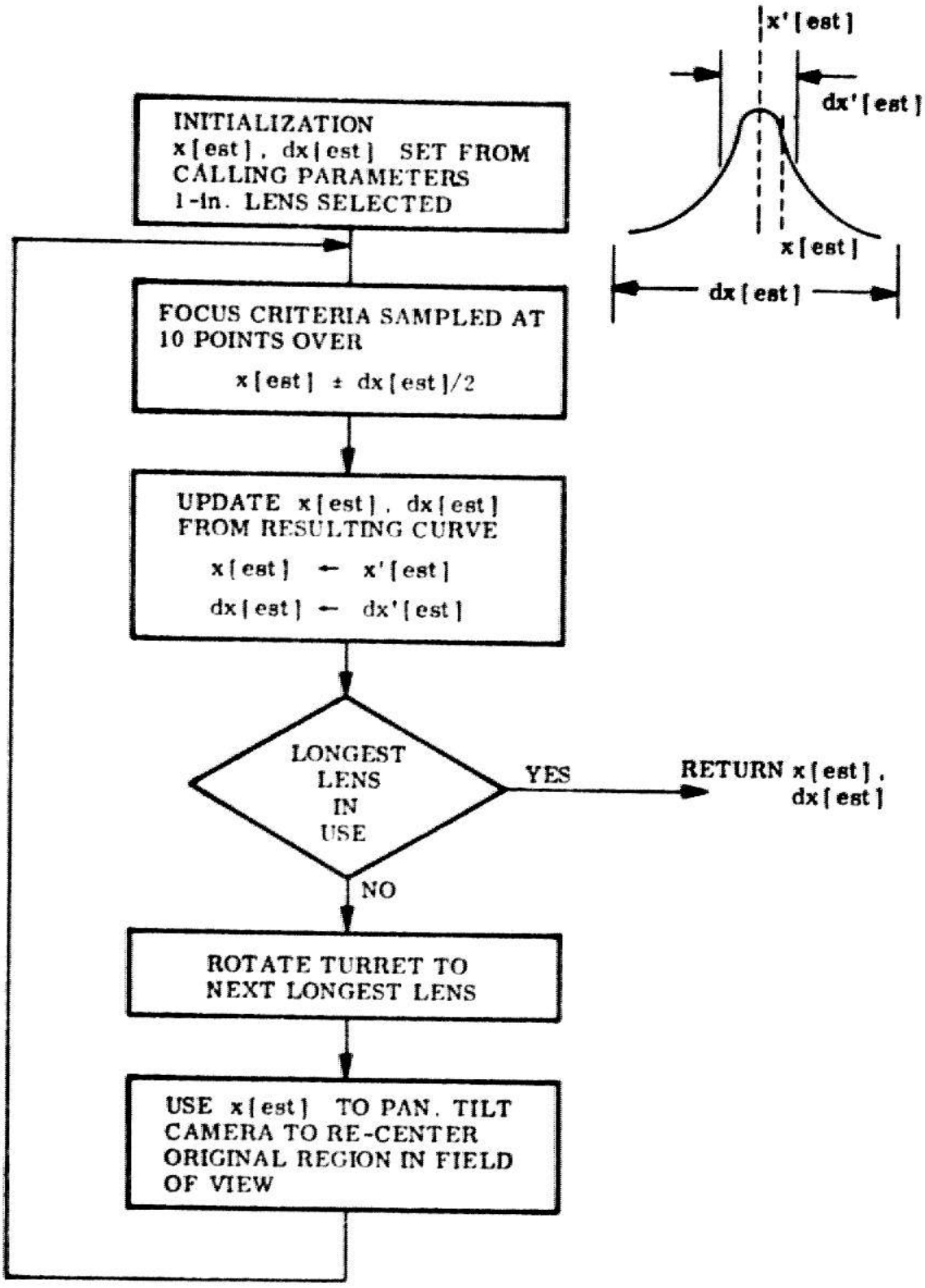


Fig. 6.13 Flow Chart of Basic Range Refinement Bootstrapping Cycle

increased. This use of servoing would improve the efficiency of the bootstrap loop.)

VI.6.1.1 ADVANTAGES OF BOOTSTRAPPING

Bootstrapping has many features that are desirable in an accommodative perceptual strategy. The most careful focusing is only performed over the narrow interval to which the global optimum has been previously localized by inexpensive, coarse searches. When the initial uncertainty is greater than a few inches, it is also more efficient to do the initial localization using a shorter lens. Suppose a 3" lens was used to narrow the focus optimum from an initial 1' uncertainty. The relative sharpness of the focus peak observed with a 3" lens would necessitate considerably more than 10 samples to insure that the true global maximum was found.

The focus strategy also contains a second example of bootstrapped optimization. Recall that the sharpness of the focus peak depends strongly on the clips, sensitivity, and operator threshold. These parameters must be set according to the observed intensities. However, the intensity range is not well-defined in an out of focus scene. We approach the globally optimal settings for these accommodations by refining them in terms of the scene characteristics observed at each peak. As focus improves, so will these

accommodations, thus facilitating an even sharper focus.

VI.6.2 THE IMPORTANCE OF PRECISE FEATURE SPECIFICATION

The focus program is very costly in terms of real time. It takes almost 30 seconds to move the vidicon over the .5" length of the threaded focus drive. Consequently, focus is used mainly to resolve specific recognition ambiguities. (In this regard, Falk [1970] formalized the useful result that the depth at a few selected points on a planar-faced object constrains the depths at all other points.)

Both the required time and attainable depth uncertainty are inversely related to the precision with which the calling parameters bind the initial search range. The size of the initial uncertainty, dx[est], determines, for instance, what lens to use for the first search. If dx[est]≤1", the 3" lens can be used directly, eliminating up to two cycles of bootstrapping for each shorter lens. In general, the more that is known a priori, the less bootstrapping is necessary.

In the absence of a precise feature specification the size of the region over which the focus criterion is accumulated has a profound affect on the shape of the focus peak. Several factors are involved, all related to the fact that the computer has no high level conception of the

feature within the region that is actually of interest. Signal/noise will be compromised if, because of uncertainties in localization, the region is specified much larger than the actual feature. For example, suppose a vertical edge passes somewhere through a wide rectangular region. Near good focus only those raster samples directly adjacent to the actual discontinuity will contribute useful information. All of the other gradients, calculated over the boundary surfaces, contribute only noise. We plan to eventually use the available knowledge about a feature to tailor the region over which the focus criterion is applied.

In the meantime, the focus program does achieve some discrimination of the gradients that enter into the sum by controlling the gradient cutoff threshold. This accommodation is included as part of the bootstrap sequence. Far from focus, all gradients will be small. Therefore, the cutoff is initially set low. At each focus peak, the highest gradients will be those associated with the dominant feature. On the assumption that this feature is the desired one, the threshold is raised to eliminate all gradients more than 25% below the maximum values recorded at the peak (An overly large region entails the risk that undesired features will be included.). As focus improves on successive cycles, the peak gradients will get bigger, justifying yet higher thresholds. This accommodation contributes to a sharply defined, low noise focus peak.

If the feature does not lie entirely in a single depth plane, the focus peak will be unnecessarily broadened by a large region. Consider focusing on the interior vertex of a cube (see Figure 6.14). This feature is defined by the intersection of three planes. Each of these edges slopes away from the camera. Thus, the larger the region, the larger the inherent depth uncertainty. Of course, too small a region is also undesirable because of insufficient samples for noise-smoothing.

We have considered but not yet implemented the possibility of resolving this conflict by adaptively accommodating the size of the region. Initially, when focus is poor, a larger region could be used to build up a significant sample and avoid noise. As focus becomes better defined, the gradients at the actual vertex will get larger. The region can then be compressed about the point (h,v) in a manner that will enclose the maximum gradient in the smallest area. This will decrease the range uncertainty contained in the region for the next iteration. The field of view in object space corresponding to a constant raster area depends on lens magnification. Magnification is proportional to focal length. Consequently, the bigger the lens, the narrower the field of view through a given area of the raster. This relationship tends to reduce inherent scene ambiguities as the focus peak is approached.

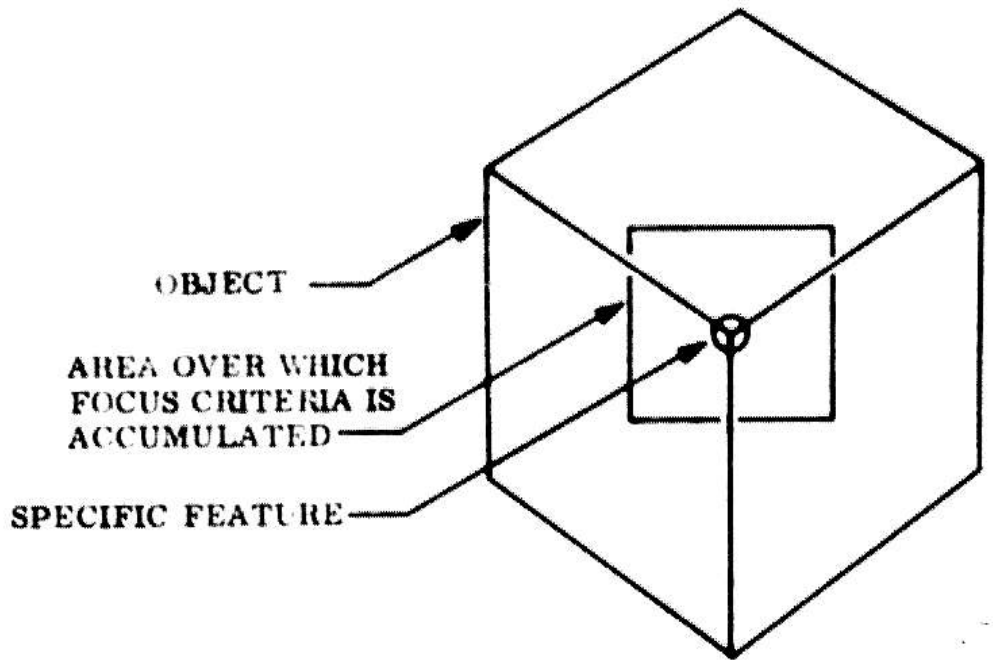


Fig. 6.14 Depth Uncertainty Introduced by Finite Window Size

VI.6.3 DETAILED DISCUSSION OF FOCUSING PROGRAM

The focus ranging program is based on the simultaneous bootstrap refinement of the range estimate and of all accommodations contingent on the quality of focus. The program is flow-charted in Figure 6.15. On entry, the initial range and uncertainty estimates, x[est], dx[est] are used with Chart 6.1 to select the shortest lens whose depth of field at x[est] is less than dx[est]. (This will often be the 1" lens.)

The focus criterion is then evaluated at 10 image distances corresponding to the range x[est] +/- dx[est]/2. A refined estimate of dx is determined from the shape of the focus peak (see Figure 6.16). If dx'[est] is smaller than the required uncertainty dx[req.], the program returns its latest estimate. If not (after satisfying a hypothetical cost constraint), accommodations are refined, based on the scene characteristics observed at the current focus maximum. If a positive improvement in the width of the peak can be expected as a result of an accommodation, the focus curve is recompiled over the refined uncertainty. Otherwise, the program exits, short of its goal.

The focus program, like the verifier, fits well into the basic accommodation paradigm described in Chapter 1. This similarity is emphasized by the dashed boxes in Figure 6.15. The sharpness of the focus peak might thus be

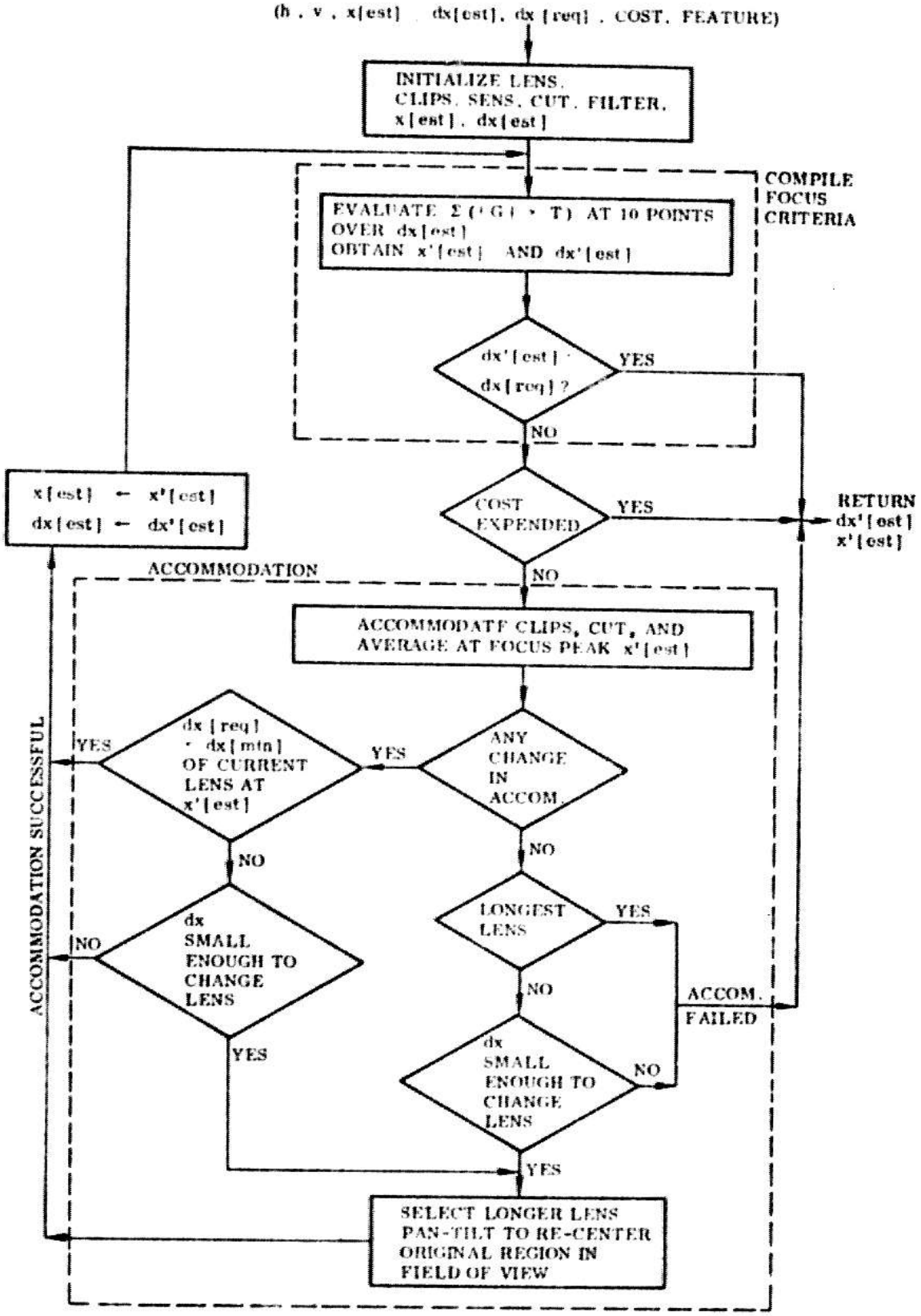


Fig. 6.15 Flow Chart of Focus Ranging Program

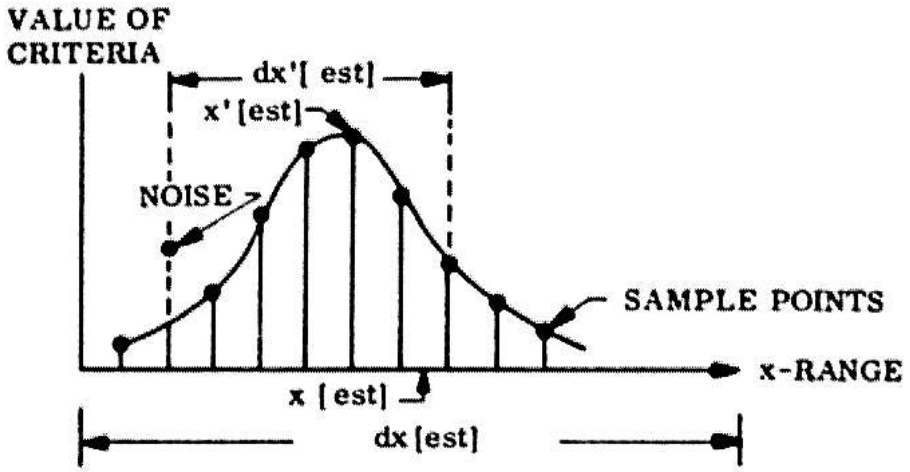
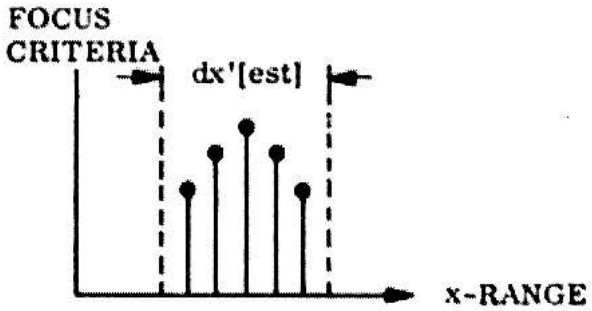
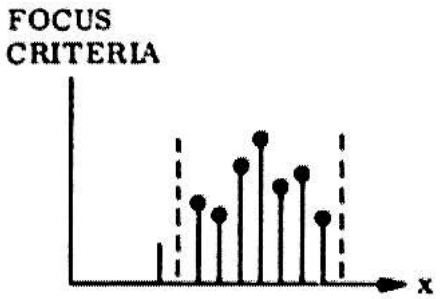


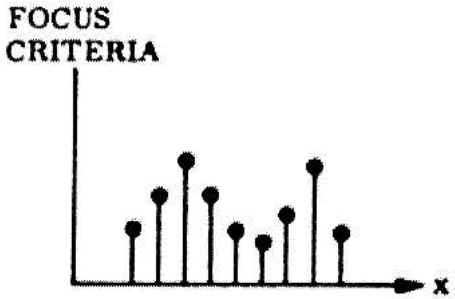
Fig. 6.16 Empiric Determination of x'[est] and dx'[est]



a. NO NOISE ≡ 3 HIGHEST SAMPLES ARE ADJACENT; ≤ 3 SLOPE REVERSALS OVER dx'[est]



b. NOISY



c. MULTIPLE DEPTHS (NOT NOISE)

Fig. 6.17 Determining the Noisiness of The Focus Peak

considered as a high level criterion of accommodation. The narrower the peak the more appropriate the acommodation. The similarity with the verifier, moreover, is especially pronounced when the feature is an edge; with the exception of the quantizer window, the accommodations responsible for sharpening the edge and sharpening the focus peak, are identical.

VI.6.4 DISCUSSION OF ACCOMMODATION STRATEGY

The principal accommodation considerations have already been elaborated. Here, we need only summarize how these considerations fit into the overall strategy. When the empirically determined width of the focus maximum is wider than required, it is first determined whether a sharper peak can be obtained with the present lens, providing other accommodations are first optimized:

1. The camera is refocused at the previously observed maximum,

2. The validity of the quantizer window used on the previous search is re-established.

3. The gradient threshold is raised, if necessary to 75% of the maximum individual gradient recorded at peak focus,

4. The noisiness of the focus peak is determined (see Figure 6.17) by the number of local maxima of the focus

criteria over the interval dx'[est]. Theory states that over short intervals the focus curve should be monotonic with range. Thus, if the new focus peak is "well-defined", temporal noise should be suspected, when there are more than three slope reversals over the new peak (dx'[est]) or the three highest criterion levels are not adjacent in range. In this event, the program will increase by one, the number of television frames that are averaged together at each focus position before the criterion is evaluated.

Figure 6.17c illustrates the notion of an ill-defined peak; two local maxima are further apart in range than the depth of field of the lens. Rather than temporal noise, this condition indicates that two distinct features are in view. dx[est] must be more tightly constrained from a higher level.

This use of averaging is an example of what might be called "heuristic filtering". The desired focus curve is known to be monotonic and smooth. Departures from this ideal indicate the need for corrective action. This approach is particularly efficient, because the expense of averaging is confined to ranges near the optimum where a precise evaluation of the focus criteria is most important.

If any of the above accommodations were performed, it is likely that a sharper focus peak can be found with the current lens. Such a repetition is justified in two cases:

1. The theoretical depth of field of the current

lens is less than dx[req.] at the current range estimates,

2. dx is not yet well enough defined to allow a change of lens.

A single re-trial with the current lens will almost always exhaust all possibilities for further improvement. Thereafter, if dx'[est] is still not small enough to switch lenses, the ranging program terminates. This condition will usually be caused by low contrast or lack of high frequency texture within the specified field of view. Otherwise, a longer lens is selected, if one remains, and dx'[est] is used to re-center the original region. (Currently, these steps are manually implemented, because the camera-centering software is not yet ready.)

VI,6.5 PERFORMANCE OF FOCUS RANGING PROGRAM

The focus program has successfully achieved focus peaks equal in width to the theoretical, minimal depths of field, expressed in Chart 6.1. A maximum uncertainty of .15" has been obtained with the 3" lens at a range of 25" for a variety of high contrast edges and textured surfaces. There was not time in the present research program to complete a more formal performance evaluation. The effect on range accuracy of each of the accommodations, discussed in this chapter, were observed in practice and found to conform, at least qualitatively, with theoretical

expectations.

A very realistic evaluation of the ultimate practicality of focus ranging will be available shortly; this program will soon be incorporated into the emerging strategy of the hand-eye system. At that time, its performance can be systematically tabulated for a wide variety of feature characteristics.

VI.7 COMPARISON OF ACCURACY POTENTIAL OF FOCUS RANGING IN MAN AND MACHINE

It is known that the shape of the lens in the human eye is also accommodated to achieve sharp focus at a current range of interest. Yet it appears that when people estimate depth, this curvature, if considered at all, is a relatively minor factor. A human cannot, in any case, consistently establish range to better than a few inches (at one meter) without mechanical aids. An interesting perspective on the results of this chapter can be obtained by comparing the relative values in man and machine of the parameters we have found to limit ranging accuracy. This comparison will illustrate why the machine is at an advantage in determining range from focus information.

1. Amplitude Resolution: At finest quantization, the machine is able to distinguish as many as three levels of intensity over a surface judged homogeneous

by human observers. (This result is based on the response of several observers who were asked whether they thought the brightness of a sloping surface, shown on the television monitor, was homogeneous. Their affirmative answers were no doubt biased, to some extent, by their perception of a uniform entity.)

2. Aperture Size: The aperture of the human eye measures approximately 2mm, in normal room lighting (SCIENCE OF COLOR [1958]). The widest aperture of the 3" lens is over 2.7", while the human aperture can be forced open to about 4mm, by darker light levels, low illumination compromises object contrast. One of the main advantages the computer has in this respect is the ability to coordinate its accommodations to nullify extraneous conditions. Thus, if a wide aperture is needed in bright sunlight to get accurate depth, the aperture can be opened, and the camera sensitivity reduced to avoid saturation. The human eye, despite its great accommodation range, cannot force its iris to open in bright light or avoid spatial averaging in dim light.

3. Focal Length: The effective focal length of the eye is about 17mm,, contrasted with 3" for our longest lens.

4. Spatial Resolution: The human eye is unsurpassed in this respect. The cones in the high acuity regions of the retina are spaced on 2 micron centers. This compares favorably with the 37 micron spacing of sample

points in the video raster. Unfortunately, the human's advantage in sampling density is not effectively utilized in focus ranging because of the mentioned limitations in just noticeable intensity discriminations.

According to Equation 2.55c, the eye's depth of field (for high resolution tasks) is about 3" at a range of one meter. (This may be observed by simultaneously focusing on the finger prints of one finger on each hand, held at arm's length. Slowly move one finger closer until simultaneous focus is lost.) Fortunately, the human does not need better range estimates because of his facility for real time servoing. (Sobel [1970] presents models which allow the stereo and statiametric ranging capabilities of the eye and machine to be compared using the data presented above.)

CHAPTER VII: MACHINE COLOR PERCEPTION

VII.1 INTRODUCTION

Color is an extremely valuable descriptive property. Deprived of this sense, the machine operates at a considerable handicap. Recall, for example, an earlier remark that objects, easily distinguishable to the human eye by their hue, may have the same grey-scale value. Conversely, homogeneous regions may appear disconnected to the computer, because of intensity variations caused by illumination gradients or shadows. These are usually correctly perceived as single entities by the human eye, due largely to continuity of color. Finally, color is important as an adjective that enlarges the class of tasks that can be described to the computer (eg."Pick up the RED block" or "Line up the cubes so that a DIFFERENT COLOR appears on each top face,").

In Chapter 2, it was noted that the spectrum of the light incident from an object onto the camera lens depends on the product of the source spectrum and the spectral reflectance of the object. However, due to an innate sense of color constancy, humans tend to see an object with the same hue (corresponding to white light) under a wide variety of illuminations. The initial intent of this work was to develop a general model of this phenomenon so that the

machine could emulate the color perceptions of the human experimenter with whom it had to communicate.

This model was successfully formulated in theory. During the course of implementation, however, the unsuitability of the vidicon for photometric color measurements (detailed in Chapter 2) was discovered. This deficiency made it unfeasible to implement the complete general theory with our present hardware. (Many television cameras without this shortcoming do, however, exist, for instance, image dissectors.) Our emphasis then shifted to developing a limited system that would function acceptably for a single standard source of illumination, such as that normally present in the computer room.

The final implementation, though of limited applicability, nicely illustrates the generality of our basic accommodation paradigm. To overcome inconsistancies within the vidicon, accommodation was optimized by maximizing color separations in a complex decision space.

VII.2 ANALYTIC MODEL FOR ACQUISITION OF COLOR INFORMATION

The illumination incident on the lens of the camera is described by

$$f(\lambda) = S(\lambda)O(\lambda) \qquad (7.1)$$

We are interested in determining $O(\lambda)$, the object's intrinsic reflectance characteristic. $O(\lambda)$ represents the perceived color in white light ($S(\lambda) \equiv$ constant). From $O(\lambda)$ one can then obtain any of the more convenient composite characteristics (eg, hue, saturation) that are normally used to describe an object's appearance.

Color information is acquired by viewing a specimen sequentially through N color filters. Each filter allows the computer to determine the total energy contained in a particular spectral range of the incident illumination. The N filters characterize the spectrum in terms of an N component vector, each of whose terms is, from Equation 2.60, given by

$$I_i = \int_{\lambda} S(\lambda)O(\lambda)F_K(\lambda)W(\lambda)\,d\lambda \quad . \quad K = 1 \ldots N \qquad (7.2)$$

The accuracy with which $O(\lambda)$ can be recovered from these N intensities is a function of N, the number of available filters, and of the extent to which the illumination source, $S(\lambda)$, is known.

To avoid unnecessary complication, the analysis will proceed in two stages. First, the combined spectrum of source and object, $f(\lambda)$, is obtained. This spectrum will then be normalized at each wavelength by a known source spectrum, $S(\lambda)$, to isolate the desired reflectance characteristics, $O(\lambda)$. Let Gamma$_1(\lambda)$ represent the

composite filter functions formed by combining the spectral transmission of each color selective filter with the spectral sensitivity W(λ) of the vidicon.

$$\Gamma_K(\lambda) = W(\lambda) F_K(\lambda) \qquad (7,3)$$

For the initial goal of characterizing the overall spectrum, Equation 7,2 can be rewritten in the simplified form

$$I_K = \int_\lambda f(\lambda)\Gamma_K(\lambda)\, d\lambda \qquad (7,4)$$

VII,2,1 ANALYTIC DETERMINATION OF f(λ)

Equation 7,4 has reduced the problem of determining f(λ) from an N component intensity vector to the solution of N simultaneous integral equations. Since the computer is confined to numerical methods we must, in practice, solve for a discrete approximation to f(λ), using Equation 7,5

$$I_K = \sum_{\lambda_i} f(\lambda_i)\, \Gamma_K(\lambda_i) \qquad (7,5)$$

(λ,i will typically be taken at intervals of 20 milli-microns, corresponding to the interval at which the filter characteristics have been tabulated by the

manufacturer in Figure 2.17.)

VII.2.1.1 SPECIALIZED SOLUTIONS

When the functional form of $f(\lambda)$ can be assumed from a priori knowledge, it is often possible to obtain specialized solutions to Equations 7.4 and 7.5. For example, if the spectrum is expected to be very broad, it might be reasonable to approximate $f(\lambda)$ by a piecewise-constant function. With n filters, a solution can be obtained for a spectral approximation with n degrees of freedom. Equation 7.6 illustrates the solution method to obtain a piecewise-constant approximation of the spectrum over the passbands of three filters. The filters sample the red, green, and blue energy at a total of 3N wavelengths over the spectrum.

(7.6)

$$\begin{bmatrix} \text{Red} \\ \text{Green} \\ \text{Blue} \end{bmatrix} = \begin{pmatrix} \Gamma_1(1) \ldots \Gamma_1(N). 0 \ldots 0 \ldots 0 \ldots 0 \ldots 0 \\ 0 \ldots 0. \Gamma_2(N-2) \ldots \Gamma_2(2N+2). 0 \ldots 0 \\ 0 \ldots 0 \ldots 0 \ldots, \Gamma_3(2N-2), \ldots \Gamma_3(3N) \end{pmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_N \\ c_{N+1} \\ \vdots \\ c_{2N} \\ c_{2N+1} \\ \vdots \\ c_{3N} \end{bmatrix}$$

The components of the left hand column vector (in 7.6) are the unnormalized intensities viewed through the red, green, and blue filters. $c+1,..,c+3N$ represent the values of $f(\lambda)$ at each spectrum frequency. (The overlap of the red filter (Gamma+1) and the blue filter (Gamma+3) by the green filter (Gamma+2) is typical of the actual overlap for the filters described in Figure 2.17.)

To solve Equations 7.6 we introduce the constraint that this spectrum be approximated by constant values over each third of its range.

$$\begin{aligned} c_1 &= c_2 \;\; . \; . = c_N = C_R \\ c_{N+1} &= c_{N+2} \; . \; . = c_{2N} = C_G \\ c_{2N+1} &= c_{2N+2} \; . \; . = c_{3N} = C_B \end{aligned} \qquad (7,7)$$

Using Equation 7.7, Equation 7.6 can be transformed into three linear equations in three unknowns.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{pmatrix} \sum_{i=1}^{N} \Gamma_1(i) & 0 & 0 \\ \sum_{i=N-2}^{N} \Gamma_2(i) & \sum_{i=N+1}^{2N} \Gamma_2(i) & \sum_{i=2N+1}^{2N+2} \Gamma_2(i) \\ 0 & \sum_{i=2N-2}^{2N} \Gamma_3(i) & \sum_{i=2N+1}^{3N} \Gamma_3(i) \end{pmatrix} \begin{bmatrix} C_R \\ C_G \\ C_B \end{bmatrix} \qquad (7,8)$$

Equations 7,8 can then be solved directly to yield C↓R, C↓G, C↓B.

Wolfe [1959] solved Equation 7.4 for the special case when Gamma↓K(λ) and f(λ) were both Gaussian functions. He obtained the mean and standard deviation of the best Gaussian approximation to f(λ) for a given set of Gamma↓K(λ).

VII.2.1.2 SOLUTION WITHOUT AN ASSUMED FUNCTIONAL FORM

The more general problem is: given a set of measured intensities, (I↓K), solve Equations 7.4 or 7.5 for f(λ) in the absence of any a priori expectations about its functional form. An optimal solution, in the "least squares" sense can be obtained by expressing f(λ) as a linear combination of orthonormal functions, Gamma'↓K(λ), derived from the composite filter functions Gamma↓K(λ).

$$f(\lambda_i) = \sum_k c_K \Gamma'_K(\lambda_i) \qquad (7,9)$$

The goodness of this approximation is measured by the error norm

$$\Delta = \| f(\lambda_i) - \sum_k c_K \Gamma'_K(\lambda_i) \| = \sum_i \left[f(\lambda_i) - \sum_k c_K \Gamma'_K(\lambda_i) \right]^2 \qquad (7,10)$$

It can be shown (Hildebrand [1963]) that when the Gamma$_K(\lambda_i)$ are orthonormal functions over the visible range of λ, Delta will be minimized by choosing

$$c_K = \langle f(\lambda_i)\Gamma'_K(\lambda_i) \rangle = \sum_i f(\lambda_i)\Gamma'_K(\lambda_i) \qquad (7.11)$$

The three filter functions Gamma$_K$ (λ) used in most of our work have minimal spectral overlaps (Figure 2,17) and are thus approximately orthogonal. A comparison of Equations 7,5 and 7,11 shows that the c_K can then be obtained by simply scaling the corresponding I_K,

$$c_K = \langle \Gamma'_K(\lambda_i)\, f(\lambda_i) \rangle = \frac{\langle \Gamma_K(\lambda_i) f(\lambda_i) \rangle}{\| \Gamma_K(\lambda_i) \|^{1/2}} = \frac{I_K}{\| \Gamma_K(\lambda_i) \|^{1/2}} \qquad (7.12)$$

(The scaling effectively normalizes the orthogonal Gamma$_K$ as required by 7.11. Normalization compensates for variations among the individual filters with regard to absolute attenuation and width of passbands.)

VII.2.1.3 OVERLAPPING FILTER FUNCTIONS

The assumption of orthogonality weakens as more filters with overlapping passbands are added to enhance spectral resolution. However, the Gramm-Schmitt procedure

can be used to orthonormalize the enlarged set of functions (Hildebrand [1963]). This procedure is conveniently summarized by the recurrence relation,

$$\Gamma'_K = \frac{\Gamma_K - \sum_{s=1}^{k-1} \langle \Gamma'_s \Gamma_K \rangle \Gamma'_s}{\left\| \Gamma_K - \sum_{s=1}^{k-1} \langle \Gamma'_s \Gamma_K \rangle \Gamma'_s \right\|^{1/2}} \qquad (7.13)$$

The Gamma↓K'(λ) given by Equation 7.13 can now be used to expand f(λ) in a series analogous to that given by Equation 7.9. The c↓K in this case are not simple multiples of the corresponding I↓K. To obtain the c↓K we use Equation 7.13 to express Gamma↓K' in Equation 7.11. Performing this substitution and simplifying yields an iterative formula for determining the c↓K from the set of measured I↓K.

$$c_K = \left\langle f(\lambda) \frac{\left[\Gamma_K - \sum_{s=1}^{k-1} \langle \Gamma'_s \Gamma_K \rangle \Gamma'_s \right]}{DN} \right\rangle \qquad (7.14a)$$

$$c_K = \frac{\langle f(\lambda) \Gamma_K \rangle - \sum_{s=1}^{k-1} \langle \Gamma'_s \Gamma_K \rangle \langle f(\lambda) \Gamma'_s \rangle}{DN} \qquad (7.14b)$$

$$c_K = \frac{I_K - \sum_{s=1}^{k-1} \langle \Gamma'_s \Gamma_K \rangle c_s}{DN} \qquad (7.14c)$$

(where DN is the denominator of Equation 7.13)

An immediate consequence of this generalization is that the intensity observed without any filter can be used as a fourth input to improve the characterization of $f(\lambda)$. In this case, the composite spectral response is simply the spectral characteristic of the camera, $W(\lambda)$ (given by Figure 2.16). $W(\lambda)$ spans the entire spectrum. It is, by definition, non-orthogonal to all other composite filter functions. $W(\lambda)$ provides useful information in the gaps between the responses of the red, green, and blue filters.

VII.2.1.4 WEIGHTED NORM

A refinement to Equation 7.10 is to add a weighting factor, $r(\lambda)$, to emphasize the error at selected spectral frequencies.

$$\Delta = \sum_i r(\lambda_i) \left[f(\lambda_i) - \sum_k c_K \Gamma'_K(\lambda_i) \right]^2 \qquad (7.15)$$

A weighted norm is helpful when it is important to have an especially accurate characterization at particular points in the spectrum. For instance, it would be useful in discriminating among several specific colors when the alternatives are known a priori from an external source of information.

VII.3 INTERPRETING SPECTRUM INFORMATION

In typical tasks we are primarily concerned with three descriptive properties of a region: hue (color sensation, eg,red); saturation (strength of coloration); and brightness (total energy). The detailed spectral composition of $f(\lambda)$ is not needed. For this purpose, it is convenient to consider an alternative interpretation of Equation 7.9; $f(\lambda)$ can be represented by a linear combination of primary colors. The spectra of these primaries are given by the orthogonalized filter functions, $\Gamma_K'(\lambda)$. The c_K represent the relative proportions of each primary included in the mixture (This interpretation is an admission that N filters can only specify N independent variables.).

VII.3.1 RELEVANCE OF CLASSICAL COLORIMETRY

The representation of arbitrary colors by mixtures of standard primaries is a cornerstone of classical

colorimetry. By comparing the mixture coefficients of the specimen with the known coefficients of the pure spectral colors, a particularly convenient characterization of hue and saturation can be obtained.

Most colorimetric representations are based on the use of three primaries, since the human eye characterizes color with just three degrees of freedom. The most common primaries have energy peaks in the red, green, and blue portions of the spectrum. (These match the broadest range of visible colors.) Because most of our work involves color filters with these hues, we will review the relevant results of colorimetry in terms of these primaries.

It must be emphasized, however, that the representation to be discussed is applicable to any set of primary colors which satisfy two conditions:

1. No one primary can be matched by a combination of the other two.

2. Some combination of the three primaries will give white light.

There exist an infinity of suitable primary systems all interconvertible by linear transformation. In the absence of physical reasons for the choice of a standard primary system, criteria of convenience may be used. Thus, any reasonable set of composite filter functions Gamma$K' could be used. Furthermore, the methodology can be mathematically extended to accommodate an arbitrary number

of orthogonalized primaries. This generalization will be developed later.

Let R be the value of c•K corresponding to the intensity observed with the red filter for a given stimulus, $f(\lambda)$. Similarly, let B,G represent the coefficients derived for the blue and green filters. Physically, these coefficients can be interpreted in terms of a Gedanken color match experiment. They represent the relative proportions of white light that must be projected through each of the composite filter functions to synthesize the unknown specimen.

If these three primaries are superimposed on a white screen, in the proportion established by the c•K, the chromatic content of the resulting mixture will be indistinguishable from the original stimulus by a perceptual system using the same composite filters. In this sense, the triplet (R,G,B) constitutes a specification of $f(\lambda)$. R, G, B are called the "tristimulus values" of $f(\lambda)$. They are based on the primary system

$$\left\{\Gamma'_1(\lambda) \ , \ \Gamma'_2(\lambda) \ , \ \Gamma'_3(\lambda)\right\} \tag{7.16}$$

To concentrate on the chromatic nature of a stimulus, its hue and saturation, it is desirable to eliminate the irrelevant dimension of brightness. The behavior of color mixtures are governed by linear relations

known as Grassman's Laws (discussed in Sheppard [1968]). Two useful consequences of these relations are:

1. Luminosities at different wavelengths can be added algebraically to obtain the total brightness.

2. All tristimulus values may be multiplied by the same positive factor without altering the chromatic content of the specification.

Total brightness is thus expressed by the sum of the tristimulus coefficients. The luminance information can be removed by dividing each of the tristimulus terms by their sum.

$$r = \frac{R}{R + G + B} \tag{7.17a}$$

$$g = \frac{G}{R + G + B} \tag{7.17b}$$

$$b = \frac{B}{R + G + B} \tag{7.17c}$$

r, g, and b are known as chromaticity (or trilinear) coefficients. They are not independent; they must sum to one. Consequently, the relative proportions of the primaries (which determine the color properties of a specimen) can be expressed by two independent variables and represented by a single point in a plane.

VII,3,1,1 CHROMATICITY DIAGRAM

This two-dimensional color space is commonly known as a chromaticity diagram. In Figure 7.1 the axes correspond to the relative contributions of the intensities, seen through the red and green filters, to the total luminance. The contribution of the blue primary is represented implicitly through the relation r+b+g=1. A point in this plane reflects the relative proportion of the total energy emitted by a stimulus that lies in the passband of each (orthogonal) filter function. For example, the point labeled W signifies an equal contribution from each filter. This point is conventionally known as the "white point".

VII,3,1,2 INTERPRETING HUE AND SATURATION FROM CHROMATICITY COORDINATES

The chromaticity coordinates will be used to relate the analytically determined c+K to the descriptive properties of hue and saturation.

Hue and saturation are, strictly speaking, the subjective impressions elicited by a specimen in a particular observer, under specific viewing and illumination conditions. The machine must, of course, base all its decisions on hard physical evidence. Consequently, the

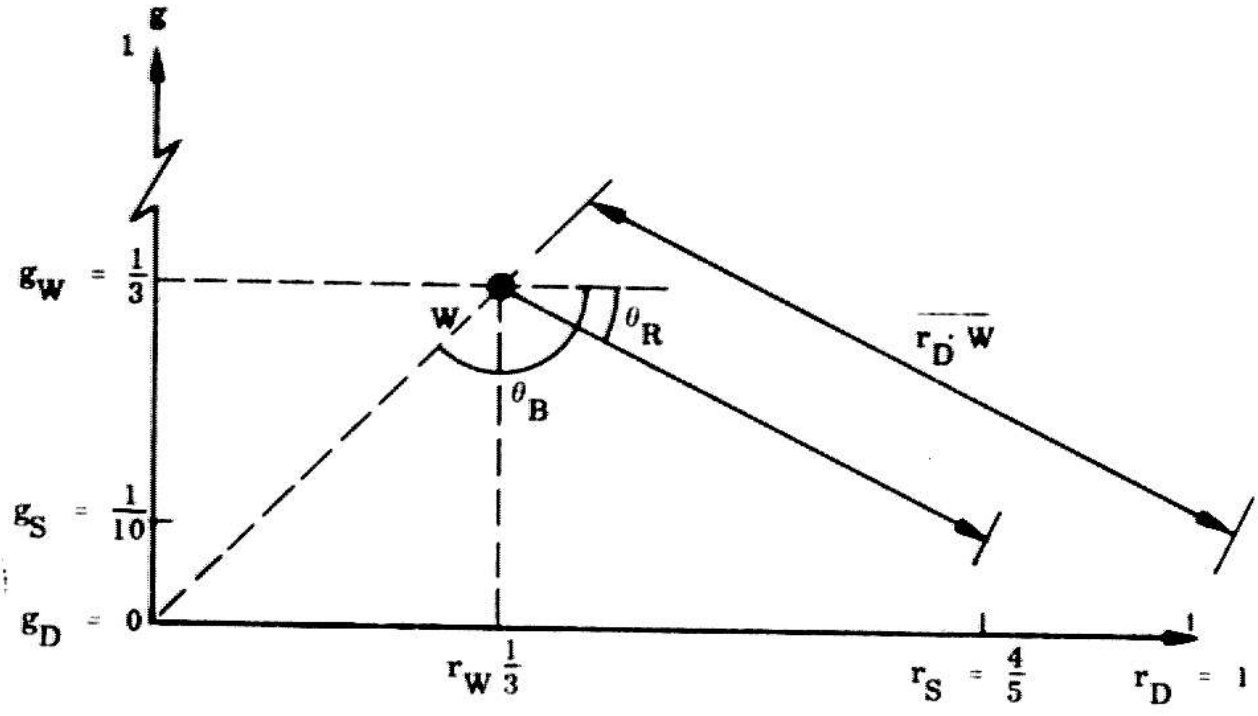


Fig. 7.1 Chromaticity Diagram

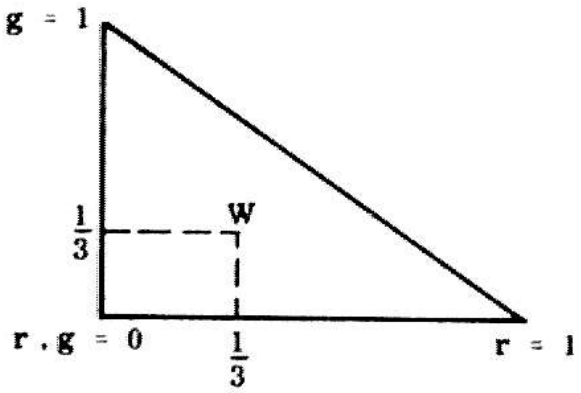


Fig. 7.2 Degenerate Representation of Pure Spectral Colors

chromaticity coefficients will actually be used to determine dominant wavelength and purity. Under most circumstances, these are close physical correlates of the psychophysical variables.

Let us begin by interpreting the significance of the white point. As its name implies, this point corresponds to no hue or zero saturation: physically, the energy is presumed to be equally distributed over the entire spectrum. A human would identify such a color to be either black, grey, or white. The choice depends on the luminance level of the local region relative to the average luminance over the entire image.

The computer must make a similar comparison. Since no luminance information is retained in the chromaticity diagram, it is necessary to preserve the unnormalized total intensity (the sum of red, green, and blue) seen at that point. An attractive alternative is to observe the scene intensity without any color filter. The voltage reference provided by the autotarget circuit of the camera is a measure of the average scene brightness. The intensity, observed at the particular achromatic location, can then be compared with this global average to establish a relative tonal value.

Point "A"(see Figure 7.1), on the other hand, represents a distinct chromatic bias: 80% of the total luminous energy observed for this stimulus was obtained

through the red filter. A human observer would most likely classify the hue of the corresponding stimulus as red.

VII,3,2,1 DOMINANT WAVELENGTH

Suppose that the original stimulus were now diluted by mixing it with white light. We know from experience that this addition will lighten the red, making it appear less saturated. It will not, however, alter the underlying hue. Physically, the white light increases the energy level at each spectral frequency by an equal increment. In terms of the chromaticity coefficients, the effect of incrementing each tristimulus value is to reduce the relative importance of the energy peak in the red end of the spectrum. This is shown in the chromaticity diagram by a shift in the mixture coordinates from "A" towards "W" along the line connecting them.

We conclude that all points extending from "W" along the line defined by angle theta$_R$ correspond to the hue, "red". (Similarly, all points representing blue specimens at various levels of saturation would line up along another locus defined by angle theta$_B$.)

More generally, the angle theta specifies what we have called the "dominant wavelength" of a stimulus. This phrase can be formally defined as that spectral color which, when combined with white (uniform spectrum), produces a

mixture equivalent in appearance to the actual stimulus. (An exception to this definition is that no pure color, corresponding to red-blue mixtures, exists in nature.)

VII.3.2.2 PURITY

The position on W-A at which the red-white mixture will plot depends on the percentage of white that was added. Purity is the term that expresses the relative proportions of white and the corresponding dominant spectrum color that are needed to match a particular stimulus. This proportion is represented geometrically in the chromaticity diagram by the ratio of the distance of a stimulus from the white point relative to the distance for the corresponding pure color. This ratio is expressed most easily with vectors:

$$P = \frac{|\bar{S} - \bar{W}|}{|\bar{D} - \bar{W}|} \quad , \quad \begin{array}{l} \bar{S} = \text{coordinates of stimulus} \\ \bar{W} = \text{coordinates of white point} \\ \bar{D} = \text{coordinates of Dominant wavelength} \end{array} \qquad (7.18)$$

Purity varies from 0 to 1 as S-bar moves between W-bar and D-bar.

In two dimensions, we can establish a computational equivalence between the ratio expressed by Equation 7.18 and the corresponding ratios obtained using only the r or g coordinates of the vectors

$$P = \frac{|r_S - r_W|}{|r_S - r_W|} = \frac{|g_S - g_W|}{|g_D - g_W|} \qquad (7.19)$$

Pure red is represented by chromaticity coordinates of (1,0). Thus, for instance, the purity of point "A" is given by

$$P = \frac{(0.8 - 0.33)}{(1.0 - 0.33)} = 0.7 \qquad (7.20)$$

VII,3,1,2,3 DEGENERACY OF CHROMATICITY COEFFICIENTS OF PURE COLORS

In general, pure colors cannot be adequately represented by the intensities obtained from a finite number of non-overlapping filters. Any pure spectral frequency must lie exclusively in the passband of a single filter. Consequently, all completely saturated colors will be mapped into one of the vertices of the color triangle illustrated in Figure 7,2. Short wavelengths will be mapped into vertex (r, g=0) (ie. energy passed by blue filter), long wavelengths into vertex (r=1), and middle wavelengths into (g=1).

As a result of this degeneracy, Equation 7.18 must be modified. The denominator should express the distance

from the white point to the vertex (in Figure 7.2) that contains the actual saturated hue. Note that the degree of degeneracy will decrease if more than three filters are used, because the spectrum will be more finely partitioned.

The inability to represent pure colors is not an important limitation in practice, except for the compromise introduced into the definition of purity. Real world objects of the type encountered in the hand-eye environment typically have very broad spectral reflectance characteristics. When these objects are viewed with broad band illuminations (such as room lighting), some energy will always be registered through each of the filters (see Figure 7.3).

VII.3.1.2.4 RELATION OF DOMINANT FREQUENCY AND PURITY TO f(λ)

A rough correspondence can be established between dominant wavelength and purity and the detailed composition of the original waveform, f(λ). The dominant wavelength will usually coincide with the wavelength λ•max at which f(λ) is maximum. This coincidence is especially likely for well-behaved (smooth, continuous, monotonic) spectra typical of simple real world objects.

Purity refers to the degree of predominance of the strongest wavelengths, relative to the average energy levels

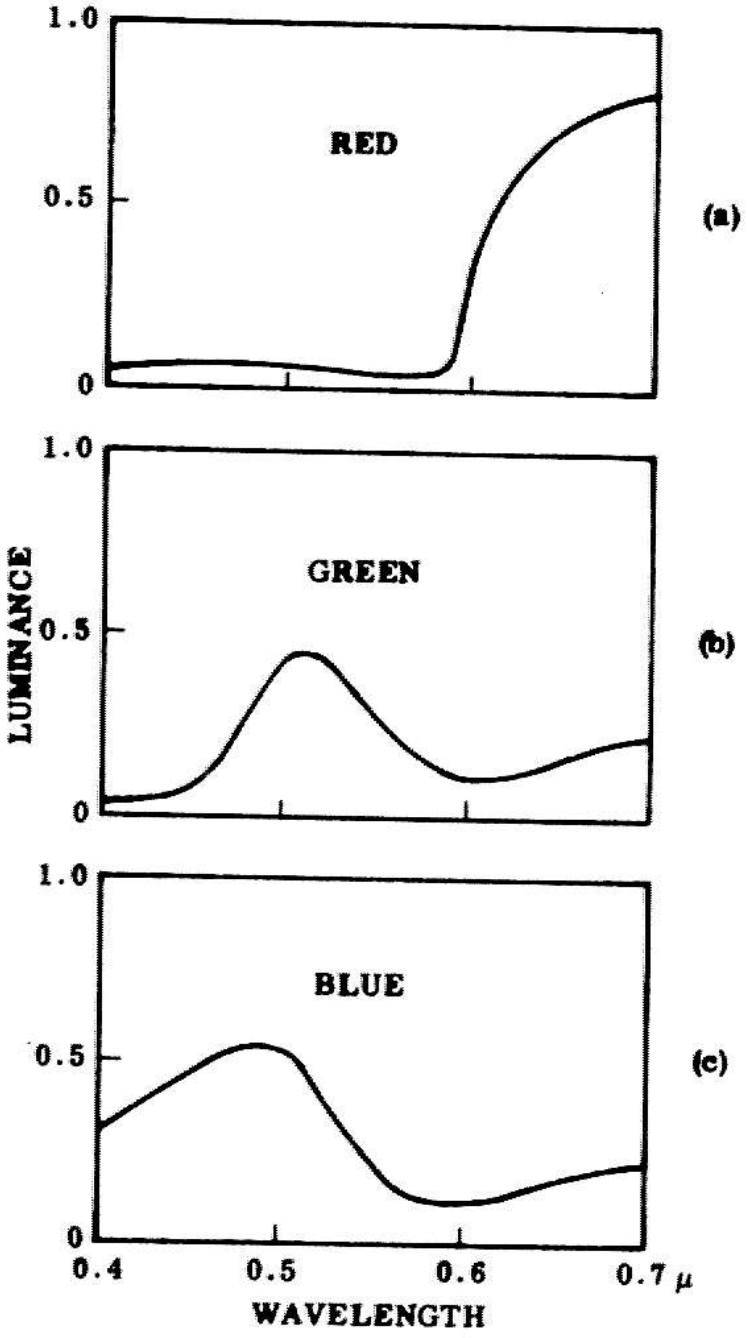


Fig. 7.3 Spectral Reflection Curves for Typical Red, Green, and Blue Pigments (Adapted From Wright [1958])

found over the spectrum (see Figure 7.4). A statement of this predominance is expressed by

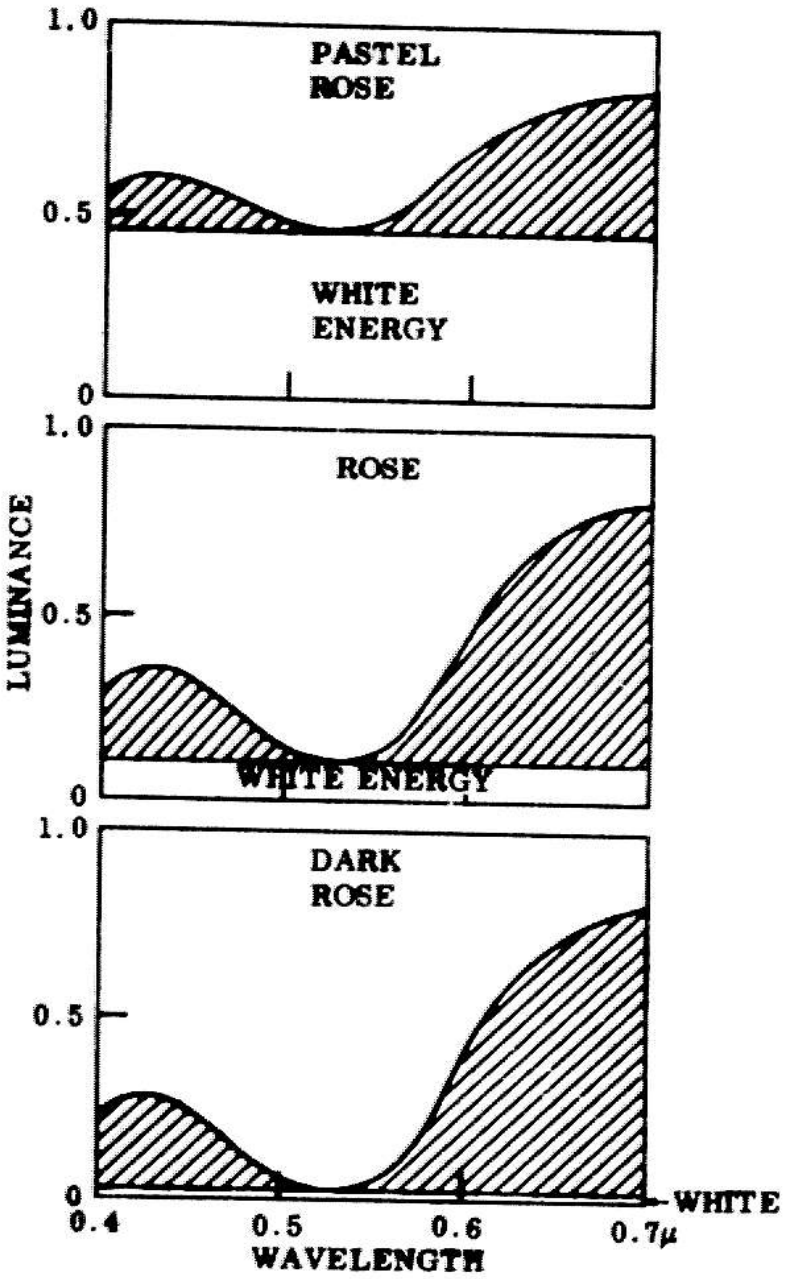
$$P' = \frac{\text{Total energy above white level}}{\text{Total energy}} \qquad (7,21)$$

P' bears a qualitative association with purity defined by Equation 7,18, P', like P, varies from 0 to 1 as the dominant peak becomes more sharply defined.

VII,3,2 BEYOND TRIVARIANCE

The visual system of a normal observer is trivariant, With the exception of brightness, two independent dimensions specify chromatic properties, Dominant wavelength and purity thus constitute a necessary and sufficient set of descriptors for a human observer.

The color discrimination of the human vision system is fundamentally constrained by its trivariant nature. The transformation between a complex spectrum and the two color descriptors can be modeled with three filters. There appears then to be no necessity for a machine to employ more than the three color filters required to achieve a comparable level of discrimination, On the other hand, there is nothing that limits the number of filters a machine can employ. This freedom motivates us to at least consider what advantages, if any, accrue from the use of more filters.



EFFECT OF DECREASING SATURATION ON THE SPECTRAL REFLECTION OF A MAGENTA PIGMENT
(Adapted from Wright [1964])

Fig. 7.4 Effect of Decreasing Saturation on the Spectral Reflection of a Magenta Pigment

If sufficient justification exists, we must then generalize the concepts of dominant wavelength and purity to apply to the representations obtained with an arbitrary number of filters.

VII.3.2.1 METAMERS

The main advantage of using more filters arises from the increased ability to resolve metameric matches. Metamers are two specimens that appear to have the same color description, when in fact, their detailed spectral compositions are dissimilar. For example, when a spectrally uniform white light is matched by mixing appropriate proportions of pure red, green, and blue light, the match is said to be metameric (see Figure 7.5).

Metamers result, because there are not enough independent parameters from three filters to register the fine structure in the spectrum; in electrical engineering terminology, the spectrum is undersampled. Metamers are always defined with respect to the color sensitivity of a particular observer. Thus, another person with a higher inherent sensitivity to red might see the discrete spectrum in Figure 7.5 as pink rather than white. The dependency of metamers on the characteristics of a hypothetical mono-variant observer is illustrated in Figure 7.6.

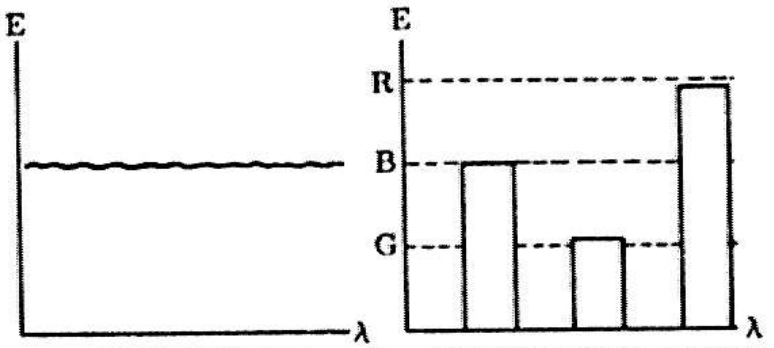
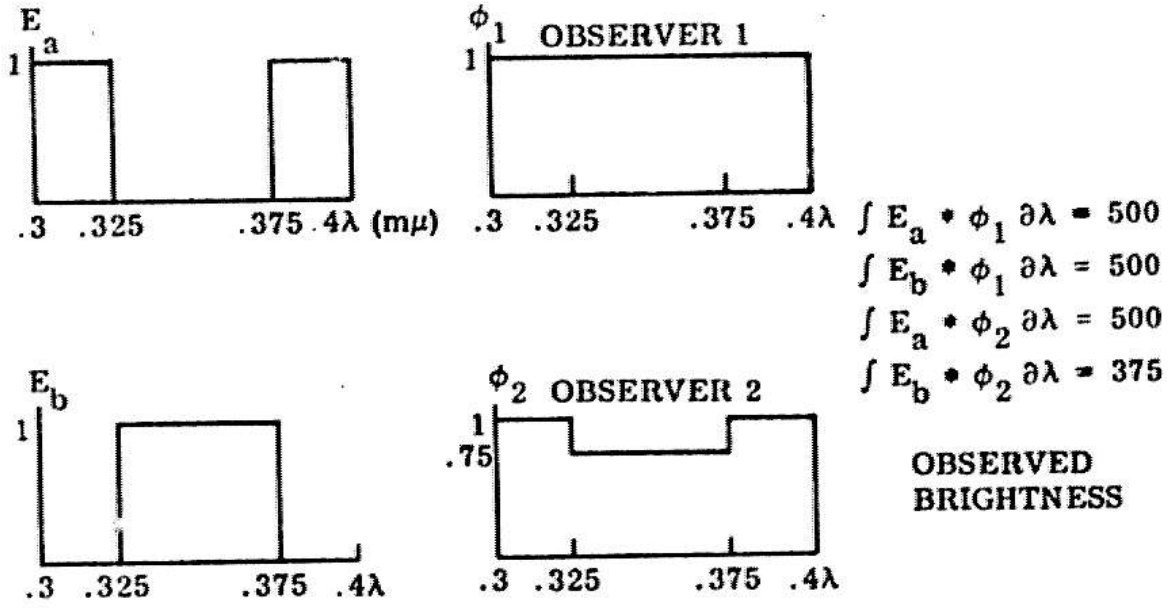


Fig. 7.5 Metameric Match



SPECTRA E_a AND E_b ARE METAMERIC TO OBSERVER 1 BUT DISTINCT TO OBSERVER 2

Fig. 7.6 Dependency of Metameric Match on Characteristics of Observer

VII.3.2.2 THE NEED FOR MORE THAN THREE FILTERS

It is highly improbable that the three filters chosen for use by the machine will exactly match the spectral characteristics of the corresponding "filters" in the head of whoever is trying to describe colors to it. The choice of red, green, and blue filters insures that the correlation will be close enough so that most color distinctions which a particular person can resolve, can also be discriminated by the machine. It is also the case, however, that some spectral pairs will appear metameric to the computer but not to a person. One of the principal reasons for considering the use of more than three filters is to minimize the contingency that the machine will be unable to discriminate color distinctions described to it by any of a large population of people. (Conceptually, one might say that the machine can use combinations of n filters, three at a time, to see whether two colors can be resolved using any of these filter combinations.)

It is suspected (but as yet unsubstantiated) that the higher dimensionality of additional filters is helpful in resolving small overall color differences in the presence of noise. Assume two spectral distributions are similar, except in a small interval of wavelengths. This distinction would clearly be most pronounced in the output of a narrow filter centered on that part of the spectrum. In

pattern recognition terms, the parameters describing these two spectra will be further apart in a higher dimensional measurement space,

Additional filters, however, will also detect disparities in spectra that should be contained in the same equivalence class, by human standards. Thus, scatter is introduced into the cluster of parameters corresponding to each color, Whether or not a net gain in the ratio of inter-cluster/intra-cluster distance is achieved must be determined experimentally for a particular set of objects and observers,

It is, of course, not necessary to restrict a machine to spectral discriminations that correspond to human judgement, Resolving a person's metameric ambiguities might well be a powerful way for the machine to discriminate between different stimuli that a person must distinguish by cues other than color,

VII,3,2,3 DEFINITION OF DOMINANT WAVELENGTH AND PURITY FOR n FILTERS

The concepts of dominant wavelength and purity can be generalized by extending the geometric interpretation of these terms (developed in Figure 7,1) to a symmetric n-1 dimensional space,

In Figure 7,7 the blue chromaticity coordinate is

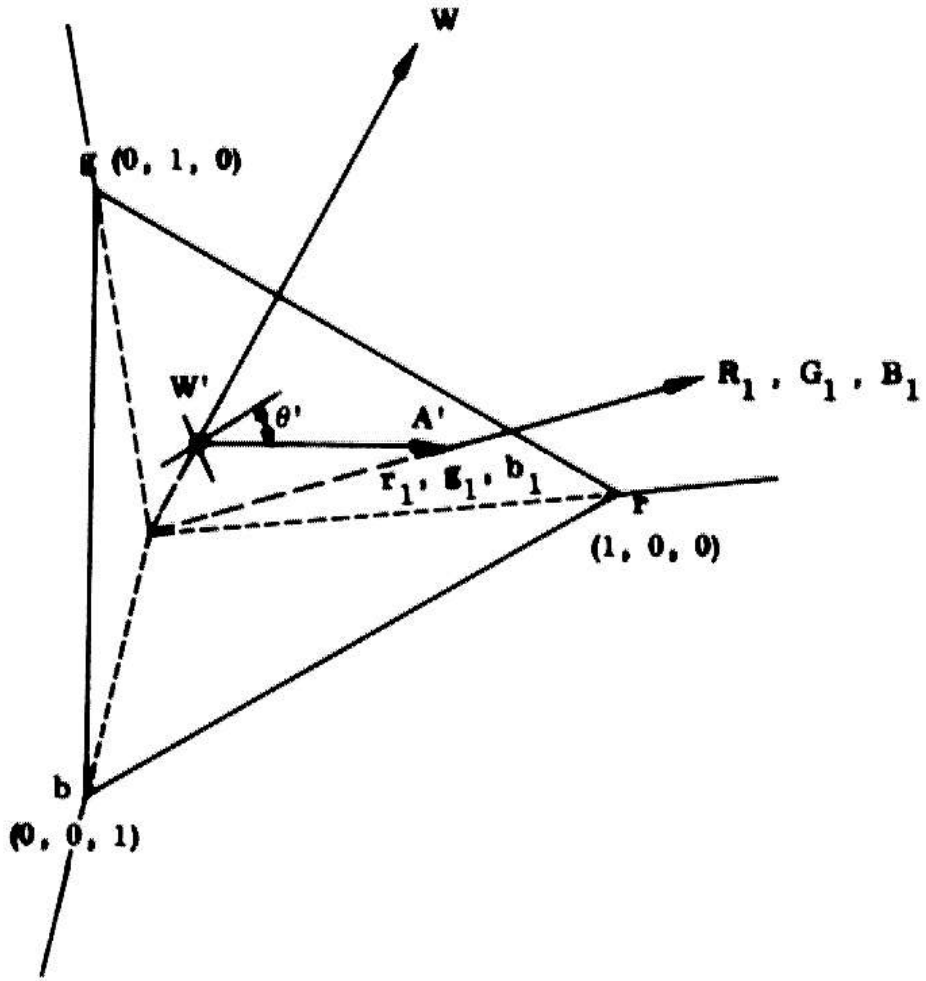


Fig. 7.7 Dominant Wavelength and Purity Defined in Symmetric 2-Dimensional Color Space

explicitly used in order to represent the color characteristics obtained with three filters in a symetric color space. The space is constructed on the surface defined by the plane

$$r + g + b = 1 \qquad (7.22)$$

The point at which the vector, determined by the tristimulus coefficients, pierces this plane defines the normalized trilinear coordinates. (The projection of this point onto plane r-g defines the equivalent point in the chromaticity diagram, Figure 7.1.) The angle theta' defined by the vector from the effective white point W' to a sample point (r,g,b) is used as a measure of hue. Similarly, the distance from W' to a sample point, normalized by the distance to any of the vertices (They will be equidistant with normalized filters.), is a measure of purity. This construction is directly extendable to higher dimensions. Given the intensities from n filters, we intersect this vector with a normalized n-1 dimensional surface drawn in an n-dimensional hyperspace. The direction from the white point to the piercing point is again taken as a measure of hue, while the normalized euclidean distance is a measure of the saturation.

VII,4 COLOR RECOGNITION

We now describe an approach to automatic color recognition based on the chromaticity representations developed in the last section. For simplicity, the discussion will be based on three orthogonal color filters. All aspects, however, can be generalized to accommodate an arbitrary number of such filters.

The basic recognition process involves simply mapping an unknown stimulus into the chromaticity space illustrated in Figure 7.7. The computer then determines the recognizable color whose direction from the white point is closest to that of the specimen. For example, in Figure 7.8 the unknown would be called red. Purity is then measured. If the stimulus is sufficiently saturated, the closest hue is returned. Otherwise, hue is not significant; the sample is assumed to be achromatic. In that case, the relative brightness is returned, as described earlier.

VII,4,1 ESTABLISHING RECOGNIZEABLE COLOR CATEGORIES

The transmission characteristics of our color sensors are not calibrated to any standard observer. It would involve considerable effort to define formal correspondences between color categories and angular position in Figure 7.8. We have opted to avoid standardized,

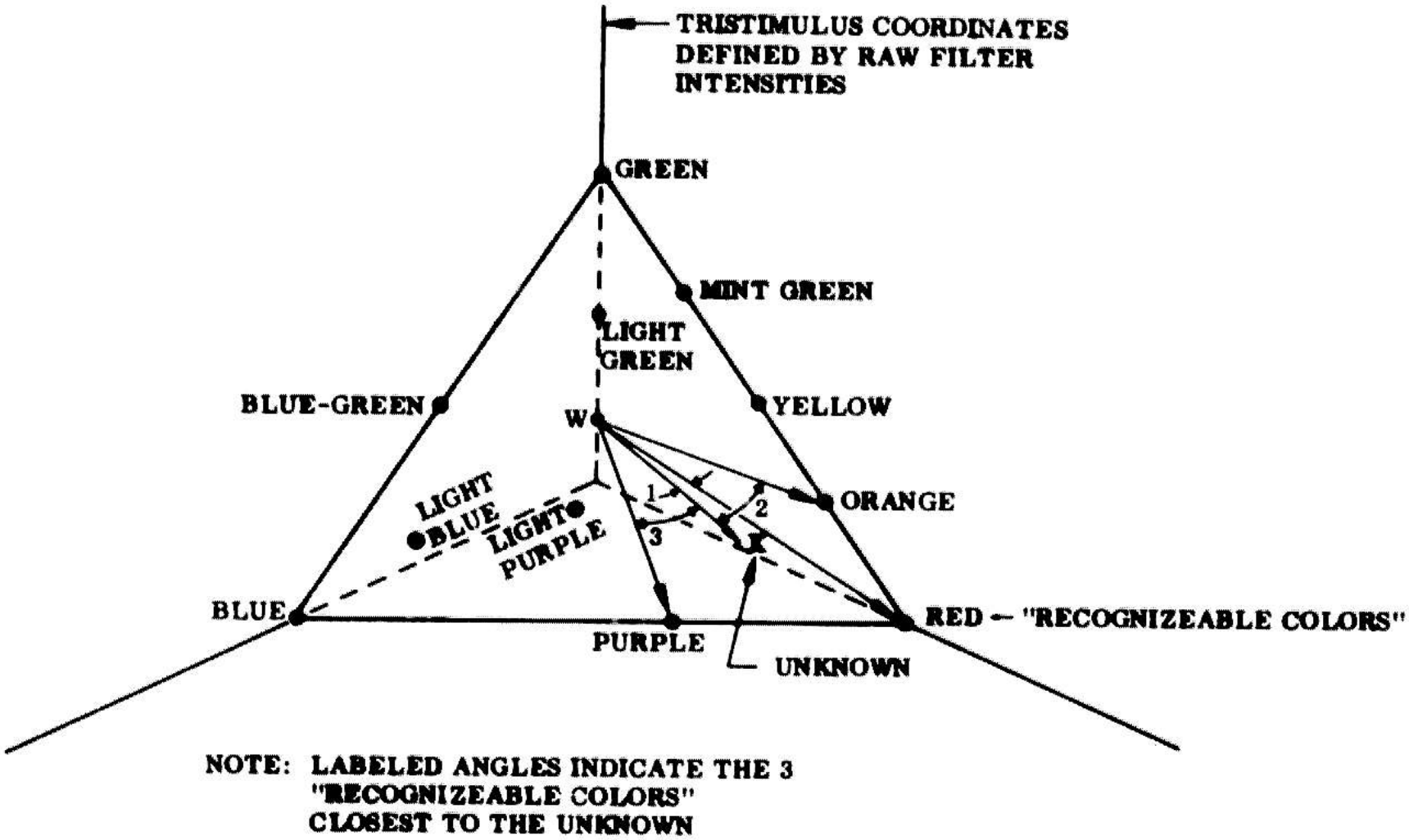


Fig. 7.8 Basic Recognition Paradigm

color notations in favor of empirical definitions. The machine is taught the desired color discriminations by the human operator:

1. The operator shows the machine a specimen belonging to an unknown color category.

2. The machine computes the chromaticity coordinates of this sample and assigns to them whatever color name is entered by the operator.

3. Each color category can be refined by additional samples. The center of gravity of the cluster coordinates, representing all previous samples of a given color, explicitly defines what is meant by the associated color name.

The interactive nature of color specification insures that, in most cases, the machine and the experimenter will agree on what is meant when color is mentioned in a task description.

VII.4.2 A DETAILED DESCRIPTION OF THE RECOGNITION PROCESS

An initial color vocabulary must be established. This can be done by showing the computer samples of several colors and identifying them by name. This procedure need not be repeated at each session, because the cumulative experience of past sessions, represented by cluster center of gravities can be loaded from disk storage.

When the program is called upon to identify a color, it first obtains the specimen's chromaticity coordinates using Equations 7,17. These coordinates are used to define a direction relative to the white point, which is then compared to the direction of known colors. A computationally expedient way to determine the best match is illustrated in Figure 7,9.

1. Transform the vectors from the white point to all cluster centers and to the unknown specimen into unit vectors by dividing each one by its length.

2. Perform a nearest neighbor match on the hypersphere defined by the end points of these normalized vectors.

Purity is determined using the original length from W to the unknown, as shown in Figure 7,10. The symmetry of the space allows the distance corresponding to full saturation (at any hue) to be represented by the distance from the white point to the periphery of a circle drawn through the vertices of triangle r,g,b.

The distance of the unknown to the closest color category, relative to its distance from the second best, reflects the confidence of the match

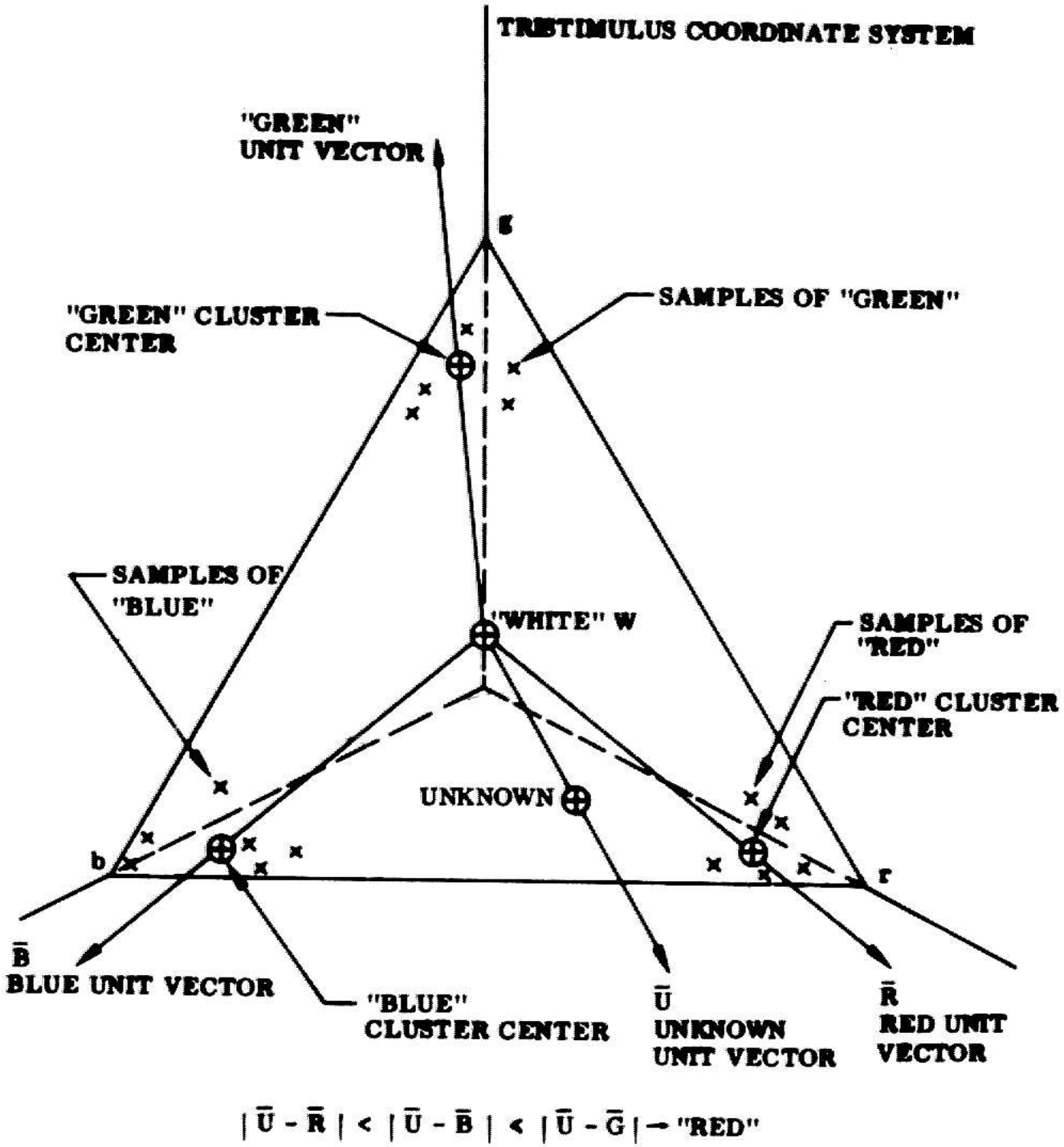


Fig. 7.9 Nearest Neighbor Color Recognition

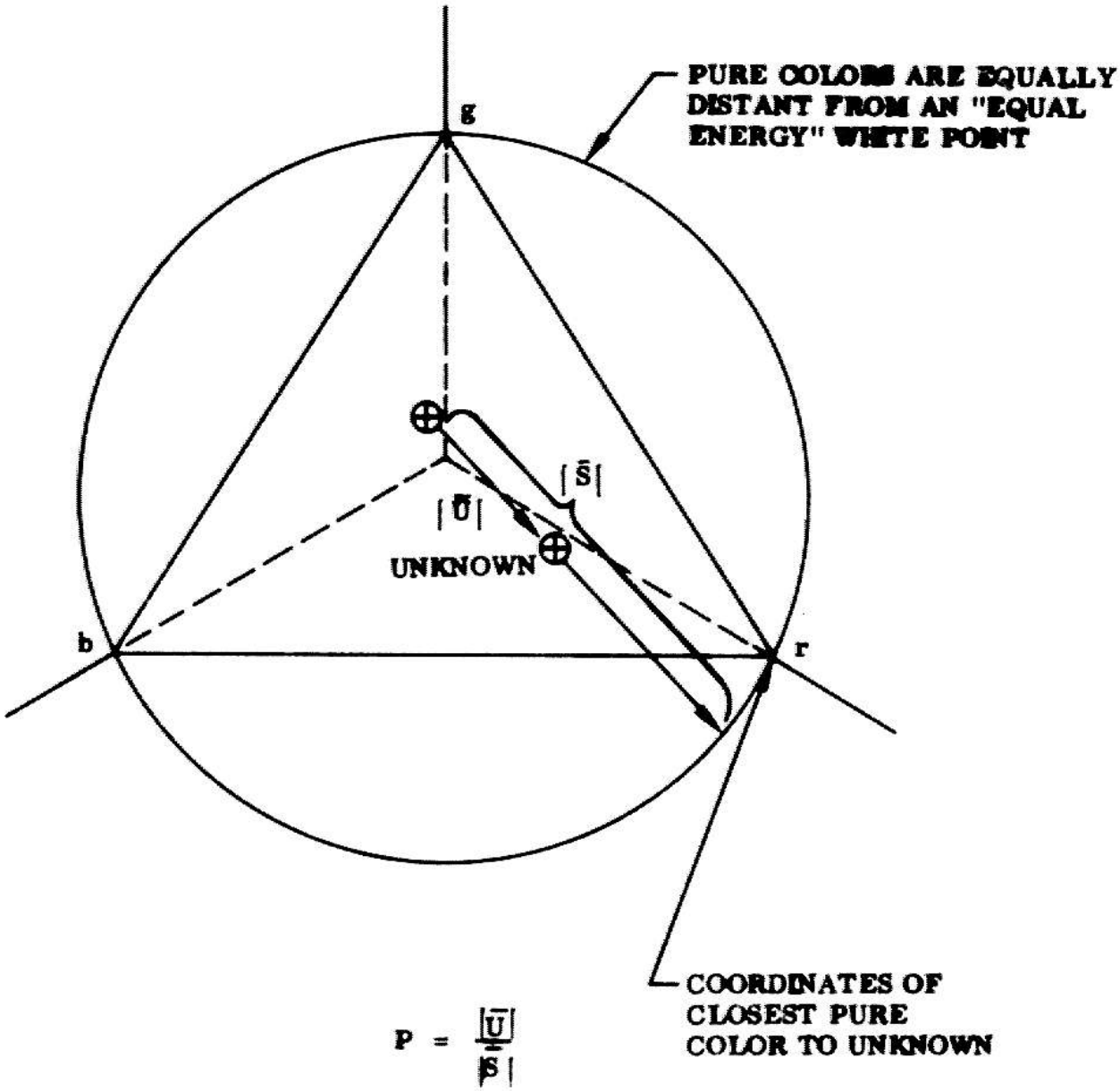


Fig. 7.10 Color Purity of Unknown

$$C = 1 - \frac{|\overline{P}_1 - \overline{U}|}{|\overline{P}_2 - \overline{U}|} \quad (7.23)$$

$\overline{U}$ ≡ coordinates of tip of "unknown" unit vector

$\overline{P}_1$ ≡ coordinates of tip of unit vector from best match category

$\overline{P}_2$ ≡ coordinates of tip of unit vector of second best matching category

Thus, in the example illustrated in Figure 7.9,

$$C = 1 - \frac{|\overline{R} - \overline{U}|}{|\overline{G} - \overline{U}|} \quad (7.24)$$

C=1 for an exact match and decreases to zero when the unknown is exactly halfway between established cluster centers.

When confidence of a color identification falls below a level set by the program requesting color information, the color routine can interact with the operator to improve its performance. It can ask, for example, for the correct identity of the questionable hue. If the actual color corresponds to an existing cluster center, that center of gravity is shifted towards the new sample point, insuring a higher confidence match the next time this variant is encountered.

On the other hand, if the color had not been seen

before, a new cluster is established with the coordinates of this sample as the initial center of gravity. In these ways, the program, much like a child, is able to sharpen its color discrimination ability as it accumulates experience. The following example typifies this behavior. When the program was first implemented, it was taught the gross distinction between red and green objects. Naturally, the first time it was shown an orange object, it confused it with red. After the program was informed of the distinction, it never repeated that particular mistake.

VII,4,2,1 LEARNING DISTINCTIONS WITHIN CLUSTERS

As the program acquires more sophisticated color discriminations, it is also likely that some of the previously acquired clusters will become unwieldy and need to be partitioned. For example, when only a few widely separated colors were known, it might have been acceptable to lump all shades of blue (from greenish to purplish) under that one heading. Because of these variants the cluster corresponding to blue might become widely scattered over a multi-lobed pattern. In a crowded space with many neighboring clusters the direction defined by the center of gravity, denoting all shades of blue, might provide a very poor match confidence for certain of the shades. Recognizing this contingency, the program could request the

experimenter to refine the categorization by providing identifying names for association with the major sub-clusters within blue.

VII.5 RECOGNIZING OBJECT COLORS IN SPECTRALLY BIASED ILLUMINATION

We have specified a method for obtaining

$$f(\lambda_i) = S(\lambda_i)O(\lambda_i) \qquad (7.25)$$

This product is equivalent to object color, $O(\lambda)$, only when the illumination is provided by a spectrally flat white source (eg, $S(\lambda) = 1$). In removing this restriction we must come to grips with the psychological phenomenon known as color constancy.

VII.5.1 COLOR CONSTANCY: AN INTRODUCTION

Chromatic adaptation or color constancy refers to the well-known and remarkable ability of humans to perceive an object in its natural hues in spite of considerable spectral bias that may be present in the illuminant. The implication that perceived hue is not strictly a function of physical wavelength has been a suspicion of students of color for at least several centuries. Edwin Land's

relatively recent experiments with two primary color projections have provided astonishing evidence that perceived colors of all hues can be elicited by combining light from just two relatively narrow spectral bands, where the choice of bands is substantially arbitrary. The most tenable explanation for Land's observations was anticipated by H,von Helmholtz [1924] in his classic work on color adaptation in colorimetry. Helmholtz would probably have disagreed strongly with Land's contention that

"There is a discrepancy between the conclusions one would reach on the basis of the standard theory of color mixing and the results we obtain in studying total images. And this departure from what we expect on the basis of colorimetry is not a small effect but is complete."

Helmholtz concurred that colors perceived to belong to an object (as opposed to an isolated patch of light) may indeed depend on other factors besides the spectrum of the radiant flux. A strong influence will obviously be exerted by the spectral composition of the illumination source. Helmholtz, however, assumed that the nature of this bias could be inferred from the overall appearance of the objects and background comprising a total image. He hypothesized that the influence of the illuminant on the measured spectrum could be systematically discounted to determine the

subjective hue associated with any particular specimen using conventional laws of color mixture. In his own words,

"We form a judgement about the colors of bodies, eliminating the differences in illumination by which a body is revealed to us,...By seeing objects of the same color under these various illuminations we learn to form a correct idea of the color of bodies; that is to judge how such a body would look in white light; and since we are interested only in the color that the body retains permanently, we are not conscious at all of the separate sensations which contribute to form our judgement."

What is special about colors, perceived to be localized within a non self-luminous object, that challenges the classic colorimetric models? The term "object color" implies that the judgement of hue is probably based on information from a number of light patches. This information is processed in the cortex to yield the perception of an object isolated from the background. Judd [1960] conjectures that, simultaneously with this perception of object color, there must be a perception of the color of light by which that object is illuminated. We will shortly suggest specific techniques by which a machine can utilize global knowledge to obtain the required spectral characteristics of the source.

Objections have been raised to the term "discounting the illuminant" on the grounds that it implied a deliberate intellectual judgement on the part of the observer. Most psychologists believe that constancy phenomena are unconscious and involuntary (Helson [1943]). (Effects of color constancy have been observed, for example, in other mammals to whom we are reluctant to ascribe human powers of reasoning.) Helmholtz, however, had intended his hypothesis to be a model of an unconscious process: the measured color coordinates of the object are modified by the inferred coordinates of the illuminant to yield the classical color representation of the perceived hue.

Machine color perception is realized through deterministic processing of spectral data obtained from physical sensors. The basis of our mechanization of this function emerged as a direct consequence of the spectral analysis model of color perception developed earlier. The simplicity, with which we can obtain and normalize out the influence of the source spectrum to obtain an object's inherent spectral reflectance characteristics, lends support to the plausibility of von Helmholtz' approach.

VII.5.2 REQUIREMENTS FOR CONSTANCY IN MACHINE COLOR PERCEPTION

When a man describes an object by its color, he is

inclined by his sense of constancy to mean the hue (as seen in white light) that the body appears to retain permanently. He is usually not aware of the reflected wavelength that happens to be dominant under the current illumination. It is imperative that we provide a comparable constancy for our machine's color sense. This preserves the common perceptual frame that both must share if communication between man and machine is to be possible.

VII.5.2.1 SOURCE ADAPTATION FOR COLOR CLUSTERING

Independent of this need to communicate, source normalization is implicitly required to obtain consistent identifications of object color using the adopted cluster-distance recognition paradigm. The coordinates of an observed specimen must be transformed to compensate for any spectral dissimilarities that exist between the illuminants used for learning and for recognition. Our entire basis of identification would otherwise be obsolete, whenever the lighting differed from that under which the learned clusters were first observed. Furthermore, since color discriminations are intended to be refined over several sessions, it is also necessary that we be able to combine cluster coordinates learned under various illuminations on a common scale.

Both requirements are met by establishing a standard

illumination to which all observed color coordinates are normalized. The natural choice for such a reference is spectrally flat-white light. The coefficients will then correspond to the spectral reflectance of the object and the analysis of the preceding sections can be applied directly.

(One could imagine a brute force approach to constancy as a direct extension of the cluster concept. All clusters would simply be partitioned according to the illumination conditions that pertained at the time they were observed. Subsequent recognitions would then be attempted only in that sub-space whose illumination corresponded to the currently prevailing conditions. Parsimony, however, is important for survival.)

VII.5.3 SOURCE NORMALIZATION OF CHROMATIC COEFFICIENTS

The intensity in the reflected spectrum at each wavelength ($f(\lambda)$) is the product of the incident illumination and the relative reflectance of the object, both taken at the wavelength in question (see Equation 7.25). If the spectral distributions of the reflected and incident light are both available, the desired object reflectance function, $O(\lambda)$, can be recovered by dividing $f(\lambda)$ by the incident illumination on a frequency by frequency basis.

(7.26)

(Equation 7.26 is only defined for continuous, broadband source spectra of the type ordinarily found in room environments. Illuminations with strong spectral bias are not considered. Nelson [1943], however, has performed relevant psychological studies.)

The c_K needed to describe $O(\lambda)$ are obtained by using Equation 7.26 instead of $f(\lambda)$ in Equation 7.12. If $f(\lambda)$ and $S(\lambda)$ are each already expressed as a series of orthonormalized filter functions (see Equation 7.9), then the c_K corresponding to $O(\lambda)$ can be directly found; each c_K in the expansion of $f(\lambda)$ is divided by the corresponding coefficient in the series for $S(\lambda)$.

$$O(\lambda) = \left[\sum_k \frac{(c_K)_{f(\lambda)}}{(c_K)_{S(\lambda)}}\right]\Gamma_K \qquad (7.27)$$

Source Calibration

The source is easily calibrated by using an object for which $O(\lambda)$ is known. The process of determining $S(\lambda)$ given $f(\lambda)$ and $O(\lambda)$ is completely analogous to the procedure for obtaining $O(\lambda)$ given $f(\lambda)$ and $S(\lambda)$. The illumination,

$f(\lambda)$, reflected from the reference object is measured. $S(\lambda)$ is then given by

$$S(\lambda) = I(\lambda)/O(\lambda) \qquad (7.28)$$

VII.5.3.2 CHOOSING A CALIBRATION OBJECT

The calibration object can be any convenient fixture of the environment. In the specific context of hand-eye, the gold base of the arm is appropriate. This function could also be served by any recognized object whose color is intrinsic and known from past experience (eg, All wedge-shaped blocks are blue.) The only constraint is that the object's reflectance be spectrally broad to insure that the source spectrum will be observed at all wavelengths.

Without loss of generality, a flat-white reference surface (eg, a piece of matte paper attached to the base of the arm) can be used to insure this last condition. The reflectance function of this surface is a constant, typically 80%, at all wavelengths. The spectral distribution of light observed on a white surface is thus a direct "reflection" of the characteristics of the illumination source.

VII,5,3,3 CALIBRATION IN THE ABSENCE OF KNOWN REFERENCES

Psychologists have found that it is often possible to assess the nature of a source by considering illumination highlights reflected from glossy objects. In the absence of highlights, one can assume a random distribution of spectrally selective objects and take the average chromaticity compiled over the whole scene as a characterization of the source. (This, of course, assumes uniform illumination over the scene.) These techniques, however, are more applicable to a robot exploring outer space than to one that lives in the constrained environment of a computer room.

A deterioration in the recognition confidence is a cause to suspect alterations in the illuminant. This hypothesis is easily confirmed by recalibrating the source. Any object whose color had previously been determined with high confidence can be used as a secondary standard.

VII,6 ACCOMMODATION IN COLOR PERCEPTION: THEORY

The accommodation considerations involved in obtaining accurate chromaticity coordinates are theoretically those involved in obtaining accurate photometric intensity measurements. To optimize signal/noise the camera sensitivity must be peaked for each filter so

that the signal level corresponding to the desired intensity is just shy of saturating the video amplifier. The clips should then be set to window this signal level at maximum resolution.

Sensitivity adjustments are especially necessary to compensate for the wide range of attenuations encountered when looking at a colored stimulus with several spectrally selective filters. It is rare that the intensities seen through all filters will simultaneously fit into the widest quantization window at a single setting of camera sensitivity.

VII.6.1 SPECIAL CIRCUMSTANCES

Signal/noise may not always be a problem; red can be reliably distinguished from green with any grossly suitable accommodation. On the other hand, it is possible to distinguish particular cases which warrant careful accommodation to insure adequate results. Obviously, at low illumination levels, color differences are obscured by low signal/noise. A special instance of this problem occurs when attempting subtle discriminations of small color differences.

Subtle variations in hue are usually most distinguished by the presence of minor spectral components at wavelengths considerably removed from the dominant color,

As a result, the most stringent requirements for measurement resolution exist for readings obtained with filters whose passbands exclude most of the energy reflected from the object. In Chapter 3 we demonstrated an inverse relationship between attainable signal/ noise and the prevailing illumination level. In detecting small color differences, we are thus faced with the unfortunate situation in which we require the most signal/noise in those areas of the spectrum in which the light energy is weakest.

Accommodation is also crucial when determining a weakly saturated hue. The problem is best illustrated with reference to the color-cluster space depicted in Figure 7.11. The tolerable error in measuring the coordinates of the specimen must be established in terms of the errors they induce in the direction of the line, defining dominant hue. This error sensitivity is obviously most critical in the case of unsaturated hues because of the leverage generated by their close proximity to the achromatic point. A fortunate coincidence is that unsaturated hues are usually characterized by high reflectance over a broad spectral band. This enhances attainable signal/noise. When brightness is low, however, the effects of noisy measurements are especially serious.

It should be remembered that measurement accuracy can always be improved somewhat by reducing the noise with temporal or spatial averaging. We have indicated

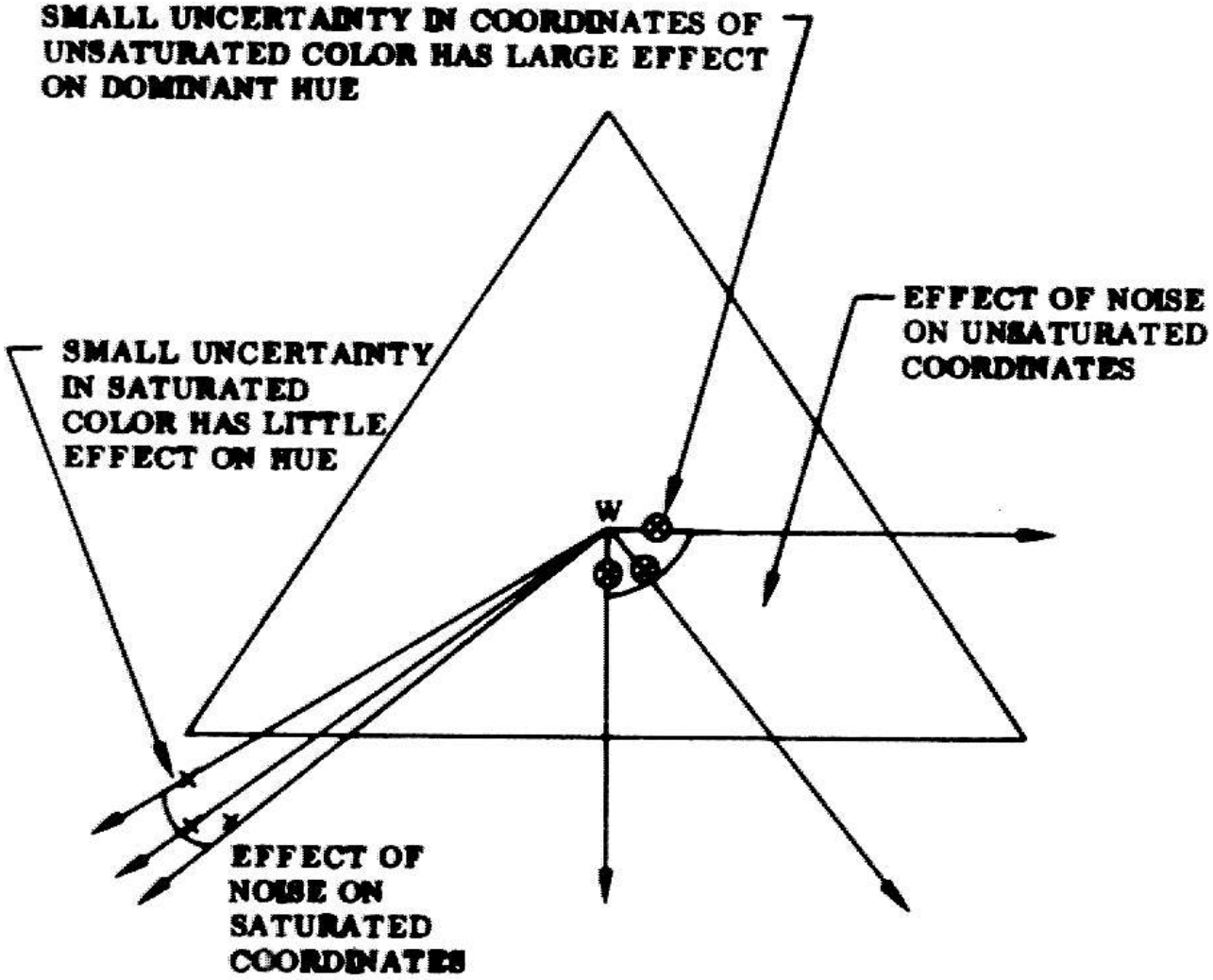


Fig. 7.11 Difficulty of Accurately Determining the Hue of Unsaturated Colors

situations in color perception where such enhancement may be desirable.

VII.7 COLOR RECOGNITION SYSTEM

A system organization that encompasses the major considerations raised in this chapter is shown in Figure 7.12. The organization again reflects the concept of performance feedback. After initialization, the color coordinates of an unknown stimulus are obtained and a tentative color decision is offered. The decision is evaluated in terms of the measured saturation and the recognition match confidence. If there is any reason to doubt its validity, the diagnosis routines in the lower left-hand corner determine whether confidence can be improved by better accommodation or by recalibrating the source. Unlike previous sub-systems, there is no failure exit. If the machine remains unsure after considering these two possibilities, it interacts with the experimenter to refine its comprehension of color categories. This assures the improved confidence.

VII.7.1 EVALUATION CRITERIA

When saturation is very low (typically less than .1), hue is not defined. The stimulus is regarded as

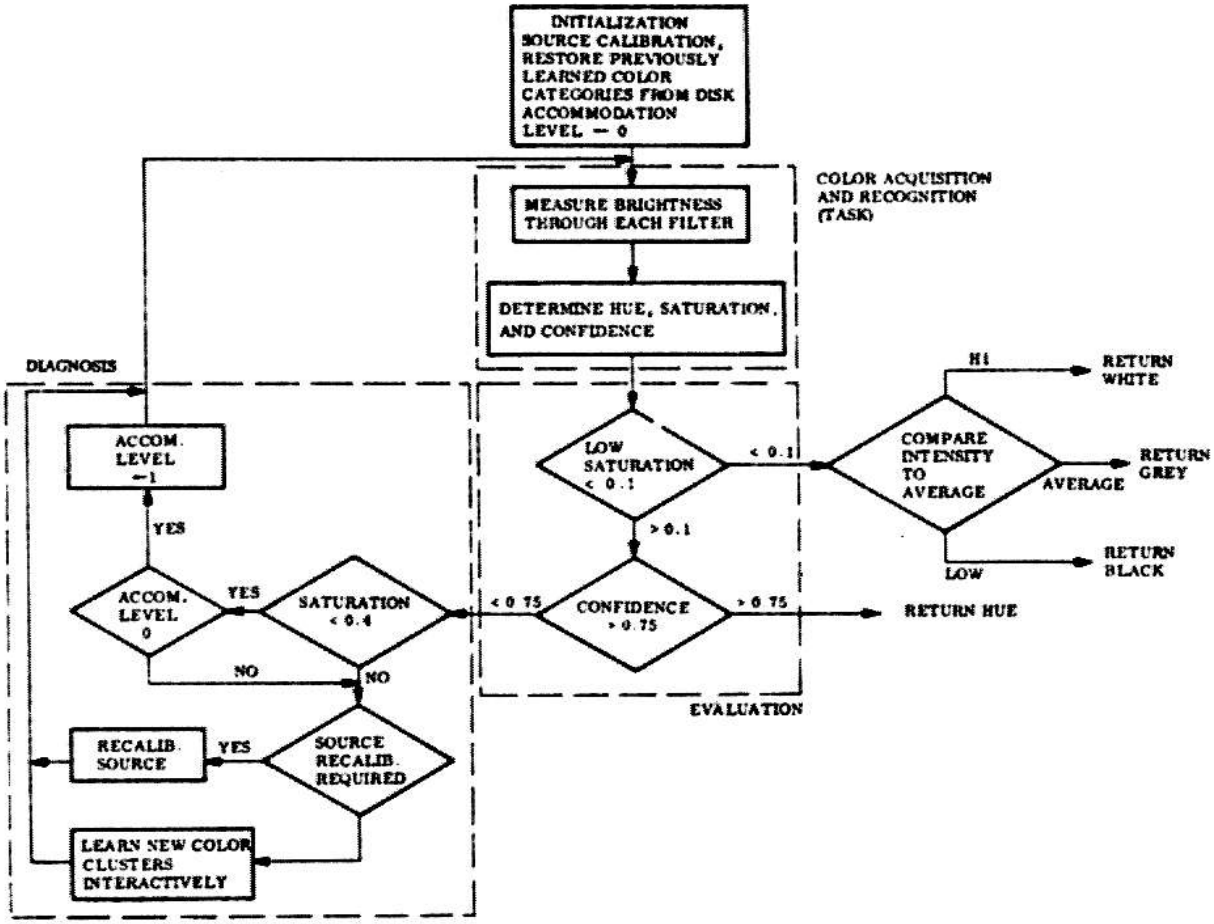


Fig. 7.12 Organization of Formal Color System

achromatic and a tonal grey value is determined. If the stimulus is not achromatic and the match confidence was greater than .75, the chosen hue is considered correct and returned. Otherwise, the diagnosis routine is called to determine whether confidence and/or saturation can be increased by more careful accommodation.

VII.7.2 DIAGNOSIS

Signal/noise, as shown in Figure 7.11, is a serious consideration with weakly saturated signals. When $P \leq .4$, optimal accommodation can significantly improve confidence.

Two levels of accommodation are distinguished. The program is initialized to level zero. In this mode, the quantization window is opened as wide as possible to encompass the largest range of intensities. The sensitivity is then adjusted only when the intensity observed through a filter is hard-clipped by the quantizer. This is an efficient strategy for the majority of cases that do not require optimal tuning. When confidence and saturation are both low, there is a reasonable likelihood that careful accommodation will be able to restore the confidence. In this case, the recognition is repeated using clips and sensitivity optimized for each filter.

If accommodation is not at fault, a likely possibility is that the illumination source has changed

since the last calibration. This suspicion is easily tested. If confirmed, the recognition stage could be repeated using the updated source model to re-normalize the original tristimulus coefficients. (This feature has not yet been implemented. Consequently, the environment is currently constrained to ensure a stable source.)

If the original source calibration was valid, the program finally assigns the fault to an incomplete color categorization. This situation is remedied by interacting with the human operator (or conceivably with a program that may have a reason for preferring one of the two best matches). The interaction may have three possible outcomes:

1. If the color of the stimulus had not previously been seen, a new category is created.

2. If the color of the stimulus is a variant of an existing color, the stimulus coordinates can be added to the cluster for that color. This action will shift the center of gravity towards the stimulus, causing the confidence to improve.

3. If the confidence does not improve sufficiently after the stimulus is added to an existing cluster, a new sub-cluster should be established, distinguished by its own color name. Previously seen samples of this new color should be partitioned from the original cluster and re-assigned to the new category. This refinement is also not yet available.

VII.7.3 ATTEMPTED IMPLEMENTATION OF THE SYSTEM: FAILURE OF PHOTOMETRIC CAMERA MODEL

When this color recognition paradigm was first implemented, its initial testing was performed without benefit of accommodations. The sensitivity was manually set such that the intensities, passed by the three filters, were distributed over the linear range of the quantizer (which had been set for maximum width).

After the feasibility of the approach was established, the standard sensitivity and quantization accommodation routines (used in the verifier,etc.) were installed with the intent of maximizing signal/noise for each filter. However, when sensitivity was adjusted, performance deteriorated.

The problem was attributed to the anomalous reaction of the vidicon's spectral sensitivity to changes in target voltage. This caused the breakdown of the photometric camera model (Equation 2.10) necessary to relate intensities observed at different target voltages to a common brightness scale.

This discovery drastically curtailed the vidicon's effectiveness as a colorimetric device; absolute comparisons of intensities observed through different color filters at optimized target voltages were impossible. For example, the ratio of intensities from a white surface seen through a

red and blue color filter would differ by 30% depending on the target voltage (equal for both filters) at which the comparison was made.

The attempt to implement a general color theory was abandoned because of our inability to make photometrically accurate comparisons over a suitably wide dynamic range. The changes in the vidicon's spectral characteristics with target voltage were experimentally investigated. It was found that certain color discriminations were facilitated by relatively high voltage levels, while others were enhanced by low voltages. We decided to take advantage of this phenomenon by implementing a heuristic "sequential decision" paradigm that converged on the correct color by a process of elimination.

VII.8 HEURISTIC COLOR RECOGNITION

The dependence of the vidicon's relative spectral sensitivity on target voltage poses a dilemma. $W(\lambda)$ enters each of the composite filter functions, $Gamma_i K(\lambda)$. To insure consistency between the filter functions, used in training and in recognition, it becomes desirable to take all measurements with each color filter at the same voltage. Unfortunately, this is inconsistent with the practical requirement that the intensity seen through a filter lie in the linear range of the quantizer.

In general, it will not be possible to observe the intensities from all filters at a single voltage; the sensitivity must be accommodated to achieve the required dynamic range. Thus, accurate tristimulus coefficients cannot always be obtained. However, in many cases, these coefficients are not necessary to determine the closest color match. For example, if the intensity through the red filter is substantially higher than that seen through the blue and green filters, then the color is most likely red. Since the precise intensity margin is not important, the approximate accuracy available with Equation 2.10 is usually sufficient.

VII.8.1 ROLE OF ACCOMMODATION

When discriminating between specific colors, the dynamic range need not encompass the intensities viewed through all filters. In the above example, as long as the intensity viewed through the red filter was relatively high, the decision would not be affected, if the intensities from the blue and green filters fell below the available dynamic range.

On the other hand, to distinguish blue from green, the sensitivity must be set so that the intensities, seen through the corresponding filters, are in the linear quantizer range. However, the reading through the red

filter is then not important.

Color recognition can be organized as a sequential decision process. The dynamic range is set to successively determine whether the unknown color belongs to particular parts of the spectrum. By partitioning the decision in this way, the dynamic range need, at any time, only cover the intensities in, at most, two of the filters.

The goal of accommodation is no longer the optimization of signal/noise in a conventional sense. Its primary purpose is rather to maximize the confidence of a complex decision in color cluster space. The limited dynamic range is adjusted to amplify the contrast between particular colors at the expense of resolution in the rest of the spectrum.

VII.8.2 EVALUATION

Decision confidence is judged on the basis of three criteria. The basic evaluation rests on saturation and confidence which are used in a manner similar to Figure 7.12. However, since the accuracy of the color coordinates is compromised by the limited dynamic range, it is desirable to independently evaluate the reasonableness of the decision. Higher level criteria, such as an expected color, can be used for this purpose.

In the absence of a cohesive vision system, we

relied, for this function, on a heuristic determination of the most likely color, The heuristic decision was obtained by comparing the ordering of intensities observed through the various filters with the orderings associated with particular colors. This technique is especially helpful in eliminating obviously bad choices. For example, "red" need not be considered, unless the intensity seen through the red filter is sufficiently higher than that observed through any other filter.

VII.8.3 IMPLEMENTATION OF A HEURISTIC COLOR RECOGNIZER

We will now describe a specific heuristic color program. The development of this program led to the concepts contained in the last section. This program was created to provide color information for the computer solution of a geometric puzzle known as "Instant Insanity". (The object of this game is to stack four plastic cubes so that four different colors show on each side of the stack.) In this context it was only necessary to distinguish four colors: red, green, blue, and white. The concept of controlling accommodation to maximize a decision criterion can be cogently illustrated in this limited domain.

VII,8,3,1 VIDICON FAULTS

Chart 7.1 indicates the basic problem. In part a, the brightnesses as observed from an achromatic (grey) specimen through four filters are tabulated at five target voltages. Part b was prepared to help interpret this data. Here, the intensities from the colored filters have been normalized by the intensity of the neutral filter at each voltage, providing a consistent standard of comparison. Three characteristics are significant:

1. The effective saturation of the color filters, ie, the ratio between the colored intensity and the neutral intensity, is inversely related to target voltage.

2. The ratio of the normalized contrasts of the red and green filters is relatively independent of target voltage.

3. The normalized contrasts of these two filters, relative to the blue filter, decreases steadily with increasing voltage.

These properties strongly influence the program's accommodation strategy.

VII,8,3,2 DETAILS OF OPERATION

The program (see Figure 7.13) concentrates initially on the red-green end of the spectrum. These colors were the

Filter	Target Voltage				
	10	14	18	22	26
	(volts)				
Red	.191	.321	.476	.697	.943
Green	.275	.445	.634	.849	1.0 *
Blue	.096	.206	.322	.456	.634
Neutral density	.055	.115	.187	.318	.496

* clipped

Chart 7.1a

Video Signal Generated by White Object Through Various Filters as a Function of E_T

Signal Ratios	Target Voltage				
	10	14	18	22	26
R/N	3.47	2.79	2.54	2.19	1.9
G/N	5.00	3.869	3.39	2.669	2.+
B/N	1.74	1.79	1.72	1.433	1.278

N = Signal through neutral density figure

R = Signal through red filter, etc.

Chart 7.1b

Video Signal Through Colored Filters Normalized by Signal Through Neutral Density Filter

(using data from Chart 7.1a)

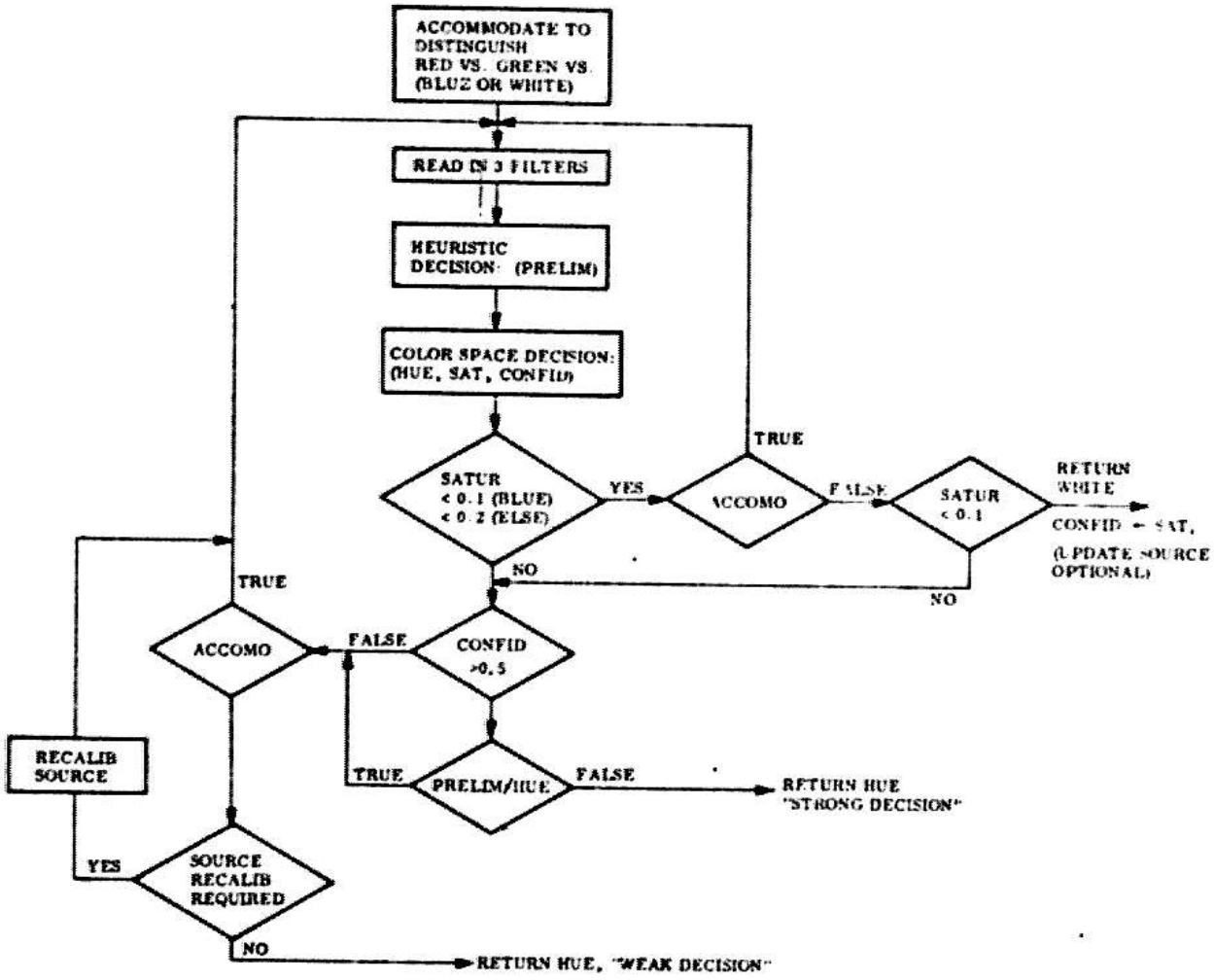


Fig. 7.13 Heuristic Color Program

easiest and most reliable to discriminate, because they happened to be more highly saturated than the blue and obviously the white hues. (Saturation was a particularly troublesome detail, because the glossy surfaces of the Instant Insanity cubes tended to pick up illumination highlights. These diluted the purity of the dominant hue. The blue face, already the least saturated, was the most adversely affected.)

We mentioned the desirability of using a consistent target voltage level. This ideal is compromised in practice by the overriding requirement of obtaining intensities that lie in the linear range of the quantizer. The following strategy consistently selects the lowest practical voltage.

The sensitivity is initially adjusted so that the intensity measured through the blue filter is at the lowest linear quantization level. This setting maximizes the contrast of green and red relative to blue and white (Chart 7.1b). Since blue is nominally the densest filter (Chart 7.1a), the lowest sensitivity at which the blue intensity is observable will maximize the likelihood that the brighter intensities from the other filters will concurrently be in the quantizer window.

The last consideration dictates that the widest quantization window (clipping levels at 0,7) be used. Fine resolution is not needed, because the required color discriminations were not particularly subtle ones. The

crucial limitation was the dynamic range of the quantizer relative to the range of intensities seen through the three colored filters. To illustrate the problem, we note that the intensity of a red object will always be highest (by a comfortable margin) through the matching red filter. However, if the sensitivity is too high, this intensity could easily exceed what the quantizer can handle. The resultant clipping will reduce the contrast between the red filter and the others, decreasing effective saturation. This accommodation error will reduce the confidence of "red" causing potential confusions with "white" and "blue".

Following the initial sensitivity adjustment, the intensities from all filters are observed without re-accommodation. In the event of low confidence, any hard-clipping will arouse strong suspicions in the accommodation diagnosis routine.

VII.8.3.2.1 PRELIMINARY DECISION

A preliminary (heuristic) recognition decision is formulated directly from the relative values of these intensities. The decision process is outlined in Chart 7.2.

In general, the observed intensities may satisfy more than one of the conditions in Chart 7.2. The tests, however, are ordered according to their reliability. Red and green, for example, are usually strongly saturated and

Condition	Decision
1. Green > 1.5 Red	Green
2. Red > 1.5 Green	Red
3. Blue > Green	Blue
4. (Blue > .4 Red) ∧ (Blue > .4 Green)	Blue
5. Else	White

Chart 7.2

Preliminary Decision Table

(exit at first match)

not likely to fail the conditions associated with them. The first successful test determines the preliminary recognition decision. The overall error rate of the preliminary decision was about 9%. Most errors were the result of omissions by the initial tests setting up errors of commission by the later tests (eg. Red was rejected, because it was only 1.4*green.).

This error rate could no doubt have been reduced, if the program were allowed to modify the decision thresholds, based on the success of the final recognition decision. These current thresholds were empirically determined by observing specimens of the four colors under a variety of lighting conditions. Over many trials certain obvious patterns are established as significant indicators of a color, while others were noted to be coincidental. For example, a red object always appeared brighter through a red filter than a green or blue filter. This fact became a necessary condition for "redness". On the other hand, there was no similarly consistent constraint on the relative brightness of a nominally red specimen, viewed through the blue and green filters.

The decisions expressed in Chart 7.2 represent a well-structured inference problem whose solution could conceivably be automated. Becker [1970] and Winston [1970] have proposed schemata that appear to be adaptable to this problem. A program could hopefully handle a more general

formulation, that is, to infer which orderings among an arbitrary number of filters are statistically significant indicators of a color (from a specified set of alternatives).

VII.8.3.2.2 EVALUATION OF CONVENTIONAL RECOGNITION

After the preliminary decision has been made, the conventional recognition process formulates an independent conclusion, based on matching color coordinates. The reasonableness of this conclusion is then determined by considering four factors:

1. match confidence,
2. saturation,
3. hard clipping flag, and
4. agreement with preliminary decision.

The matching decision is accepted forthwith under the following conditions:

1. The saturation is less than .10. This condition must not be an artifact caused by a clipped intensity through one or more of the filters. If the low saturation is legitimate, "white" is returned, independent of what hue was found.

2. The saturation is greater than .1, the match confidence greater than .5, and the preliminary and matching decisions agree. In this case, the recognized hue is

returned.

VII.8.3.2.3 ACCOMMODATION AND DIAGNOSIS

If neither acceptance condition is satisfied, the program attempts to determine whether the problem is related to the initial choice of accommodation. The accommodation diagnosis is also organized in a decision table format (see Chart 7.3). The tests are ordered by the precision with which the corresponding problem can be pinpointed. The most glaring deficiencies, in other words, are detected first.

The accommodation decisions are designed to detect specific problems. The presence of hard-clipping is, for example, not, in itself, sufficient grounds to re-accommodate. Because of the limited decision space, it is only necessary to call the accommodation routine once the initial sensitivity was either too high (causing top end clipping with the red or green filters) or too low (diluting the saturation of a blue sample).

The first two tests are designed to insure that the recognition of red and green were not undermined by hard clipping in the respective matched filters. For example, if the hue was not recognized as red but the intensity in the red filter was at least partially above the quantizer's range, the original decision must be judged inconclusive.

Symptom	Action
1. Accommodation already attempted	1. Return FALSE
2. ((HUE ≠ RED) ∨ ((HUE = RED) ∧ SATUR < .2)) ∧ (I_R > .875 volts) I_R = Tristimulus value observed through Red Filter	1. Switch to Red filter 2. Adjust E_T so that .56 < I_R < .775 3. Return TRUE
3. ((HUE ≠ GREEN) ∨ ((HUE = GREEN ∧ SATUR < .2)) ∧ (I_G > .875 volts)	1. Switch to Green filter 2. Adjust E_T so that .56 < I_G < .775 3. Return TRUE
4. I_B < .125 volts	1. Switch to Blue filter 2. Adjust E_T so that .125 < I_B < .187 3. Return TRUE
5. ELSE	Return FALSE

Chart 7.3

Accommodation Decision Table

Similarly, if the hue was identified as red, but with low saturation, one cannot decide whether the specimen was really white, until the intensity, seen in the red filter, is not artificially clipped. The remedy in both of these cases is to select the red filter and lower the sensitivity, until the intensity is comfortably in the quantizer's range. The recognition cycle is then repeated with the original source of error eliminated.

If neither test 1 or 2 applies, it is improbable that the actual hue was red or green. The program must now decide whether the stimulus is most likely to be blue or white. If the intensity distribution seen through the blue filter is initially below the quantizer's range, this would have an adverse effect on the saturation level necessary to distinguish it from white. Furthermore, as can be seen from Chart 7.1, a higher target voltage improves the relative sensitivity of the camera to blue. Thus, after red and green have been eliminated, the target voltage can be increased to maximize the likelihood of detecting blue, if it is present.

VII.8.3.2.4 OTHER SOURCES OF ERROR

If none of the above situations applies, it is more likely that source calibration or an illegally colored specimen (rather than poor accommodation) are responsible

for the weak results. If the source calibration is confirmed, the program currently returns its best guess, with an indication that the result is questionable.

A strategy program can apply higher level constraints to determine whether the color is reasonable. For instance, if the internal models of the Instant Insanity cubes are available, the returned color may be allowed or disqualified on the grounds of whether it corresponds, with other observed colors, to a known cube. If the color does not make sense at a higher level, we plan to have the hand reposition the cube such that the questionable face is directly illuminated by the calibrated light source. This correction tends to neutralize the principal source of error, introduced by reflections from neighboring objects of different color, picked up by the surface gloss.

CHAPTER VIII: CONCLUSION

Accommodation is inseparable from perception. The performance of every perceptual function depends implicitly on the appropriateness of the visual information available to it.

Every function has unique information requirements. Resolution is most important for some purposes, dynamic range for others. The limited capabilities of the camera and data channel make it unfeasible to include the information needs of all functions in a single image. The serial nature of present digital computers make it undesirable to do so. By matching the information provided by the sensor with that required by a specific function, accommodation ovecomes hardware limitations while improving performance reliability and efficiency.

Acommodation should not be treated in isolation, as a preliminary to the actual recognition process. The information requirements of a function depend on the nature of the scene and, in general, cannot be anticipated. Furthermore, as the machine's characterization of a scene improves, the accommodation can be refined. Accommodation, to be effective, must thus be an integral part of the perceptual strategy.

In this thesis we have proposed a general paradigm for accomplishing this integration and have demonstrated it

in a variety of perceptual functions. Accommodation is set so that the image predicted by a comprehensive model of the visual channel satisfies the anticipated information requirements of a specific perceptual function. The paradigm rests on the assumption that the most reliable evaluation of accommodation in this context is based on the performance of the current function.

VIII.1 ADVANTAGES OF AN ACCOMMODATIVE VISION SYSTEM

1. Level of effort proportional to difficulty of task. Expensive accommodations can be confined to specific regions where more detail is essential. This strategy is not only efficient. It also minimizes the influence of noise by using high sensitivity only when necessary to confirm the presence of an expected feature.

2. Selective Attention. The machine, like a person, comes in contact with much more visual information than it can process. It is only by selecting the most important information that either is able to perceive. The ability of a physical sensor to supply information is also limited. A camera, for example, cannot simultaneously provide high spatial resolution and a wide field of view.

Accommodation allows the computer to concentrate the available camera resources on those aspects of a scene in which it is most interested. Accommodation eliminates

unnecessary detail at the data entry level, in parallel over the entire image. In conventional picture-processing systems the camera is manually tuned to provide the most detailed image. The computer must then eliminate all of the irrelevant detail with (serial) pre-processing algorithms. These systems are overwhelmed in all but the simplest environments. Accommodation can emphasize selected portions of a complex environment to provide images that are suitably simplified for specific perceptual functions.

Accommodation enables the computer to utilize what it knows, a priori, about a feature and the environment to improve its ability to see that feature.

3. Performance limitations. The performance of a perceptual function can be improved either by applying more sophisticated processing to a given image, or by adapting the accommodation to obtain a more appropriate image. Most research in this field has been directed toward the development of more powerful algorithms to cope with inappropriate images. Accommodation, on the other hand, attacks the source of the problem, the quality of the image which poses the fundamental limitation on reliability. Furthermore, the amount of effort devoted to optimizing accommodation will be, in many cases, considerably less than would be necessary to achieve a corresponding increase in performance through processing complexity.

4. Autonomous Operation: Automatic accommodation

eliminates the need for manual intervention in picture processing. In previous systems, the camera was manually adjusted to emphasize the overall image of an object of interest. There was no provision to optimize accommodation for the requirements of specific functions and no possibility of altering the original image if an algorithm failed. Both of these features are implicit in our performance feedback paradigm.

VIII.2 GENERALITY OF ACCOMMODATION

Many of the detailed accommodation algorithms, presented in this thesis, are closely tied to the specific capabilities and limitations of our vidicon camera. Accommodation, however, can be used to advantage in any perceptual system that depends for information on a physical sensor.

Accommodation is necessary when any of the following conditions apply:

1. The useful information in the environment exceeds the maximum dynamic range of the sensor.

2. The static sensor resolution is inadequate over the required dynamic range.

3. The system's information needs vary as a function of the task and contextual information already on hand. Too much information can be as bad as too little.

Conditions 1 and 2 concern limitations of a specific sensor to supply information. Condition 3, on the other hand, concerns the system's need for information. Every sensor has its own particular limitations. The significance of accommodation is not that it overcomes sensor limitations but rather that it matches the capabilities of a sensor with the information requirements for a specific task. In our work with a vidicon camera, it was often possible to capitalize on specific limitations. For example, selective acquisition of visual features could be substantially simplified by utilizing the limited dynamic range to exclude undesired scene characteristics from the image.

The performance feedback paradigm is a general model of accommodation that could be adapted to different sensors in a variety of perceptual domains. Two examples are given.

VIII.2.1 IMAGE DISSECTOR TELEVISION CAMERA

This camera (also known as a vidisector) is used by the vision research programs at MIT and Carnegie-Mellon University. As described in a memo by Horn [1968], it is distinguished from our vidicon system by three principal characteristics:

1. The vidisector is a random access device that provides an accurate photometric intensity at any

single point in the field of view. Spatial averaging is less efficient than temporal averaging because of the time required by the vidisector to change image coordinates.

2. There is no equivalent to our camera's sensitivity control. The dissector measures brightness in terms of the time taken to observe a fixed number of photons. This method inherently accommodates for dynamic range. Darker intensities, for example, simply take longer to encode. (A dark cutoff limits the maximum time that is spent on a single point.) In effect, weaker signals are subjected to more temporal averaging. It can be shown that the ratio of signal/noise will be constant for all brightness levels in this mode of operation.

3. There is no adjustable quantizer window. Intensity resolution is instead selected by setting the required photon count (which is equivalent to selecting the required signal/noise levels). Higher resolution is thus obtained at the expense of measurement time.

The interval between measurements can be reduced, if necessary, so that the number of bits required to encode the increased resolution will not overload the data channel. (The trade-off between data rate and signal quality can be resolved according to the priorities of a specific task. This advantage is not enjoyed by scanning cameras, such as the vidicon.)

Most of the accommodation algorithms described in

this thesis can be adapted to operate with a vidisector:

1. Sensitivity adjustments must be eliminated.

2. The eight degrees of quantizer resolution possible in our system must be equated with equivalent signal/noise thresholds in the vidisector.

3. The concept of an intensity window for applications like selective attention is obtained in two steps. The dark cutoff defines the darkest object of interest, analogous to the bottom clipping level of the quantizer. The equivalent of an upper clipping level must, however, be simulated by software.

4. If the vidisector is fitted with accessories similar to those of our camera (eg. filter wheel, lens turret, focus drive, pan-tilt head, etc.), the associated accommodation heuristics are unchanged.

If an optimal strategy is desired for the vidisector, it is necessary to consider in detail such factors as the relative cost-effectiveness of different modes of averaging. Even so, changes will be minor.

VIII.2.2 SPEECH PREPROCESSOR

In speech recognition, as in visual perception, the primary problem is too much raw information. To represent a speech wave with good fidelity requires about 20,000 9-bit

samples/second, Yet the intelligible information contained in one second of speech is probably less than 20 bits.

Reddy and Vicens [1968] describe a successful speech recognition system based on a preprocessor that reduces the bandwidth of the initial speech intelligence to 3600 bits/second. The speech waveform was crudely divided into three bands (150-900 cps, 900-2000 cps, and 2000-5000 cps). These bands contain the first three formants of typical adult male speech. Recognition was based only on the number of zero crossings and height of the amplitude envelope in each of the bands, sampled every 10 milli-seconds.

One of the problems in using this system with a large population is that the formants of a particular person may be not be centered in the fixed bands. Subtle sound discriminations, such as those involving fricatives, can be confused.

A significant improvement in performance can be expected, if the filter bands can be accommodated to a particular speaker. The formant frequences could be derived from vowel sounds where they are most pronounced. The filter passbands could then be tuned for that speaker, improving performance on marginal phonetic distinctions. If recognition performance shows subsequent signs of deterioration, the filters can be re-accommodated.

Accommodation can also be helpful prior to the

actual extraction of signal parameters. The current sophistication of automated speech recognizers, as typified by the system of Reddy and Vicens, demands clear articulation. The performance of these systems is severely compromised by any background noise, even if totally unrelated to speech (eg, the normal clatter of a computer room).

Humans have the ability to avoid distraction by tuning out unwanted sounds. This phenomenon is most noticeable at cocktail parties and has, in fact, been studied by Cherry [1953] in that context.

In humans most of the selection is thought to take place at a reasonably high cognitive level. However, receptor adjustments such as turning one's ear towards the sound source, obviously contribute. Experimental evidence suggests that man utilizes virtually every physical characteristic of a sound source (location, intensity, frequency, etc.) to aid discrimination.

A computer that could utilize similar characteristics to distinguish desired speech from unwanted sound energy would be at a considerable advantage. Because of the primitive cognitive development of machines, peripheral masking is presently the most convenient way to discriminate unwanted sound. One could imagine a number of easy to implement, computer-controlled accommodations (directional microphone, adjustable cutoff high and low pass

filters, intensity threshold, etc.) that would be effective. Using performance feedback, the computer could adjust these parameters to maximize its rate of recognition.

VIII.3 FUTURE DIRECTIONS

The research described in this thesis was limited chiefly by constraints of time, hardware capabilities, and the current sophistication of the hand-eye vision system. As these constraints are relaxed, the following extensions appear promising.

VIII.3.1 ADDITIONAL ACCOMMODATIONS

The present study concentrated on parameters which could be conveniently adjusted on the existing hardware. Ideally, it would be nice if all factors which influence the information content of an image were under control of the computer.

Some useful, additional accommodations for a vidicon camera might include:

1. variable spatial sampling density,
2. variable bandwidth video amplifier,
3. variable gain video amplifier,
4. variable size scanned area
5. zoom lens with power iris,

6. additional color filters,

7. other optical filters (eg. polaroid), and

8. control of lighting.

The effects of each contemplated accommodation on the characteristics of an image must be determined (qualitatively and quantitatively) in order to relate them to the requirements of perceptual tasks.

Some of the accommodations listed above will overlap some functions of existing accommodations. However, they may be more efficient or have desirable side effects for certain tasks. (For example, the ability to vary spatial sampling density is functionally analogous to changing lens magnification. However, it is likely to be more rapidly variable than a mechanical zoom lens. Furthermore, focal length and resolution affect depth of field in different ways.)

There is, of course, no need to confine attention to a vidicon sensor. The Stanford hand-eye project is currently formulating its requirements for a new camera. The alternatives under consideration range from conventional imaging-type sensors (vidicon, vidisector) to an exotic scheme in which spot illumination, provided by a laser, is detected by a non-directional photomultiplier (Earnest [1967]).

Among the requirements to be established will be a specification of desirable accommodations. In this context,

it is necessary to distinguish which of the functions, presently handled by peripheral accommodations, would be better served by improving the response range of the camera hardware. To answer this question this in a general way we return to the information flow model of selective attention (Figure 1.9).

We established the principle that selection should be deferred to the highest level of processing for which adequate channel capacity exists. Let us assume that the engineering problems regarding the transfer of information from the sensor into memory have been overcome (ie, The bandwidth of the computer's input channel is no longer the limiting factor,). The sole consideration then is what required visual discriminations can be performed more effectively by the sensors than with software. Due to the serial nature of current processing, simple intensity thresholding is the only selection, currently done with accommodation, that can be handled as efficiently by software. (Consider how much more difficult the job of isolating a simple foreground object from a complex background might be, if focus is not used to restrict attention to the range of interest,)

We thus conclude that a large brightness range is a desirable sensor characteristic. There is nothing (except memory capacity) to be gained by artifically restricting the channel data rate, the quantizer resolution, the amplifier

dynamic range, etc. On the other hand, there are strong reasons why the other modes of selectivity (color filter, depth of field, etc.) should be made as selective as possible.

VII,3,2 ADDITIONAL APPLICATIONS

We have demonstrated the important advantages of accommodation in several illustrative perceptual functions. (The development of these specific applications can be continued along lines suggested in each chapter.) Similar advantages can be realized in many other perceptual functions, such as texture recognition, region finding, and stereo range finding.

Each of these functions can be treated individually as was done for the examples in this thesis. A more ambitious goal would be to develop a general "front end" supervisory program. This program would be cognizant of the influence of each camera parameter. The requirements and performance criteria of a new perceptual function could be expressed to it directly, in terms of high level image characteristics. The supervisor would then assume responsibility for formulating a detailed diagnostic and accommodation strategy, taking into account such factors as,

1. the relative cost-effectiveness of each accommodation in the context of the requirements of the

function,

2. prerequisite information requirements (eg. A range estimate is necessary before a lens change in order to re-center the camera.), and

3. supplementary accommodation requirements (eg. Focusing must be done immediately after a lens change.).

VIII.3.3 HIGH LEVEL ACCOMMODATION CRITERIA

In this thesis, we have exploited the concept of performance feedback primarily in the context of local loops, tied to the immediate success of a specific function. The broader concept of hierarchical evaluation is significantly more general.

The program that calls a particular function has its own expectation of how that function should perform. For example, the program that requests the color of the faces on instant insanity cubes might be upset if "yellow" were returned. Similarly, a yet higher program whose job were to determine the orientation of the cube, would object if the colors of the visible faces did not correspond to any of the four known prototype cubes. Both errors raise doubts about the accommodations used in the color program.

The nature of the error can prove helpful in diagnosing the precise deficiency. Work is needed to relate these high level symptoms to the appropriate

corrective accommodations. As the hand-eye system develops, it will become very important to explore the implications of performance feedback at this more global level.

The evaluation hierarchy will extend beyond recognition to the ultimate completion of the task. For example, suppose the hand unexpectedly encounters a physical obstacle at a point in space that previously was observed to be empty. It is possible that the accommodation at that point was initially inadequate. The camera can be re-accommodated until the obstacle can be seen. (Aida and Kinoshita [1970] discuss some hypothetical strategies for tactile-visual symbiosis.)

High level accommodation criteria are necessary in complex problem domains. Human faces, for example, do not lend themselves to hierarchical representations of the type used to develop perceptions of planar-faced solids. Kelly [1972] has formulated a top-down approach to face recognition which looks for complete features (eg. the outline of a head) in a complex scene. To establish context, Kelly's program first looks for a suitably shaped blob in a simplified version of the original image.

Kelly presently tunes the camera (manually) to insure that the desired feature will appear with good contrast in the simplified image. To automate this function, it is necessary to develop a complex criterion of

accommodation that, in effect, simulates Kelly's own global conception of a head. Ideally, global recognition should be the immediate criteria of accommodation. A more practical criterion is the successful detection of a suitably shaped blob. (A defocused image can be used to obtain the necessary reduction in detail for this preliminary phase.)

VIII.3.4 STRATEGIES

Our work, so far, has stressed the effectiveness of accommodation in improving the limiting performance of isolated perceptual functions. With the emergence of a unifying vision system, it becomes important to consider

1. the efficiency of alternative accommodation strategies and
2. how accommodation can enhance the cost-effectiveness of of the overall perceptual strategy.

The following questions are provided to convey some of the complexity of these issues:

1. By what combination of processing time, real time, and reliability should cost-effectiveness be measured?
2. If several objects are required for a task, in what order should they be sought?

3. If an object is described by several visual characteristics, in what order should these be sought?
4. Should the search for an object be conducted breadth first (using coarse accommodations over a wide field of view) or depth first (looking hard in a specific area of the table)?
5. In the event of failure, how much effort should be applied to re-accommodation versus alternate processing techniques?

The accommodation capabilities of the system bear strongly on each of these questions and on the ultimate success of machine perception.

BIBLIOGRAPHY

Abbaronte, M., Johnston, E.G., Lee, Y.H., Nagel, R., Rosenfeld,A., and Thurston, M. [1970]. Edge and curve enhancement in digital pictures, Report 70-103, Computer Science Center, University of Maryland, College Park, Maryland.

Abou-Taleb, Naim, [1952], Visibility of small color differences against colored background, unpublished Ph.D. dissertation, Stanford University, Stanford, California.

Aida, Shuhei, and Kinoshita, Gen-ichiro. [1970]. Pattern recognition by the symbiosis of visual and tactile senses, PROCEEDINGS THIRD ANNUAL HAWAII CONFERENCE ON SYSTEM SCIENCES, ed. by Bertil M. Granborg, 525-528.

Arkadev, A.G., and Braverman, E.M., [1959], COMPUTERS AND PATTERN RECOGNITION, Washington, D.C.: Thompson.

Asher, Harry, [1961], EXPERIMENTS IN SEEING, Greenwich: Fawcett.

Barlow, H.B. [1957]. Incremental thresholds at low intensities considered as signal/noise discriminations, JOURNAL OF PHYSIOLOGY, 136, 469ff.

Bartley, S.M. [1958]. PRINCIPLES OF PERCEPTION, New York: Harper and Row.

Beardslee, D.C., and Wertheimer, M, (Eds.). [1958]. READINGS IN PERCEPTION, New York: Van Nostrand.

Becker, J.D. [1970]. An information-processing model of intermediate- level cognition, AIM-119, Stanford Artificial Intelligence Project, Stanford University, Stanford, California.

Beisner, H.M. [1968]. A recursive Bayesian approach to pattern recognition, PATTERN RECOGNITION, 1 (1), 13-33.

Bledsoe, W.W., et al. [1959]. Discussion of problems in pattern recognition, EJCC, 233ff.

Bell, D. [1968]. Computer-aided design of image processing techniques, PROCEEDINGS I.E.E./N.P.L. CONFERENCE ON PATTERN RECOGNITION, Institute of Electrical Engineers, London, England.

Bennett, Carol, and Franklin, Norman, [1954]. STATISTICAL ANALYSIS IN CHEMISTRY AND THE CHEMICAL INDUSTRY, New York: John Wiley.

Billingsley, F.C, [1967]. Digital image processing at JPL, PROCEEDINGS OF COMPUTER IMAGING TECHNIQUES SEMINAR, Society of Photo-Optical Instrumentation Engineers, Washington, D.C,

Binford, T, [1970]. A visual preprocesser, Internal Report, Project MAC, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Box, J.E.P, [1954]. DESIGN AND ANALYSIS OF INDUSTRIAL EXPERIMENTS, New York: Oliver Boyd.

Broadbent, D.E, [1968]. Applications of information theory and decision theory to human perception and reaction. In CONTEMPORARY THEORY AND RESEARCH IN VISUAL PERCEPTION, ed, by Ralph Norman Haber, San Francisco: Holt, Rinehart, and Winston, 53-63.

Bruinsma, A.H, [1960]. PRACTICAL ROBOT CIRCUITS, New York: Macmillan.

Buerle, R.L., Daniels, M.V., and Hills, B.L. [1959]. Effect of noise on visual pattern recognition, PROCEEDINGS INTERNATIONAL JOINT CONFERENCE ON A.I., 91-105,

Cheng, George et al, (Eds.), [1968], PICTORIAL PATTERN RECOGNITION, Washington, D.C.: Thompson,

Cherry, E.C, [1953]. JOURNAL OF ACCOUSTICAL SOCIETY OF AMERICA, 25,975ff,

Clowes, M.B, [1969]. Perception, picture processing and computers, In J.S. Collins and D. Michie, (Eds.), MACHINE INTELLIGENCE I, New York: Elsevier.

Cochran, William G. and Cox, Gertrude. [1950]. EXPERIMENTAL DESIGNS, New York: John Wiley.

Cover, T.M, and Hart, P. [1967]. Nearest neighbor pattern classification, IEEE TRANSACTIONS ON INFORMATION THEORY, IT-13, 21-27.

Cox, Arthur, [1966]. PHOTOGRAPHIC OPTICS, New York: Focal.

Dineen, G.P, [1955]. Programming pattern recognition, PROCEEDINGS WJCC, 94-100,

Earnest, L. [1967]. Choosing an eye for a computer, AIM-51, Stanford Artificial Intelligence Project, Stanford University, Stanford, California.

Engel, Charles (Ed.). [1968]. PHOTOGRAPHY FOR THE SCIENTIST, New York: Academic.

Eppler, W.G. [1967]. Electronic X-ray film reader and analyzer, Lockheed Missiles and Space Company, Palo Alto, California.

Falk, Gilbert. [1969]. Some implications of planarity for machine perception, AIM-107, Stanford Artificial Intelligence Project, Stanford University, Stanford, California.

Falk, Gilbert. [1970]. Computer interpretation of imperfect line data as a three-dimensional scene, unpublished Ph.D. dissertation, Stanford University, Stanford, California.

Feigenbaum, E., and Feldman, J. (Eds.). [1963]. COMPUTERS AND THOUGHT, San Francisco: McGraw Hill.

Feldman, J., Feldman, G., Falk, G., Grape, G., Pearlman, J., Sobel, I., and Tenenbaum, J. [1969]. Stanford hand-eye project, PROCEEDINGS INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 521-526.

Fink, D. (Ed.). [1957]. TELEVISION ENGINEERING HANDBOOK, New York: McGraw Hill.

Fischler, M. [1969]. Machine perception and description of pictorial data, Lockheed Missiles and Space Company, Information Sciences, Electronic Sciences Laboratory, Palo Alto, California.

Five papers on photoconductivity. [1951]. RCA REVIEW, XII.

von Foerster, H., and Zopf, G.W., Jr., (Eds.). [1962]. PRINCIPLES OF SELF-ORGANIZATION, New York: Pergamon.

Gibson, James J. [1950]. THE PERCEPTION OF THE VISUAL WORLD, Cambridge: Riverside.

Gill, Aharon. [1959]. Minimum scan pattern recognition, IRE TRANSACTIONS, IT-5, 52-57.

Gill, Aharon, [forthcoming], (not yet titled), unpublished Ph.D. dissertation, Stanford University, Stanford, California.

Goodman, J. [1968], INTRODUCTION TO FOURIER OPTICS, New York: McGraw Hill.

Grape, Gunnar, [1969], Computer vision through sequential abstractions, Stanford Artificial Intelligence Project, Stanford University, Stanford, California (offset).

Grape, Gunnar, [1970], On predicting and verifying missing elements in line drawings, based on brightness discontinuity information from their initial television images, Stanford University, Stanford, California (offset).

Grasselli, A, (Ed.), [1969], AUTOMATIC INTERPRETATION AND CLASSIFICATION OF IMAGES, New York: Academic .

Gregory, R.L. [1966], EYE AND BRAIN, New York: McGraw Hill (World University Library).

Gregory, R.L. [1968], Visual Illusions, SCIENTIFIC AMERICAN, 219, 66-80.

Guzman, A, [1968], Computer recognition of three-dimensional objects in a visual scene, MAC-TR-59(THESIS), MAC, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Haber, R.N, [1968], CONTEMPORARY THEORY AND RESEARCH IN VISUAL PERCEPTION, San Francisco: Holt, Rinehart, and Winston.

Hart, Peter, Nilsson, Nils, and Raphael, Bertram, [1968], A formal basis for the heuristic determination of minimum cost paths, IEEE TRANSACTIONS ON SYSTEMS SCIENCE AND CYBERNETICS, 4 (2), 100-107.

Hart, Peter, [1969], Searching probabilistic decision trees, SRI Project 7494, Technical Note 2, AI group, Stanford Research Institute, Menlo Park, California.

Hawkins, J.K., and Munsey, C.J. [1964], Eulogismographic non-linear optical image processing for pattern recognition, JOURNAL OF OPTICAL SOCIETY OF AMERICA, 54, 998-1003.

von Helmholtz, H.(Ed. by R.C.S. Southall). [1963]. HANDBOOK OF PHYSIOLOGICAL OPTICS, London: Dover Reprint.

Helson, H. [1959]. Adaptation level theory in psychology: a study of a science, in S.Koch (Ed.), SENSORY, PERCEPTUAL, AND PHYSIOLOGICAL FORMULATIONS, I, New York: McGraw Hill, 565-621.

Helson, H. [1943]. Some factors and implications of color constancy, JOURNAL OF OPTICAL SOCIETY OF AMERICA, 33, 555.

Herskovitz, A. [1970]. Verification, Project MAC-182, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Heuckel, M. [1969]. An operator which locates edges in digitized pictures, AIM-105, Stanford Artificial Intelligence Project, Stanford University, Stanford, California.

Hildebrand, F.B. [1952]. METHODS OF APPLIED MATHEMATICS, Englewood Cliffs: Prentice-Hall.

Hochberg, Julian, [1964]. PERCEPTION, Englewood Cliffs: Prentice-Hall.

Horn, B.K.P. [1968]. Focusing, Project MAC-160, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Judd, Deanne. [1960]. Appraisal of a land's work on two-primary color projections, JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, 50 (3), 254-268.

Julesz, B. [1962]. Visual pattern discrimination, IRE TRANSACTIONS, IT-8, 84-92.

Kanal, L. (Ed.). [1968]. PATTERN RECOGNITION, Washington, D.C.: Thompson.

Kelly, Michael. [1970]. Visual identification of people by computer, unpublished Ph.D. dissertation, Stanford University, Stanford, California.

Kiver, Milton. [1964]. COLOR TELEVISION AND FUNDAMENTALS, New York: McGraw Hill.

KODAK WRATTEN FILTERS FOR SCIENTIFIC AND TECHNICAL USE. [1968], Eastman Kodak.

Kolers, P.A., and Eden, M.A. [1968]. RECOGNIZING PATTERNS, Cambridge: Massachusetts Institute of Technology.

Land, E.H. [1959]. Color vision and the natural image (Part I), PROCEEDINGS NATIONAL ACADEMY OF SCIENCE, 45, 115-129.

Larmore, Lewis. [1965]. INTRODUCTION TO PHOTOGRAPHIC PRINCIPLES, New York: Dover.

Lettvin, J.Y., Maturana, H.R., McCulloch, W.S., and Pitts, W.H. [1959]. What the frog's eye tells the frog's brain, PROCEEDINGS OF THE INSTITUTE OF RADIO ENGINEERS, 47, 1940-1951.

Levin, R.E. [1968]. An investigation of controlling source spectra for maximizing small colour differences, unpublished Ph.D. dissertation, Stanford University, Stanford, California.

Linfoot, E.H. [1964]. FOURIER METHODS IN OPTICAL IMAGE EVALUATION, New York: Focal.

McCormick, B.H. [1963]. Illinois pattern recognition computer-Illiac III, IEEE TRANSACTIONS, EC-12, 791-813.

McCarthy, J., Earnest, L.D., Reddy, D.R., and Vicens, P.J. [1968]. A computer with hands, eyes, and ears, PROCEEDINGS AFIPS 1968 AND FALL JOINT COMPUTER CONFERENCE (33), Washington, D.C.: Thompson .

Marill, T., and Bloom, B. [1966]. Learning and perceptual processes for computers, ANNALS OF NEW YORK ACADEMY OF SCIENCE, 128, 1029-1034.

Miles, T.R., and E. (Eds.). [1963]. THE PERCEPTION OF CAUSALITY, London: Methuen.

Miller, W.F., and Shaw, A.C. [1968]. Linguistic methods in picture processing, PROCEEDINGS FJCC, 279-290.

Minsky, M., and Papert, S. [1969]. PERCEPTRONS, Cambridge: Massachusetts Institute of Technology.

Moles, Abraham (trans. by Joel Cohen). [1968]. INFORMATION THEORY AND ESTHETIC PERCEPTION, Urbana: University of Illinois.

Montanari, U. [1968]. A method for obtaining skeletons using a quasi-Euclidean distance, JOURNAL OF ACM, 15, 600-624.

Munn, Norman L. [1961]. PSYCHOLOGY:THE FUNDAMENTALS OF HUMAN ADJUSTMENT, Boston: Houghton Mifflin.

Nadler, M. [1968]. Empyrean, an alternative paradigm for pattern recognition, PATTERN RECOGNITION, 1(2), 147-165.

Nagy, G. [1968]. State of the art in pattern recognition, PROCEEDINGS IEEE, 56, 836-862.

OPERATING AND MAINTENANCE INSTRUCTIONS FOR 3100 SERIES OF TELEVISION CAMERAS, [1964], Conu Electronics,

Pingle, K., Singer, J.A., and Wichman, W.M. [1968], Computer control of a mechanical arm through visual input, PROCEEDINGS IFIP CONFERENCE, Edinburgh, H140-H146,

Pingle, K. [1966], A program to find objects in a picture, AIM-39, Stanford Artificial Intelligence Project, Stanford University, Stanford, California,

Pingle, K,, and Tenenbaum, J. An accommodative edge follower, Stanford Artificial Intelligence Project, Stanford, California (forthcoming).

Prewitt, Judith, [1966], Advances in biomedical computer applications, NEW YORK ACADEMY OF SCIENCE,

Pylyshyn, Zenon, [1970], PERSPECTIVES IN THE COMPUTER REVOLUTION, Englewood Cliffs: Prentice-Hall,.

PROCEEDINGS OF IMAGE ENHANCEMENT SEMINAR, [1963], Society of Photo-optical Instrumentation Engineers,

Reddy, D.R,, and Vicens, P.J. [1968], A procedure for segmentation of connected speech, JOURNAL OF AUDIO ENGINEERING SOCIETY, 16,

Reddy, D.R, [1969]. On the use of environmental, syntactic, and probabilistic constraints in vision and speech, AIM-78, Stanford Artificial Intelligence Project, Stanford University, Stanford, California,

REFERENCE DATA FOR RADIO ENGINEERS, [1956]. New York: International Telephone and Telegraph Corporation,

RESEARCH ON UTILIZATION OF PATTERN RECOGNITION TECHNIQUES TO IDENTIFY AND CLASSIFY OBJECTS IN VIDEO DATA, [n.d.]. Report SM-48464CF, Douglas Aircraft Company.

Roberts, L.G, [1965]. Machine perception of three-dimensional solids, In J. Tippet et al (Eds.), OPTICAL AND ELECTRO- OPTICAL INFORMATION PROCESSING, Cambridge: Massachusetts Institute of Technology, 159-197.

Rose, A, [1948]. Television pick-up tubes and the problem of vision, In L, Marton (Ed.), ADVANCES IN ELECTRONICS, New York: Academic, 131-167.

Rosen, Charles A, [1967]. Pattern classification by adaptive machines, SCIENCE, 156, 38-44.

Rosenblith, Walter A, (Ed,). [1961]. SENSORY COMMUNICATION, Cambridge: Massachusetts Institute of Technology.

Rosenfeld, Azriel, Huang, Han, and Schneider, Victor, [1968]. An application of cluster detection to text and picture processing, Technical Report 68-68, Computer Science Center, University of Maryland, College Park, Maryland.

Rosenfeld, A., and Pflatz, J.L, [1968]. Distance functions on digital pictures, PATTERN RECOGNITION, 1 (1), 33-63.

Rosenfeld, Azriel, Thomas, Richard B., and Lee, Yung, [1969]. Edge and curve enhancement in digital pictures, Technical Report 69-73, Computer Science Center, University of Maryland, College Park, Maryland.

Rosenfeld, Azriel, [1968]. Picture processing by computer, Technical Report 68-71, Computer Science Center, University of Maryland, College Park, Maryland.

Schade, Otto, [1948]. Electro-optical characteristics of television systems, RCA REVIEW, IX, (four articles),

SCIENCE OF COLOR, [1958]. OPTICAL SOCIETY OF AMERICA,

Schreiber, W, [1964]. Vidicon noise. PROCEEDINGS IEEE, 52 (1), 217.

Sebestyen, G, [1962]. DECISION-MAKING PROCESSES IN PATTERN RECOGNITION, New York: Macmillan.

Sevin, Leonce J, [1965]. FIELD-EFFECT TRANSISTORS, New York: McGraw Hill.

Shaw, Alan C, [1968]. The formal description and parsing of pictures, SLAC Report 84, Stanford Linear Accelerator Center, Stanford University, Stanford, California.

Sheppard, J.J.,Jr. [1968]. HUMAN COLOR PERCEPTION, New York: Elsvier.

Singer, J.R, [1961]. Electronic analog of the human recognition system, JOURNAL OF OPTICAL SOCIETY OF AMERICA, 51, 61-69.

Sobel, Irwin, [1970]. Camera models and machine perception, AIM-121, Artificial Intelligence Project, Stanford University, Stanford, California.

Soule, Harold, [1968]. ELECTRO-OPTICAL PHOTOGRAPHY AT LOW ILLUMINATION LEVELS, New York: John Wiley.

Susskind, Alfred (Ed.), [1957]. NOTES ON ANALOG-DIGITAL CONVERSION TECHNIQUES, New York: Massachusetts Institute of Technology and John Wiley.

Swets, John A., Tanner, Wilson P., and Birdsall, T.C, [1961]. Design processes in perception, PSYCHOLOGICAL REVIEW, 68 (5), 301-320.

Sutherland, Ivan, [1966]. Computer graphics: ten unsolved problems, DATAMATION, 12 (5), 22-27.

Teevan, R.C., and Birney, R.C. (Eds.). [1961]. COLOUR VISION, New York: Van Nostrand.

Tou, J.T. [1968]. Feature extraction in pattern recognition, PATTERN RECOGNITION, 1 (1), 3-13.

Triesman, Ann, [1968]. Selective attention in man, in Ralph Norman Haber (Ed.), CONTEMPORARY THEORY AND RESEARCH IN VISUAL PERCEPTION, New York: Holt, Rinehart, and Winston, 258-266.

Uhr, Leonard (Ed.). [1966]. PATTERN RECOGNITION, New York: John Wiley.

Vicens, Pierre, [1969]. Aspects of speech recognition by computer, AIM-85, Stanford Artificial Intelligence Project, Stanford University, Stanford, California.

Wald, A. [1947]. SEQUENTIAL ANALYSIS, New York: John Wiley.

Walls, Gordon L. [1960]. Land! land!, PSYCHOLOGICAL BULLETIN, 57 (1), 29-48.

Weinberg, George H., and Schumaker, John A. [1962]. STATISTICS: AN INTUITIVE APPROACH, Belmont: Wadsworth.

White, B.W. [1961]. The computer as a pattern generator for perceptual research, BEHAVIORAL SCIENCES, 6, 252-259.

White, David. [1949]. Maximizing color differences, unpublished Ph.D. dissertation, Stanford University, Stanford, California.

Wichman, W.B. [1967]. Use of optical feedback in the computer control of an arm, AIM-56, Stanford Artificial Intelligence Project, Stanford University, Stanford, California.

Winston, P. [1970]. Learning structural descriptions from examples, unpublished Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Woldrige, D.E. [1963]. THE MACHINERY OF THE BRAIN, New York: McGraw Hill.

Wolfson, M.M. [1959]. Some new aspects of color perception, IBM JOURNAL, 312.

Wright, W.D. [1958]. THE MEASUREMENT OF COLOR, New York: Macmillan.

Yarbus, Alfred L. [1967]. EYE MOVEMENTS AND VISION, New York: Plenum.

Yau, S.S., and Yang, C.C. [1966]. Pattern recognition using an associative memory, IEEE TRANSACTIONS EC-15, 944-947.

Zahn, C.T. [1966]. Two-dimensional pattern description and recognition via curvature points, SLAC Report 70, Stanford Linear Acceleration Center, Stanford University, Stanford, California.