

R-565
DEVELOPMENT OF VISUAL, CONTACT
AND DECISION SUBSYSTEMS FOR A MARS ROVER
July 1966 to January 1967

by

Louis L. Suto Roberto Moreno-Diaz
William L. Kilmer Warren S. McCulloch, et al.

July 1967

**INSTRUMENTATION
LABORATORY** ●

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Cambridge 39, Mass.

R-565

DEVELOPMENT OF VISUAL, CONTACT
AND DECISION SUBSYSTEMS
FOR A MARS ROVER

July 1966 to January 1967

by

Louis L. Sutro	Joseph J. Convers
William L. Kilmer	David Lampert
Roberto Moreno-Diaz	Robert J. Magee
Warren S. McCulloch	Charles D. Sigwart
James G. Bever	David G. Tweed
Jay Blum	

July 1967

A report of work performed from
July 16, 1966 to January 15, 1967 under
NASA Contract NSR 22-009-138,
and from May 1 to December 31, 1966
under NASA Grant NGR 22-009-140.

INSTRUMENTATION LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

*Prepared for publication by Jackson & Moreland
Division of United Engineers & Constructors, Inc.*


Approved


Professor of Aeronautical
and Astronautical Engineering

Approved


Associate Director

Approved


Deputy Director

ACKNOWLEDGEMENT

Section 1, part of Section 2, Sections 3, 4 and 5, part of 7 and Section 8 were prepared under the auspices of DSR Project 55-257, sponsored by the National Aeronautics and Space Administration, Headquarters, Office of Space Science and Applications, through Contract NSR 22-009-138.

Part of Section 2, Section 6, part of Section 7 and Section 9 were prepared under the auspices of DSR Project 76370, sponsored by the National Aeronautics and Space Administration, Electronics Research Center, Cambridge, Massachusetts, through Grant NSR 22-009-140.

All of the authors are either members of, or consultants to, the Longspur Section of the Instrumentation Laboratory. Two additional consultants are

Dr. Carl Sagan of Harvard University and the Smithsonian Astrophysical Observatory

Prof. Calvin Burnett of the Massachusetts College of Art who drew Figs. 2-3, 5-1, 5-2, 5-8, E-1 and E-2.

Further acknowledgement is given at the end of Section 5 to those who assisted in the preparation of that section.

TABLE OF CONTENTS

PART I DEVELOPMENT

<u>Section</u>		<u>Page</u>
1	INTRODUCTION	1
2	SUBSYSTEMS AND THEIR COORDINATION	3
	2.1 General	3
	2.2 Robot Subsystems	3
	2.3 The LOOK Mode of Operation	6
	2.4 Mounting a Stereo Pair of Cameras on a Vehicle	6
	2.5 Contact Subsystem	8
	2.6 Coordination of the Subsystems	12
	2.7 Digital Computers	12
3	CAMERA-COMPUTER CHAINS	14
	3.1 Objectives	14
	3.2 Camera-Computer Chain A	15
	3.3 Camera-Computer Chain B	17
	3.4 Designation of Stereo Optics and Camera Electronics	17
	3.5 Type 1 Camera Electronics	18
	3.6 Uncertainties Due to the Type 1 Camera Electronics	20
	3.7 Optics of Camera-and-Stereo-Attachment Assembly C1	20
	3.8 Anticipated Ranging Ability of Camera-Computer Chain C1-1	23
	3.9 Camera-and-Gimbal Assembly D1	25
	3.10 Anticipated Ranging Ability of Camera-Computer Chain D1-3	26
	3.11 Camera-and-Gimbal-Assembly E1	28
4	VISUAL DATA PROCESSING	32
	4.1 General	32
	4.2 Characteristics of the Visual Processing Program	33
	4.3 Initial Approach	33
	4.4 A Second Approach	35

<u>Section</u>	<u>Page</u>
4.5 The Present Approach	36
4.6 Range Finding Planned for Camera-Computer Chains C and D.	36
5 TOWARD A METHOD OF MODELLING AN ENVIRONMENT . .	39
5.1 General.	39
5.2 Proposed Method of Acquiring a Point Model	39
5.3 An Example.	40
5.4 Interrogation of a Point Model.	41
5.5 Interrogating a Point Model in the Search for Evidence of Life	43
5.6 Mapping.	43
5.7 Representation by Points and Reflectances	44
5.8 Computation of Reflectance	44
5.9 Computation of the Diffuse Reflectance of ΔFGJ	48
5.10 Interrogation of a Point-and-Reflectance Model.	48
5.11 The Earth Station.	49
5.12 Acknowledgement.	49
6 ADDITION OF A LEARNING ABILITY TO THE DECISION-MAKING SUBSYSTEM	51
6.1 Introduction.	51
6.2 Requirements.	51
6.3 Nature of the Problem	52
6.4 Approach to the Problem.	52
6.5 Implementation by Operations on P_{AB} Vectors	52
6.6 Structure of the Model	54

PART II

RESEARCH INTO THE NATURE OF ANIMAL SENSORY,
DECISION AND CONTROL SYSTEMS

7 ANIMAL COMPUTATION.	57
7.1 Background to a Computer Model of a Vertebrate Brain	57
7.2 A Possible Engineering Equivalent	60
7.3 Memory in Biological Computers	60
8 VISUAL PROCESSING IN ANIMALS	62
8.1 Development of a Computational Model	62
8.2 Generalization of Interaction Coefficients.	64
8.3 Examples of Lateral Interaction.	65

<u>Section</u>	<u>Page</u>
9 LEARNING IN ANIMALS	68
9.1 General.	68
9.2 Classes of Response	68
9.3 Classical Conditioning	69
9.4 Habituation	69
9.5 Discriminative Learning.	70
9.6 Instrumental Conditioning	70
10 SUMMARY AND CONCLUSION.	71
<u>Appendix</u>	
A MEASUREMENT OF VIDICON RESOLUTION	73
B RANGING AND RANGE ERROR	75
C OTHER STEREO CAMERA CONFIGURATIONS	83
D PROPOSED OPTICAL TRAIN FOR CAMERA E1	85
E CALCULATION OF THE DIRECTION COMPONENTS OF THE NORMAL TO THE TRIANGLE FGJ AND THE ANGLE BETWEEN THIS NORMAL AND SUNLIGHT.	87
LIST OF REFERENCES	91

PART I

DEVELOPMENT

SECTION 1
INTRODUCTION

by

Louis Sutro

The first of the two efforts described in this report is supported by the Bioscience Division of the Office of Space Sciences and Applications, NASA, and is concerned with communicating pictorial data from a remote location, such as Mars, to Earth. The second, supported by the Electronics Research Center of NASA, is to develop methods of making decisions at remote locations. Both are instrumentation efforts applicable to a Mars rover and both use biology as a source of ideas.

Earlier reports^(1, 2) indicated that the processing of television pictures at a remote location was desirable, and could be facilitated by employing a "rover"; that is, a pair of stereo television cameras on a mobile base. Since further on-board processing was required to position the cameras, the concept evolved of an assembly - to be called "robot" - of stereo cameras, vehicle, and computer. Included in the concept are inputs to the computer from senses other than the visual, namely, accelerometers to indicate the inclination of the cameras and their motion, and a tactile-kinesthetic sense of both the posture of the robot and its contact with the ground.

The word "computer" has been used in two ways in this and earlier reports. As illustrated above, it is used first to describe all equipment employed by the robot for processing or computation. For testing stationary equipment this computer will be the general-purpose machine described in subsection 2.7; for testing mobile equipment it will be special-purpose and transportable. This report presents those specifications presently available for the mobile system, while more detailed specifications are being prepared that will lead to size, weight and power estimates necessary to judge practicality.

The second use of the word "computer" is to describe the kinds of computation performed by animal nervous systems. Section 7 introduces this second use of the word.

Each of the two efforts is being approached from the points of view of both development and research. Although the first six sections of the report, in general describe development and Sections 7 through 9 describe research into the nature of animal sensory, decision and control systems, the line between research and development cannot be sharply drawn.

The sections on development consider first, in Section 2, the tying together of the subsystems. Sections 3, 4 and 5 describe three aspects of the visual subsystem: camera-computer chains, processing and representation of the environment. Section 6 describes additions being made to the decision subsystem.

The first of the research sections - Section 7 - describes Dr. Warren McCulloch's concept of the nervous system as interconnected computers. Section 8 develops mathematically a concept of visual processing, known as lateral interaction, which appears basic to the functioning of higher animal retinas. Finally, Section 9 describes the theory behind animal learning, a body of knowledge which must be understood if an engineering system such as that in Section 2 is to adapt to changes in its environment. Section 10 is a summary and conclusion.

SECTION 2

SUBSYSTEMS AND THEIR COORDINATION

by

Louis Sutro, William Kilmer and Joseph Convers

2.1 General

The system whose parts are described here includes the robot, means of transmitting pictorial and scientific data to Earth, and equipment for decoding it and displaying it to an Earth operator. Subsystems of the robot are outlined in subsection 2.2, after which follow a description of the LOOK mode of operation and proposed means of mounting the visual on the mobility subsystem. Coordination of subsystems is described in principle in subsection 2.6 and simulation within a general purpose computer in subsection 2.7.

2.2 Robot Subsystems

The visual subsystem includes a camera assembly of the kind described in Section 3, processing of the kind described in Section 4, and the making and interrogating of a model of visual and tactile space described in Section 5.

Equipment intended to contact the environment either to report on it or to affect it is called the contact subsystem - examples are the hand and arm assemblies shown at the left center of Fig. 2-1. Since vision, in many ways, is a distant method of feeling the shape of things, coordination is required of the visual and contact subsystems. This appears achievable either by a matching of models of the kind described in Section 5 or by the learning mechanisms described in Section 6.

The mobility subsystem considered during this report period is the vehicle developed by General Motors for a landing on the moon and now being redesigned for a landing on Mars. Following their practice, this subsystem includes the basic structural elements, wheels, wheel drives and steering actuators as illustrated by Fig. 2-2.

The fourth component, the decision subsystem, determines the mode in which the robot operates. For initial tests, only the following four modes will be possible:

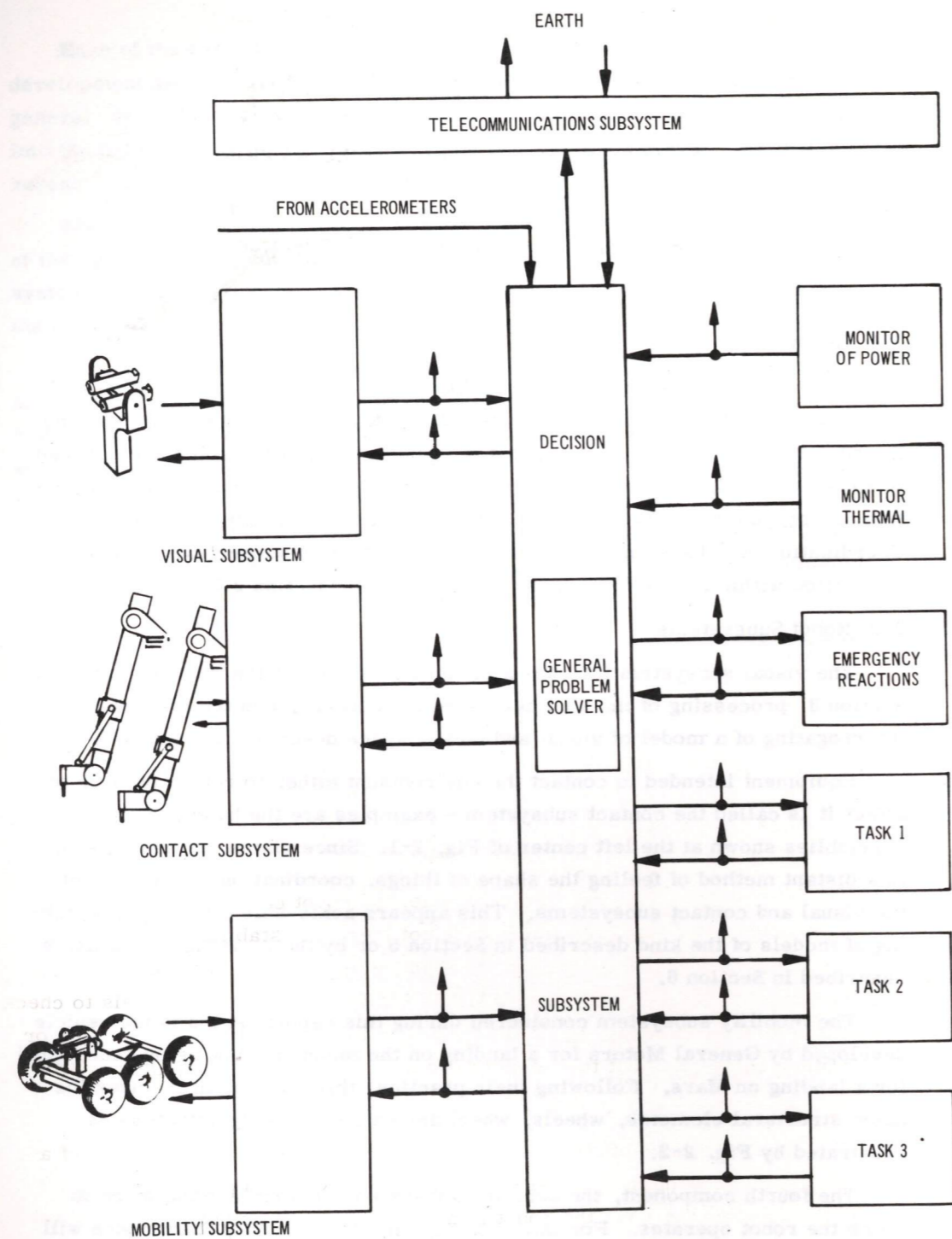


Fig. 2-1. Subsystems of the robot proposed for a Mars landing.

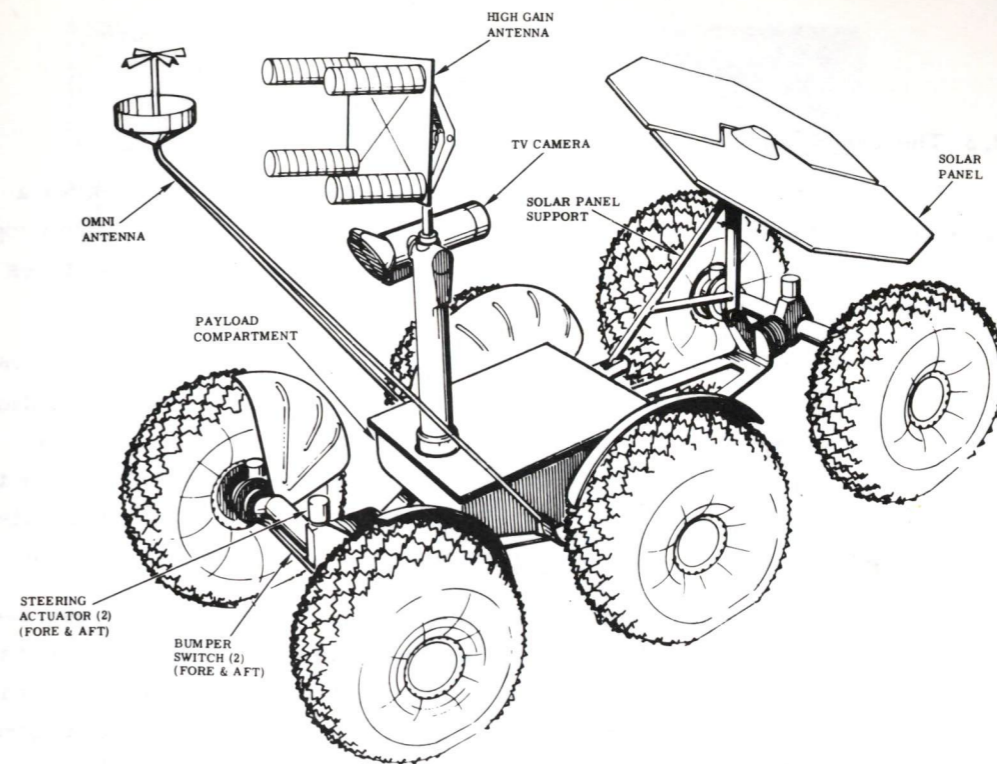


Fig. 2-2. Typical vehicle developed for lunar and planetary exploration by the Defense Research Laboratory, AC Electronics Division, General Motors. Length is 72 in., width 36 in., height of wheels 18 in.

- Look
- Move
- Rest
- Maintenance

These are mutually exclusive. For example, the robot cannot LOOK and MOVE at the same time because the cameras are not inertially stabilized during vehicle motion. The REST mode is planned for night, when lack of light makes looking and therefore movement difficult. MAINTENANCE could use the wheels to check for vehicular trouble or use the cameras either to check their own operation or check other equipment, but the intent will not be the same as the LOOK or MOVE modes.

The robot will have other modes of operation: one to try to get it out of a difficulty such as wedging a wheel between rocks, others to perform specified tasks or experiments. Yet the robot is not planned to be autonomous. The top command remains on earth.

The upward pointing arrows in Fig. 2-1 represent connections that can be made to the telecommunications subsystem, which in a test system being planned would be a microwave link to a nearby control station. A high-gain directional antenna and an omni-directional antenna, both intended for this purpose, are shown in Fig. 2-2.

2.3 The LOOK Mode of Operation

The LOOK mode of operation may take one of these forms: to look for a passage ahead for the robot to follow, to look for objects having specified properties of scientific interest, or to look at the scene or object to record pictures for transmission to Earth.

Figure 2-3 indicates how a scene might appear on Mars in two successive looks (the two are separated by the vertical line at the center of the illustration). At the center of each look is a high-resolution central field of the kind to be achieved by camera assembly E, described in subsection 3.10. Around it is the low-resolution peripheral field. In this illustration the ratio of the central field to the peripheral is made 1/5, whereas in camera E it is planned to be 1/10.

The LOOK mode begins with the levelling of the platform shown in Fig. 2-4. Possible next steps are the pitching of the cameras to scan the lowest line of the nearest rectangle; then turning of the cameras one position at a time, so that their point of convergence steps along a line the width of the rectangle. The illustration shows the cameras in the position where they receive images from the same point on the rock. If this image is an edge, as here, it can be considered a pointer in the direction to be explored next. Following such pointers, the cameras trace the edge of the rock. Then, converging further away, they can sweep the second rectangle to find the hole in the ground. Converging still further away, they can sweep the farthest rectangle to find the edge of the cliff. Since this process may be difficult to realize if images of the scene are crowded, other approaches are being studied.

Points as well as lines may define a surface. Experiments conducted at Lincoln Laboratory, MIT, indicate that white dots against a black background, located in both the left and right views of a stereoscopic picture, can perform the same function as a set of lines. Such dots could also be black against a light background or merely contrast with the adjoining area. Thus in the display on Earth a surface such as Fig. 2-4 could be defined by points without connecting lines.

2.4 Mounting a Stereo Pair of Cameras on a Vehicle

Figure 2-2 shows the proposed General Motors vehicle to which equipment is planned to be attached. A vehicle of similar but earlier design was delivered to Jet Propulsion Laboratory and is operating there.

Two approaches have been made to the mounting of a stereo pair of cameras on such a vehicle. One is to support the stereo pair, as in Fig. 2-5, on a boom, which can be raised to permit looking in the distance or lowered to permit looking immediately in front of the vehicle. The other approach is to mount the pair



Fig. 2-3. Supposed views of the surface to be explored, showing two high resolution looks surrounded by low resolution looks.

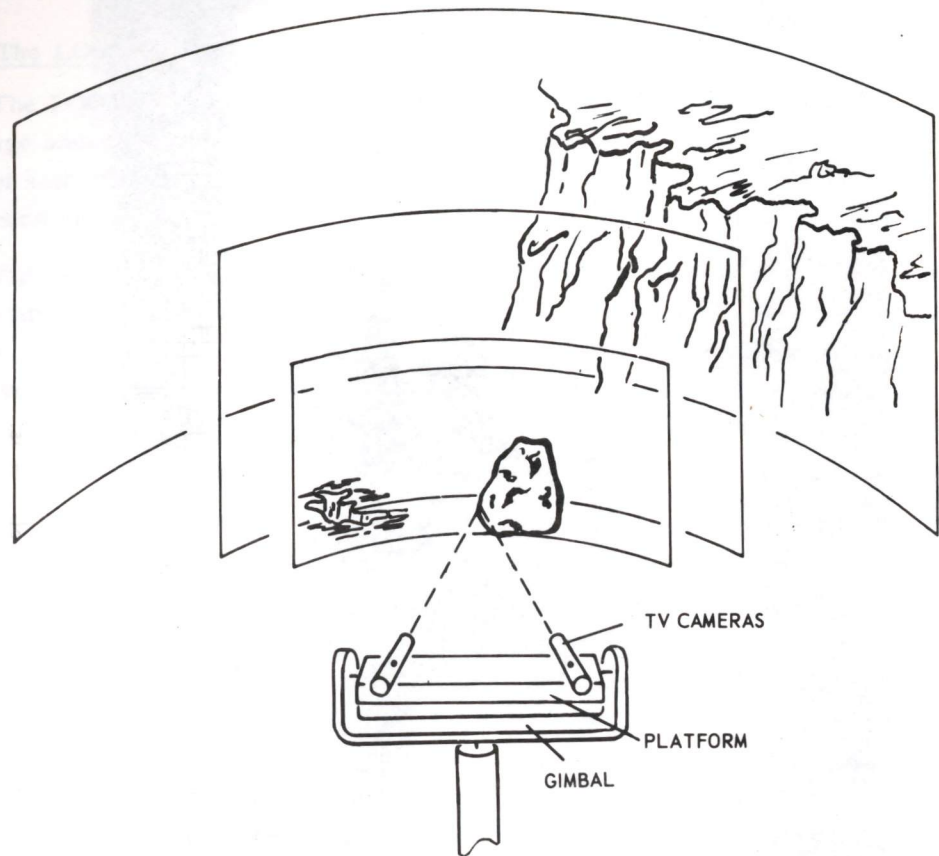


Fig. 2-4. Discrimination of objects by a stereoscopic pair of TV cameras and a computer.

as close to the ground as possible, as in Fig. 2-6. For both approaches it is desirable, after vehicle movement, to level the camera platform and then maintain the pitch axis of the cameras horizontal as they look up or down, sideward, or backward.

The first approach required more mechanical complexity than appeared wise at the start. The second placed the weight of the camera assembly too far forward. Figure 2-7 is a compromise, placing the camera assembly above the axle.

2.5 Contact Subsystem

If recognition of objects is to be performed as it is in animals, more than one sensory input is required. While information provided by a single sense is, in general, inadequate for recognition, the addition of a second sense generates what is necessary to validate the original evidence. With the integration of the contact and the visual subsystems, a reliable recognition capability should be achieved.

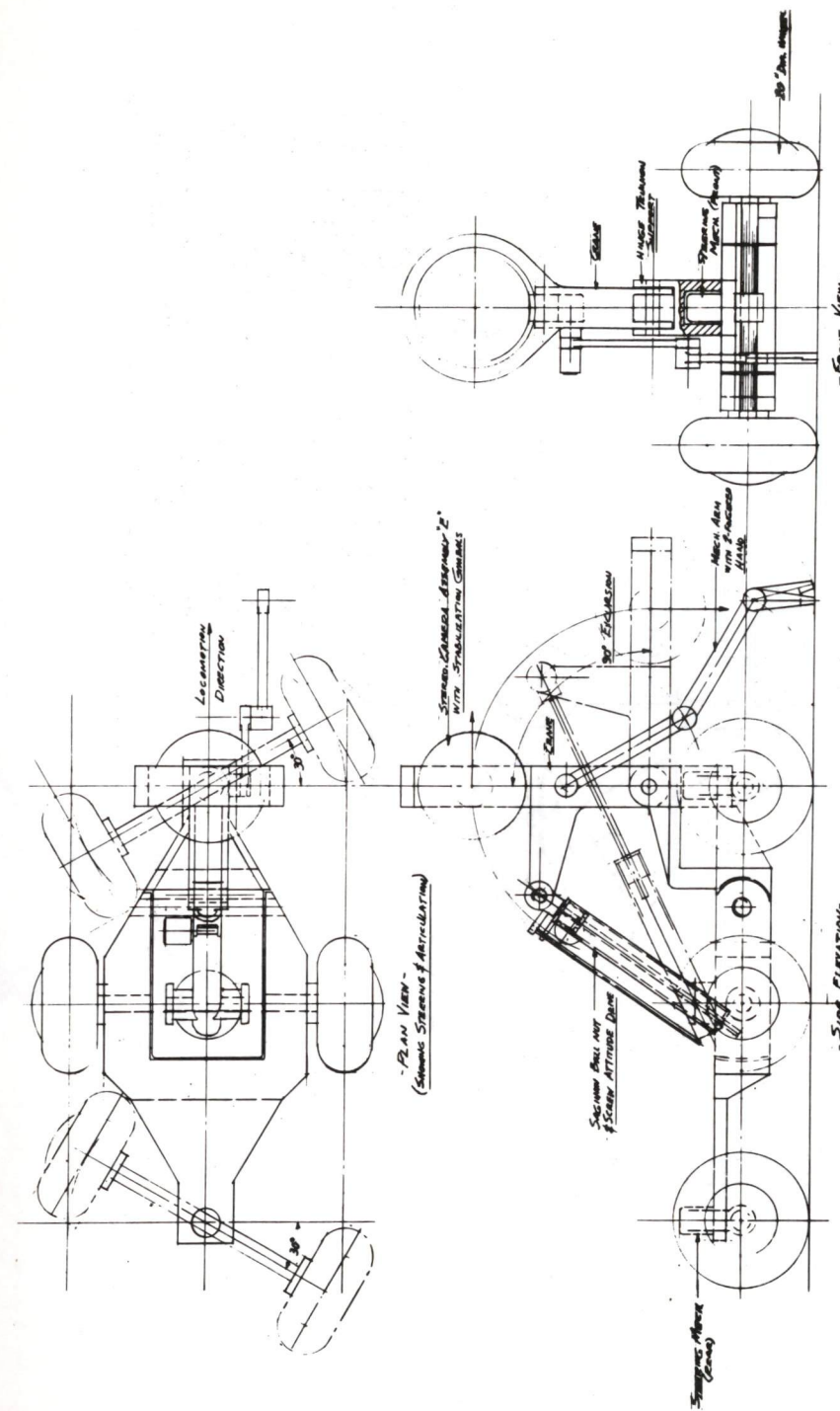


Fig. 2-5. Use of a boom (crane) to elevate stereo pair of cameras.

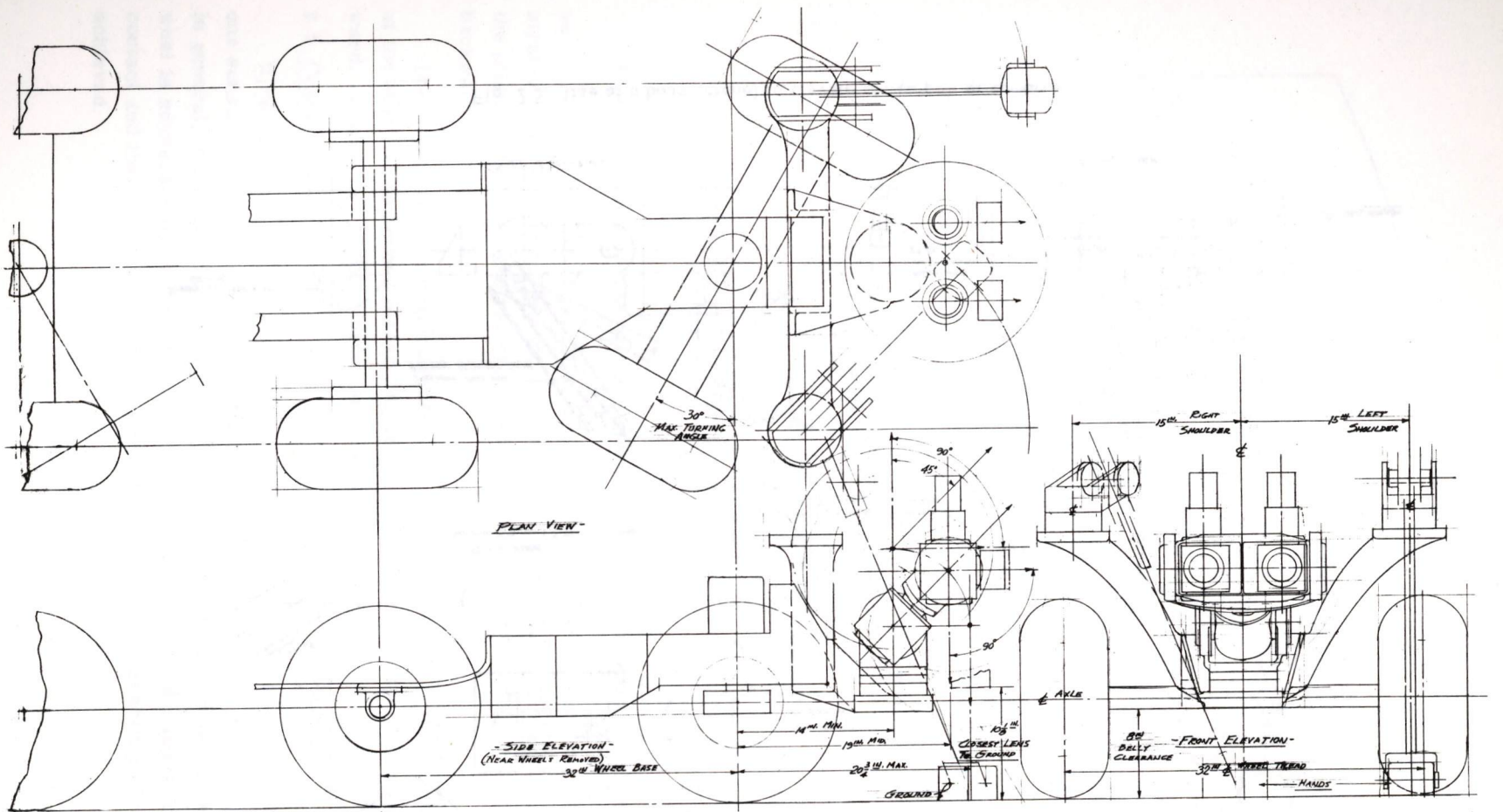


Fig. 2-6. Camera assembly E mounted in front of forward axle of General Motors six-wheeled vehicle.

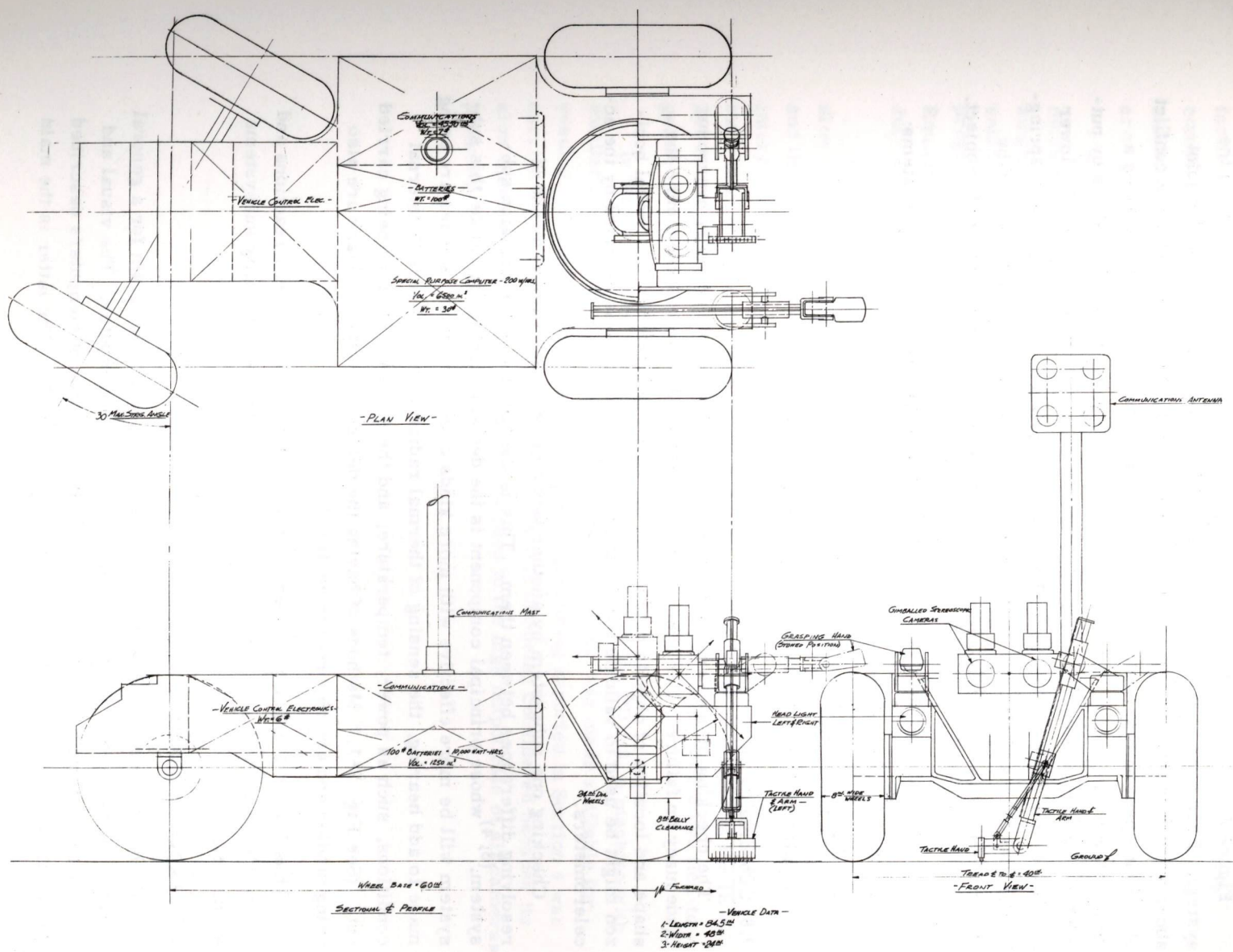


Fig. 2-7. Stereo pair of cameras mounted on front axle.

Figure 2-6 is the first attempt to include a hand-and-arm assembly in the system. Each hand is at the end of an arm which is supported through a joint above a front wheel. Two objections to this arrangement are the possible conflict of a front wheel with an arm and the mechanical difficulties of cantilevering a long arm with precision. The design of Fig. 2-7 lessens these difficulties by putting the wagon-type steering at the rear as in a ship, and the shoulders both lower and nearer the center. The hand, as presently designed, contains a line of spring-loaded pins with a shaft position encoder on each. From the line array can be determined the relative displacement of the pins and hence the contour of an object.

The following elements feed data to or receive data from other subsystems:

- a) Tactile input of the hands.
- b) Kinesthetic input of each joint.
- c) Motors which position the hand for touching.

2.6 Coordination of the Subsystems

The checking of one sensory subsystem by others should partially overcome inadequacy or failure. For example, the contact subsystem can reexamine the shape and location of objects detected by the visual subsystem. The visual horizon might be used to validate the inclination of the cameras as reported by the accelerometers.

Checking one subsystem by another would be of no use without means of resolving differences between them. This is the function of the decision subsystem,^(3, 4) whose principal component is the decision computer. Since this subsystem will be more effective with more kinds of sensory input, studies are being made to add hearing, the sensing of thermal radiation and sensors of internal conditions, such as power, temperature, and the progress of tasks being carried out. (See Fig. 2-1.) Methods of having the decision subsystem learn are also being studied and are reported in Section 6.

Thus Fig. 2-1 might be curled on itself so that the emergency reactions and tasks, shown at the right, relate to the visual, contact and mobility subsystems at the left.

2.7 Digital Computers

Each of the boxes in Fig. 2-1 is being developed as a program for a general purpose digital computer, although not for the same computer. The visual and contact subsystems are being developed in the PDP line of computers described below, the decision subsystem in the IBM 360 Model 75. The latter is the main

facility at the Instrumentation Laboratory and is provided with a very facile compiler - the M. I. T. Algebraic Compiler. However, pressure to use this machine is so great that attachment of external equipment such as cameras and arm-and-hands has not been permitted.

It is the intention of this effort to obtain first a combination of commercial machines in which all programs can be run and all equipment attached, then a single machine of equal facility but small and light enough to be carried on a vehicle such as that of Fig. 2-7. The combination of machines in which all programs can be run and all equipment attached is expected to be a Digital Equipment Corp. PDP-9 to be delivered to the Instrumentation Laboratory in July of this year, and a larger computer to be connected to it by telephone lines.

As presently planned, the larger computer would contain both the relatively slow formation of the model of the environment and its interrogation (Section 5) and the decision subsystem (Section 6). The PDP-9 would perform the faster interpretation of the output of the cameras (Section 4) and the commanding of their motion. The decision subsystem will continue to be developed in the IBM 360 Model 75 while a copy of it is programmed for the larger of the two tied-together computers.

In the interval between ordering and delivery of the PDP-9, a machine with similar instruction code, the PDP-7 in Project MAC, was made available for evening and off-hour daytime use. The programming reported in Section 4 was carried out in this machine. Finally, TX-0 is not being considered because its slow memory access rate is poorly matched to the high output rate of all cameras presently under development.

SECTION 3

CAMERA-COMPUTER CHAINS

by

Louis Sutro, David Tweed, Joseph Convers and Charles Sigwart

3.1 Objectives

While the primary task of the visual subsystem is to sense the environment, that is, to obtain pictures of the area surrounding the cameras, due to the time lag in communication and the low rate of transmission between the robot and Earth, it is also necessary that the robot construct a model of the environment, that is, an internal representation that can be interrogated to help in guidance and experimentation. As visual data flows from the cameras through successive transformations to the internal representation, it can be tapped at any point and transmitted to Earth. Thus, several choices will exist regarding not only what pictures are to be sent but also what will be their level of abstraction.

A prerequisite for dealing with visual data is a means of determining range, such as obtaining two views from two vantage points. Engineers call such an arrangement stereoscopic, while psychologists argue that stereoscopic (stereo) vision is not only range finding on one point at a time but determination of the relative range of every point in the entire field for which there is a disparity between the left and right views.⁽⁵⁾ Camera-computer chains for the purpose of finding the range of one point at a time are considered later in this section.

Beginning with subsection 3.2, the subjects considered are cameras A and B, both monocular. Then follow descriptions of three of the four types of stereo optics being considered for the visual subsystem, namely, the C, D and E. (A fourth type developed by Philco-Ford has been described elsewhere.^(6, 7)) This section then shows how the stereoscopic camera assembly and its electronics can be tied to a computer to interpret its output and direct it. An analysis of the errors inherent in each camera type is presented.

Since descriptions of hardware and software can be conveniently separated, this section will dwell on camera-computer chain A, operated in the summer of 1966, and chains C, D and E presently under development; while discussion of the programming to be employed in these chains is reserved for Section 4. A possible method of modelling the environment is presented in Section 5.

Nonetheless, the subject material of the three sections has been conceived as a single unit. For example, the window, employed for detailed examination of the image, is formed by counters that open or close gates at a window frame superimposed either on the scene or on the stored image. The counters will be either special-purpose electronics (hardware) or a routine within a general-purpose computer (software).

All of the cameras described in this section employ as their transducer the slow-scan vidicon (Type 1343-018) developed by the General Electrodynamics Corp. of Garland, Texas, for the Mariner and Surveyor spacecraft. It is being used to obtain the data reported here and may be replaced by solid-state transducer arrays as these become available.

3.2 Camera-Computer Chain A

Camera chain A was monocular; that is, it employed a single optical train. Built in 1964 and improved in the spring of 1966⁽⁸⁾ it was linked to the TX-0 computer in the summer of 1966 for experimentation in the detection of motion. (See Fig. 3-1.) The camera had been designed at Lincoln Laboratory, MIT, to be operated as fast as its output could be loaded on an analog tape (500,000 Hz). After loading, the tape was to be run at a slower speed, so that its contents could be converted from analog to digital form and stored on a second tape for computer processing.

In its use for the Instrumentation Laboratory, the camera was required to respond to four times as large a raster at the same resolution. Given the same dwell time per position in the raster, this would require that the video amplifier have four times as wide a passband as was required for the Lincoln Laboratory application.

In the actual test, due to the inavailability of high-speed analog tape, the camera was operated at the speed of digital tape, with the result that signals passed through the video amplifier at frequencies below the passband. Lowering

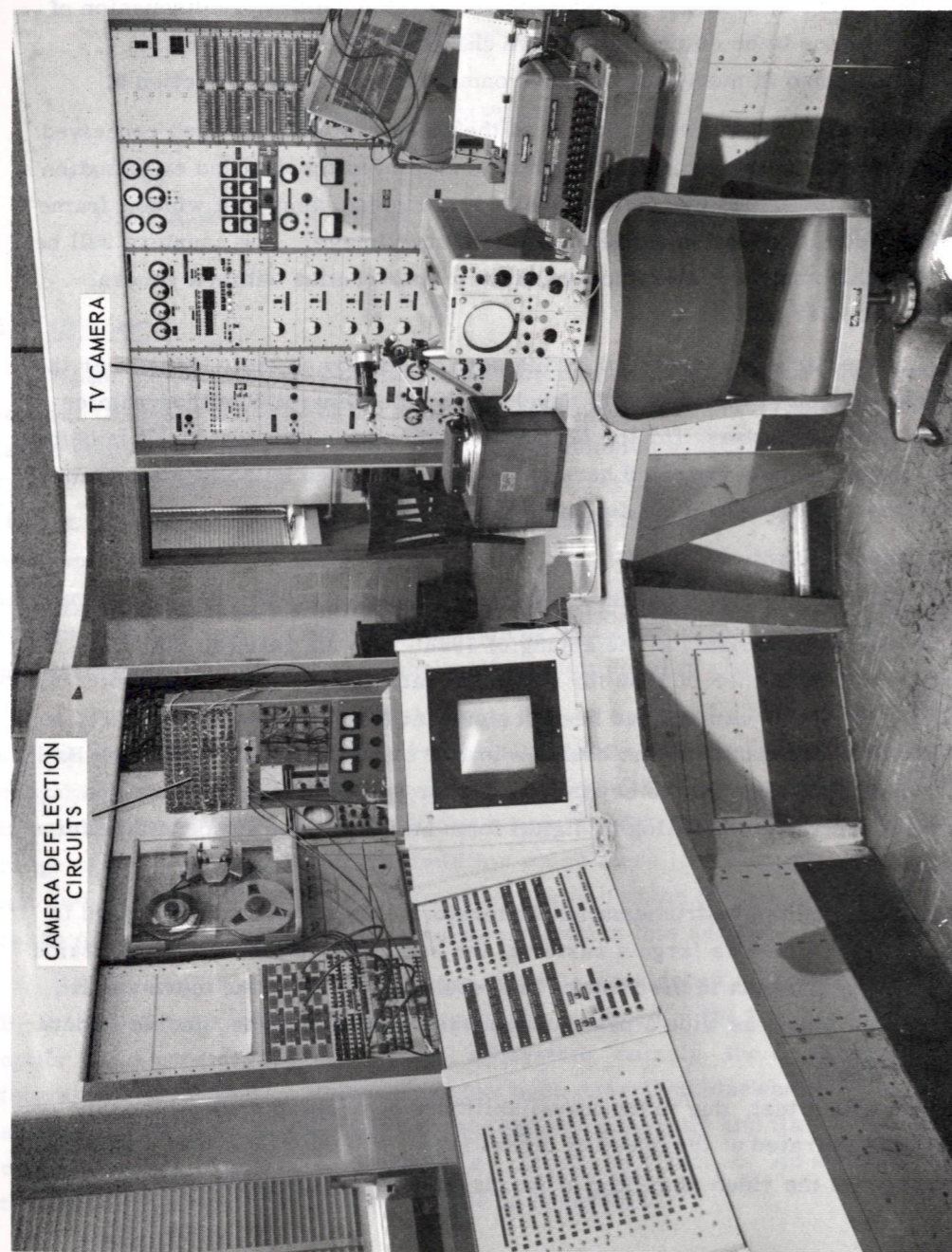


Fig. 3-1. Camera-computer chain A. The computer is the TX-0 in the Department of Electrical Engineering, MIT.

the bandpass through partial modification of the video amplifier could not prevent noise from appearing in the data, and construction of the amplifier on printed cards within a tubular case made further modification difficult. Since Camera C was already in development, further work on Camera A was dropped.

In spite of the difficulties camera-computer chain A was operated by a graduate student leading to a thesis on motion detection.⁽⁹⁾ The study consisted of recording in the TX-0 computer pairs of television frames, comparing corresponding positions in them and generating a "decision map" of changes in luminance. To lessen the effect of noise the computer was programmed to "reduce" the decision map by rejecting isolated changes.

Further experiments in motion detection have been deferred for the present. It is assumed that the method employed in camera-computer chain A could be employed also in succeeding designs.

3.3 Camera-Computer Chain B

Chain B was an assembly to implement motion detection in hardware. Un-developed beyond the sketching stage, it was to consist of two camera tubes, one of fast response, the other slow, both examining the same scene through a beam splitter. The outputs of the two tubes were to be compared to detect motion.

3.4 Designation of Stereo Optics and Camera Electronics

While camera computer chain A was being assembled, design began on stereo optics and continued on camera electronics. It became convenient to employ a different designation for each of these lines of development. Thus, the basic type of the stereo optics is represented by a letter, e.g., C, D, E; a variant of the basic type by a number and a letter, e.g., C1, C2, and the electronics by a second number separated from the letter and first number by a dash. (See Table 3-1.)

Type C stereo optics is an arrangement of fixed mirrors to bring two separated optical paths onto a single vidicon. Type C1 is an attachment for a Bolex 16 mm moving-picture camera that can also be attached to a television camera. As shown in Fig. 3-3, the axes of its two optical paths are parallel. A possible variant is a configuration with fixed convergent axes.

Stereo optics types D and E are characterized by both variable convergence and ability to mechanically scan a field of view larger than that of the lenses. For reasons given in subsection 3.8, it was decided to provide two cameras in each of these types and turn and tilt them to achieve the variable convergence.

Table 3-1. Components of camera-computer chains.

ITEM NO.	COMPONENTS	CAMERA-COMPUTER CHAIN			
		A	C1-1	D1-3	E1-3
1	Camera for A	X			
2	TX-0 Computer	X			
3	Opto-mechanical configuration Type C1		X		
4	Camera electronics Type 1		X		
5	PDP-9 Computer		X	X	X
6	Opto-mechanical configuration Type D1			X	
7	Camera electronics Type 3			X	X
8	Three gimbals, their torquers and resolvers				X
9	Six gimbals, their torquers and resolvers				X

and mechanical scanning. Type E stereo optics provides each camera with both a wide-angle coarse-resolution field and one of narrow-angle and high-resolution. In the Type E each camera is capable of the motions provided for the D, plus pitch, roll and yaw motions to erect the pair of cameras from a tilted base.

3.5 Type 1 Camera Electronics

Experience with Camera A led to the design of the Type 1 camera electronics which is planned to be used with the Type C1 stereo optics. The resulting camera chain is called C1-1. Type 2 camera electronics is a commercially available system, on order. Type 3 is the electronics of the two cameras of the D configuration, assumed here to have the characteristics of the Type 1. For purposes of comparison, Type 3 electronics is assumed to be employed also with camera assembly E1, although another type may be employed yielding twice the aperture response (resolution).

When the field of view is imaged on the face of the vidicon the brightness or luminance of each small angle in this view causes an appropriate change in the resistance of the photoconductive surface that modulates the video output waveform. The purposes of the electronics involved, as indicated in Fig. 3-2, are:

- To deflect the electron beam of the camera tube horizontally 512 positions and vertically 512 lines so as to describe a raster. Rate of advance of the beam should be a multiple of the computer clock rate.
- To amplify the video signal and convert its amplitude at each raster position into a digital word to be entered into the computer.

While the electron beam does not actually dwell at each position, it is useful to think of it as doing so.

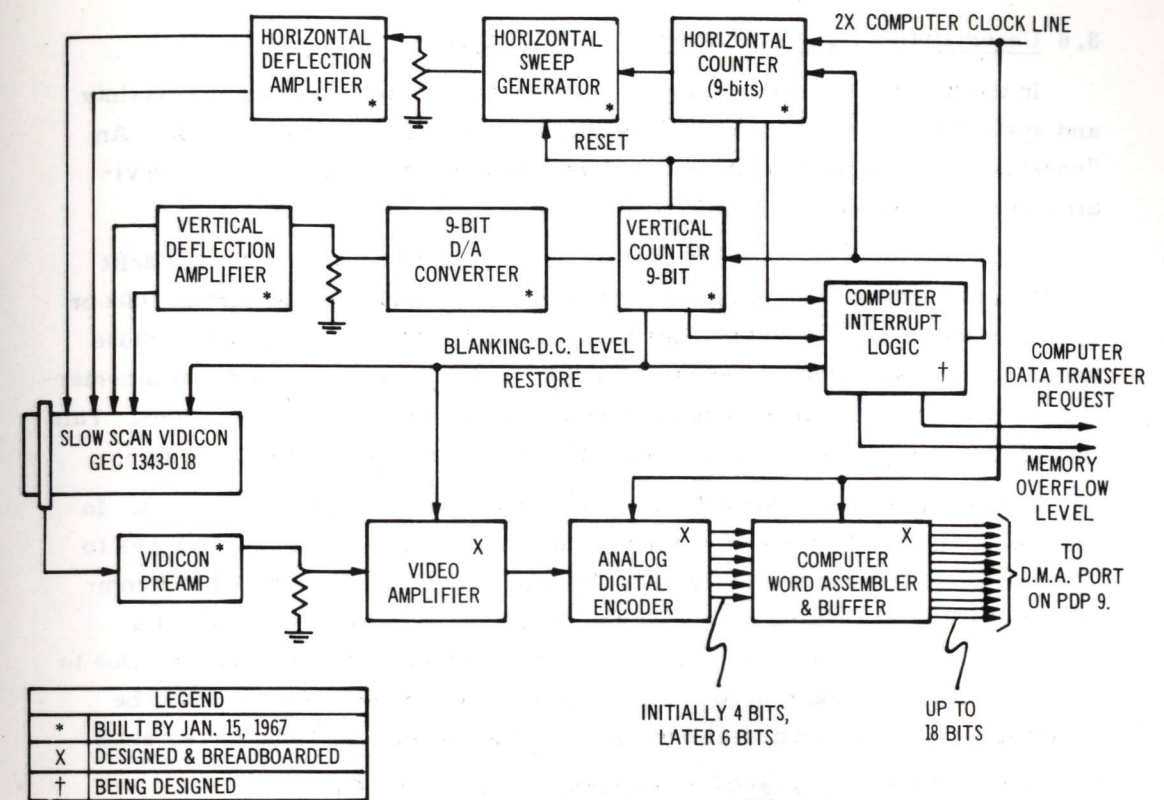


Fig. 3-2. Block diagram of Type 1 camera electronics

Deflection of the beam along one line of the raster is achieved by generating a ramp waveform in the horizontal sweep generator shown at the top center of Fig. 3-2. The beginning and end of the ramp are determined by the horizontal counter which counts from 0 to 511 at twice the computer clock rate. This multiple is chosen because two six-bit words can be fed into the computer memory in one clock time. (The remaining six bits are needed for flags, etc.). The computer clock rate is 1 MHz and the television camera frame rate is approximately 7 frames per sec.

The video signal is amplified and converted to digital words also at twice computer clock rate, i. e., 2MHz. While the sampling theorem suggested an amplifier bandpass of 1.0 MHz, it was possible to design it with a bandwidth greater than 1.5 MHz and a rapid rolloff after 2.0 MHz. Sampling distortion (aliasing) results but is moderate because the finite vidicon beam size causes the amplitude of the components of the video signal to roll off rapidly at high frequencies. While slightly reducing the S/N ratio of the amplifier, this extension of the video bandpass lessens distortion of these high frequency components.

3.6 Uncertainties Due to the Type 1 Camera Electronics

In the discussion that follows three terms will be used, error, uncertainty and systematic error. An "error" is a deviation from an expected value. An "uncertainty" is a random error. A "systematic error" is a consistent deviation from an expected value.⁽¹⁰⁾

Uncertainty in the centering of the electron beam is due to D-C bias drift which in the present design is estimated to be a maximum of one part in 1000 or 0.03 v in the bias of 30 volts. This deviation is 1/1000th of the raster, whose useful width, as reported in subsection 3.8, is 300 lines. Thus there is a centering uncertainty of 0.3 line along both the horizontal and vertical directions. This uncertainty can be reduced to one half this amount or less by careful design.

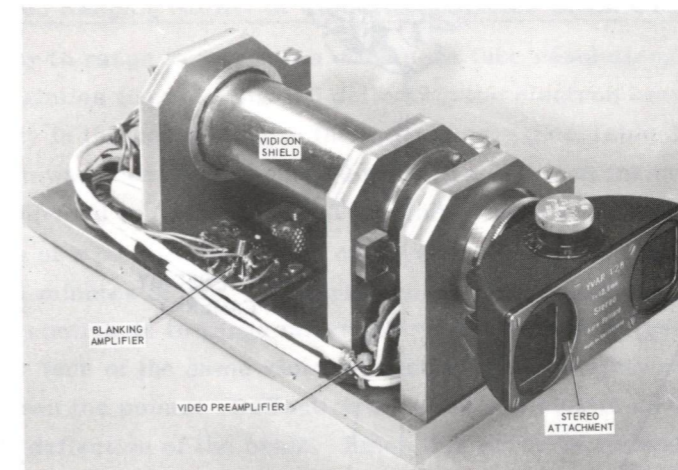
Systematic error in deflection is a function of the amount of deflection. In the present design this error is zero at the center of the raster, but it rises to 1/2% or 1-1/2 lines at the periphery. This error could be measured and compensated for in computing range. It can furthermore be reduced by about a factor of five through the use of high-gain feedback amplifier techniques. Due to lack of uniform response across the surface of the vidicon there may also be uncertainty in the indication of luminance if many levels are to be interpreted.

3.7 Optics of Camera-and-Stereo-Attachment Assembly C1

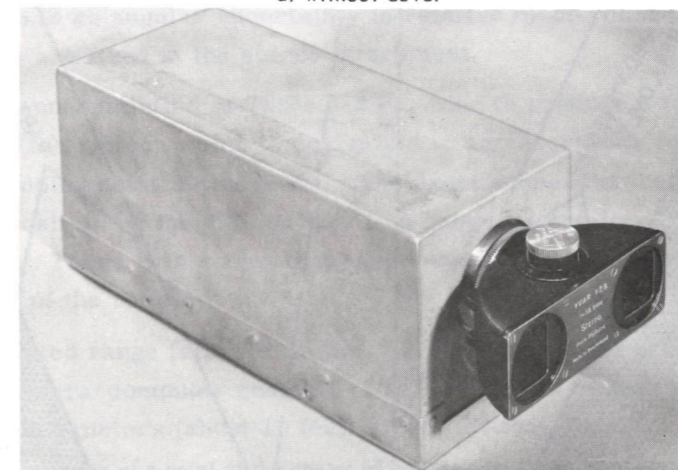
In this and succeeding subsections the word "assembly" is used to describe optics, camera case and supporting mechanisms, such as gimbals. The word "assembly" thus refers to the optical and mechanical component of a camera-computer chain, but not to its electronics.

Camera assembly C was designed to have the simplest available stereo optics.

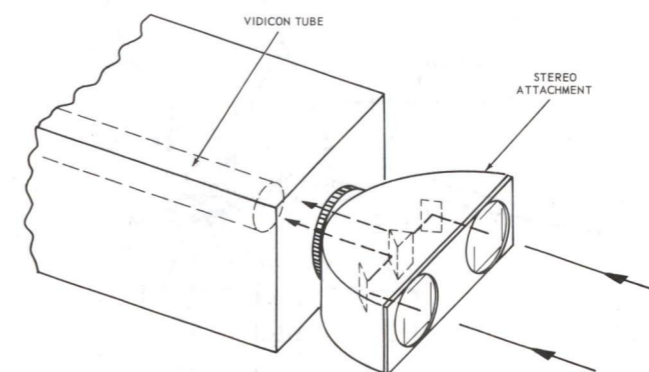
Since a vidicon camera may employ lenses intended for 16-mm movie cameras, it was possible to use the stereo attachment designed for a Bolex camera shown in Fig. 3-3. The stereo attachment receives two views of a scene, 63-mm (2-1/2 in.) apart, and forms images of the views side by side on the face of the vidicon. In considering the performance of this assembly it is convenient to straighten the two light paths and show the two views imaged as though on separate cameras as in Fig. 3-4. Each lens receives light from a rectangular pyramid in space subtending a minimum linear angle of 28°. The volume of space contained in the intersection of the pyramids is the scene that is viewed stereoscopically.



a) without cover



b) with cover



c) diagrammed to show reflecting surfaces

Fig. 3-3. Stereo camera assembly C1.

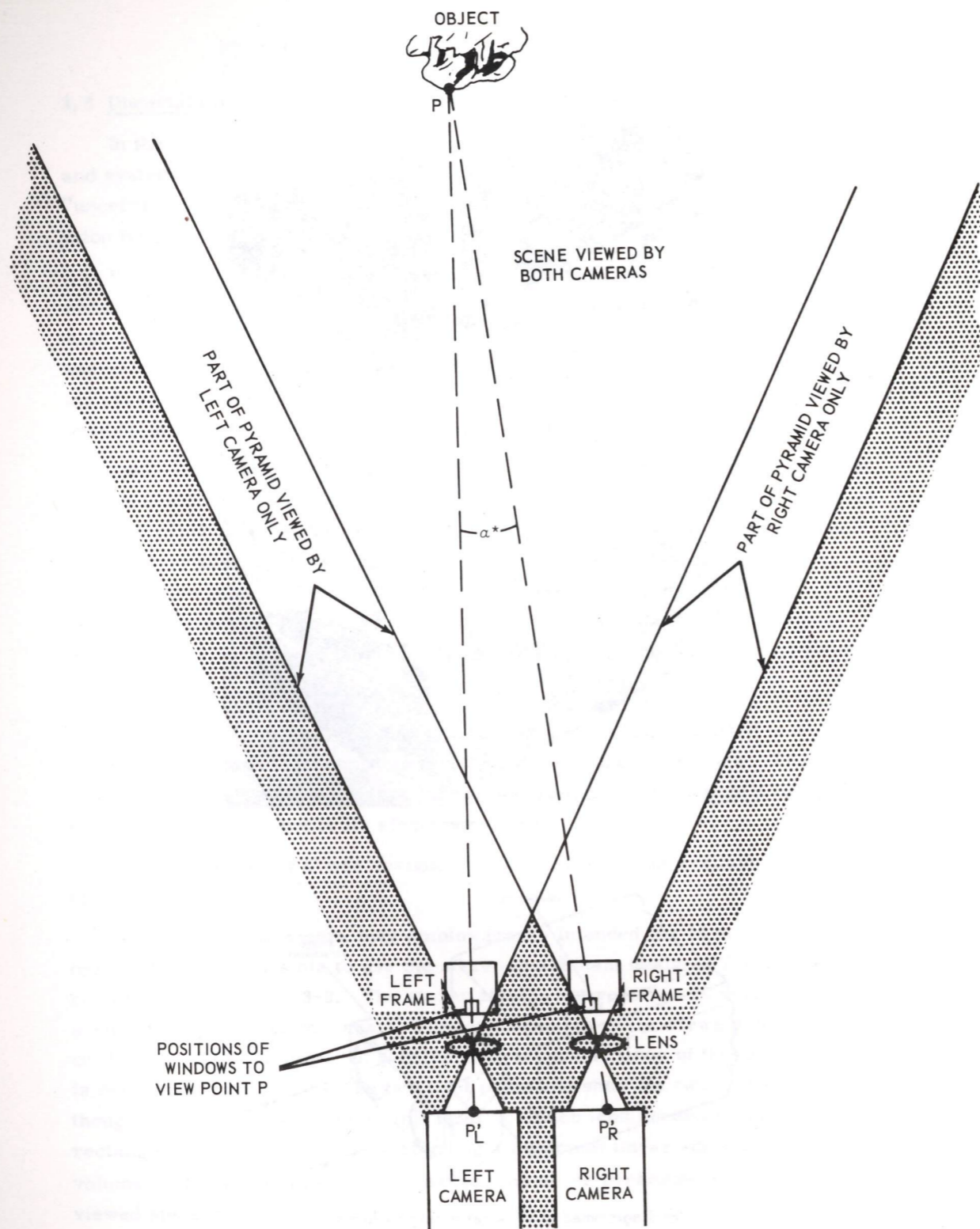


Fig. 3-4. Wide-angle parallel lenses viewing the same object.

3.8 Anticipated Ranging Ability of Camera-Computer Chain C1-1

The ability to range is a function of camera tube resolution, geometry of the optics, uncertainties in centering and deflecting the electron beam in the vidicon, and uncertainty in the positioning of the optic axes. (See Table 3-2.) Camera tube resolution was measured to be 300 lines in a 1/2 inch frame at 10% response factor, as described in Appendix A. This finite resolution produces an uncertainty in the position of a point in the image of ± 1.0 line which, in the geometry of the optics, is ± 24 minutes of arc. Centering uncertainty, described in subsection 3.6, does not contribute to range uncertainty. Since both images are on different portions of the face of the same vidicon, centering uncertainty cannot change the distance between the points. Deflection error is systematic and proportional to the amount of deflection of the beam. Since this error is consistent, measurable and can be compensated for in computing range or corrected by redesign, it is not listed. There is no angular uncertainty in relative mechanical positioning because the optic axes are fixed in the stereo attachment.

All of the uncertainties of Table 3-2 combine to produce a total uncertainty on the face of the vidicon which is equivalent to an uncertainty in angular position of a corresponding point in the scene. The exact expression relating the total uncertainty of position on the vidicon face to a range uncertainty is developed in Appendix B.3. There it is shown that range uncertainty increases approximately as the square of the range.

The indicated range falls within the curves of maximum uncertainty C-C in Fig. 3-5. Camera-computer chain C1-1 is expected to be tested in ranging on objects up to 3.0 meters (about 10 feet) away. At that distance

Table 3-2. Uncertainties of a point at the center of each image, that contribute to range uncertainty.*

	CAMERA-COMPUTER CHAIN							
	C1-1		D1-3		E1-3		E1-3	
	MIN. ARC	LINES	MIN. ARC	LINES	MIN. ARC	LINES	MIN. ARC	LINES
Uncertainty due to finite resolution of vidicon and geometry of optics	± 24	± 1.0	± 3.0	± 1.0	± 3.6	± 1.0	± 3.6	± 1.0
Uncertainty of centering of electron beam of the vidicon	0	0	± 1.8	± 0.6	± 2.2	± 0.6	± 0.22	± 0.6
Uncertainty in the positioning of the optic axes	0	0	± 1.0	± 0.34	± 1.0	± 0.28	± 1.0	± 2.75
Maximum total uncertainty	± 24	± 1.0	± 5.8	± 1.94	± 6.8	± 1.88	± 1.58	± 4.35
		min. lines		min. lines		min. lines		min. lines

*This table assumes systematic errors have been corrected

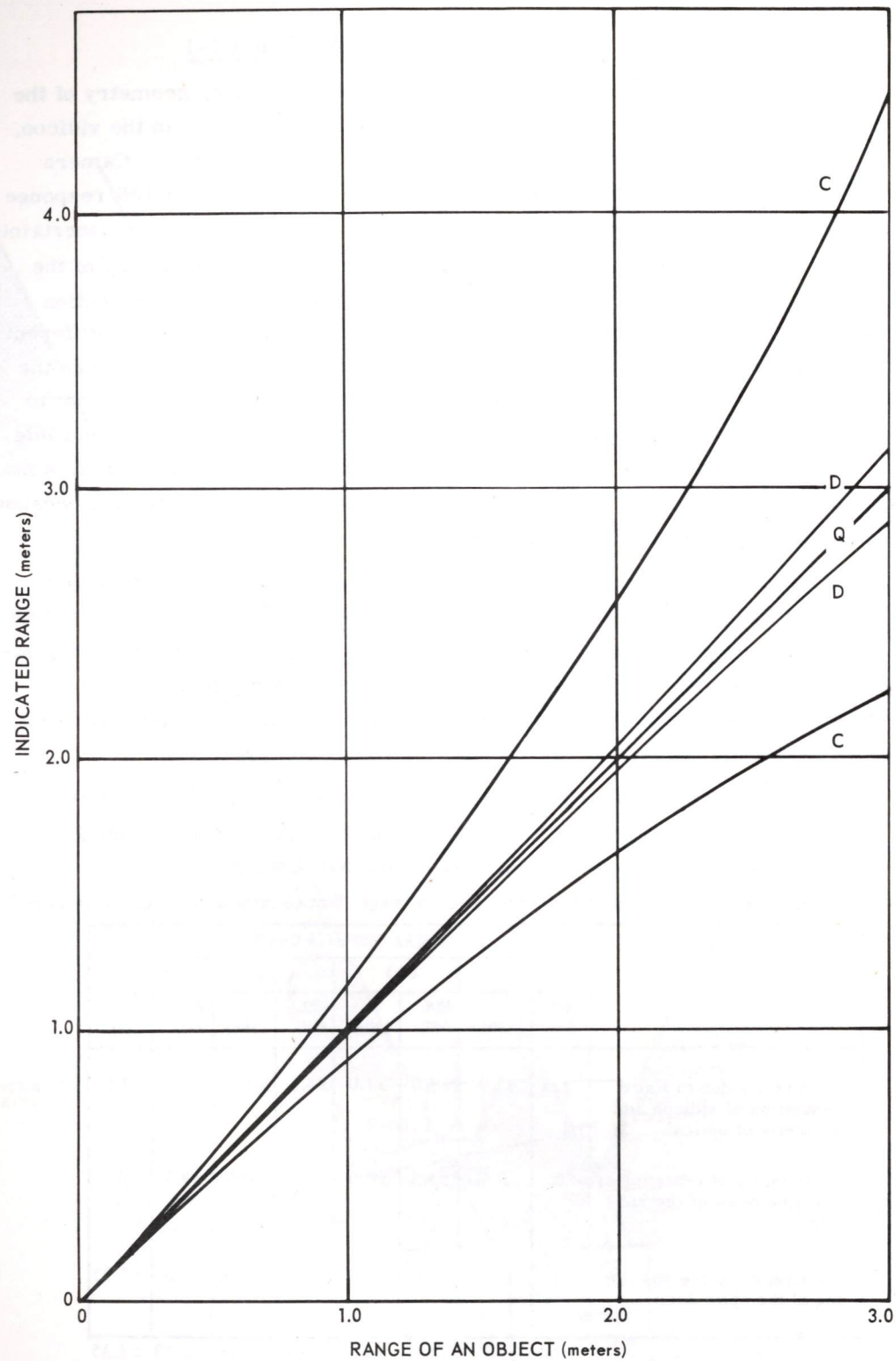


Fig. 3-5. Anticipated maximum uncertainties in camera-computer chains C1-1 and D1-3. (See Appendix B.2 and B.3 for computation of the points plotted here.)

the range uncertainty is about ± 1.0 meter. The range uncertainty may be reduced by increasing the separation of the optic axes, as is done in the D Assembly.

3.9 Camera-and-Gimbal Assembly D1

The D Assembly is a frame in which various combinations of transducers and lenses may be tested. The assembly shown in Fig. 3-6 provides for rotating each camera about its yaw axis and rotating both cameras as a unit about the pitch axis. The configuration chosen is appropriate for testing range finding with converging optic axes. The torquer-resolver platforms (gimbals) provided will make it possible to position each camera with an uncertainty of about 30 seconds of arc and position a pair with an uncertainty of about one minute of arc. (Other configurations which were considered are shown in Appendix C.)

An uncertainty of one minute of arc was also sought for each vidicon-lens assembly (camera).

$$\text{Vidicon resolvable line width} = \frac{0.5 \text{ in.}}{300 \text{ lines}} = 0.0017 \text{ in.}$$

$$\therefore \text{focal length, } f = \frac{\text{line width}}{\sin 1'} = 5.66 \text{ in.}$$

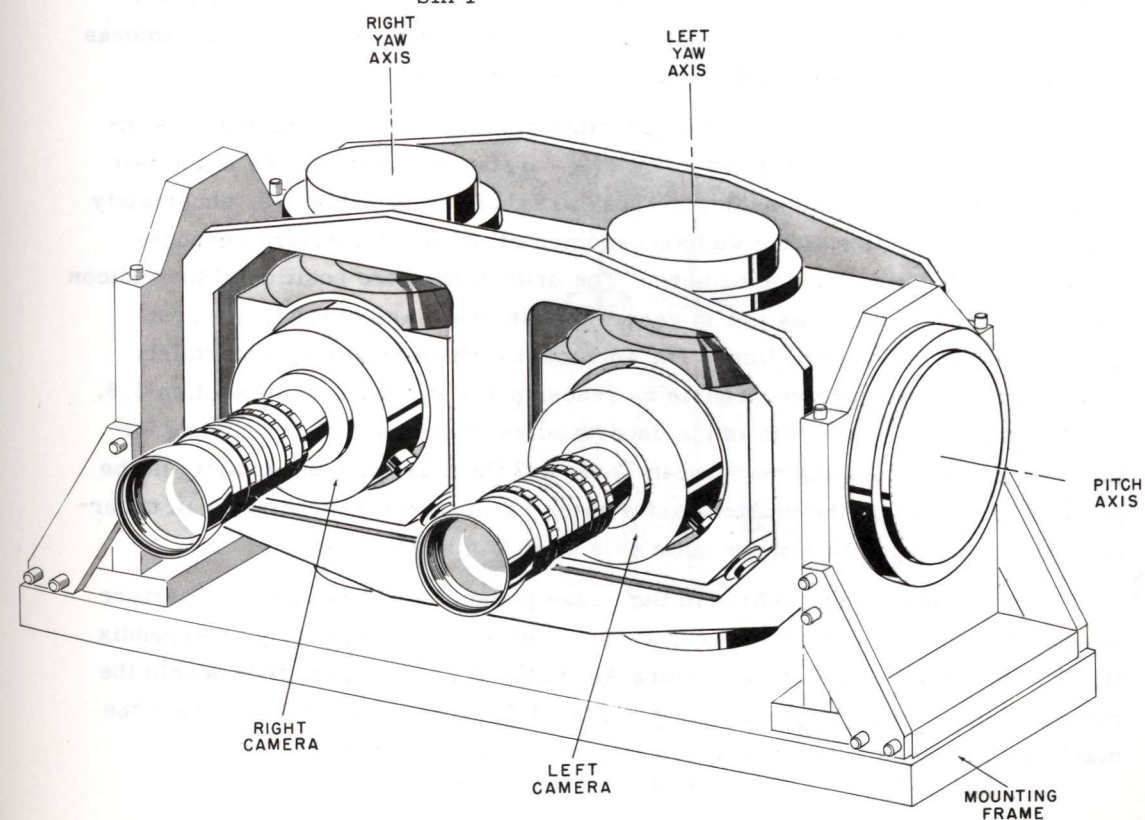


Fig. 3-6. Camera and gimbal assembly D1.

Such a lens is undesirable when considering a mechanical support of that length. The lens selected has a focal length of 3 inches and images a 5.67° square field on the raster of the vidicon. With this width of field and focal length the resolution limit is determined by the finite resolution of the vidicon.

The D1 Assembly is shown in Fig. 3-6 with 3-inch focal length lenses. The physical length of the lens is 3.9 inches and the overall length of the lens and vidicon assembly (camera) is 13.0 inches. Each camera is turned by a torque motor above it which measures 3 inches in diameter, while the position of the camera is sensed by a resolver below it measuring approximately 4 inches in diameter. The yaw axes of the cameras are separated by 6.250 inches. The cameras are mounted in a stereo box pitched by a torque motor, which is out of sight at the left. The pitch angle is sensed by the resolver at the right. Geometric relations in Camera-and-Gimbal Assembly D1 are shown in Fig. 3-7.

Motion of each camera is in increments of 1° , an amount being reduced by redesign.

3.10 Anticipated Ranging Ability of Camera-Computer Chain D1-3

Table 3-2 summarizes the uncertainties in ranging by Camera-Computer Chains C1-1 and D1-3. For the latter it is assumed that the axes of the cameras converge so that $\alpha_1 = \alpha_2$ (see Fig. B-1, Appendix B).

In determining the uncertainties in range finding with two cameras, the intrinsic uncertainties are doubled since in a worst case uncertainties in the two cameras do not cancel each other. Thus, as tabulated in Table 3-2, uncertainty due to finite resolution of the vidicon becomes ± 1.0 line of aperture response rather than ± 0.5 line for one camera. The drift of the zero position of the vidicon rasters of the two cameras are in general not in phase and result in a worst case error of ± 0.6 vidicon line. The amounts by which centering uncertainty and deflection error are reducible by redesign are indicated in subsection 3.6. In addition, there is the sum of the uncertainties of mechanical positioning for the two cameras. As summarized in Table 3-2 the maximum uncertainty in the position of a point on the vidicon raster is ± 1.94 lines corresponding to an uncertainty in angular position of the optic axis of ± 5.8 minutes of arc.

This uncertainty of position in the image plane of about ± 2 lines of aperture response results in uncertainties in range as developed analytically in Appendix B. The indicated range using Camera Assembly D as described falls within the curves of maximum uncertainty D-D in Fig. 3-5. At a range of 3.0 meters the maximum uncertainty is about ± 0.13 meters (about ± 5 inches).

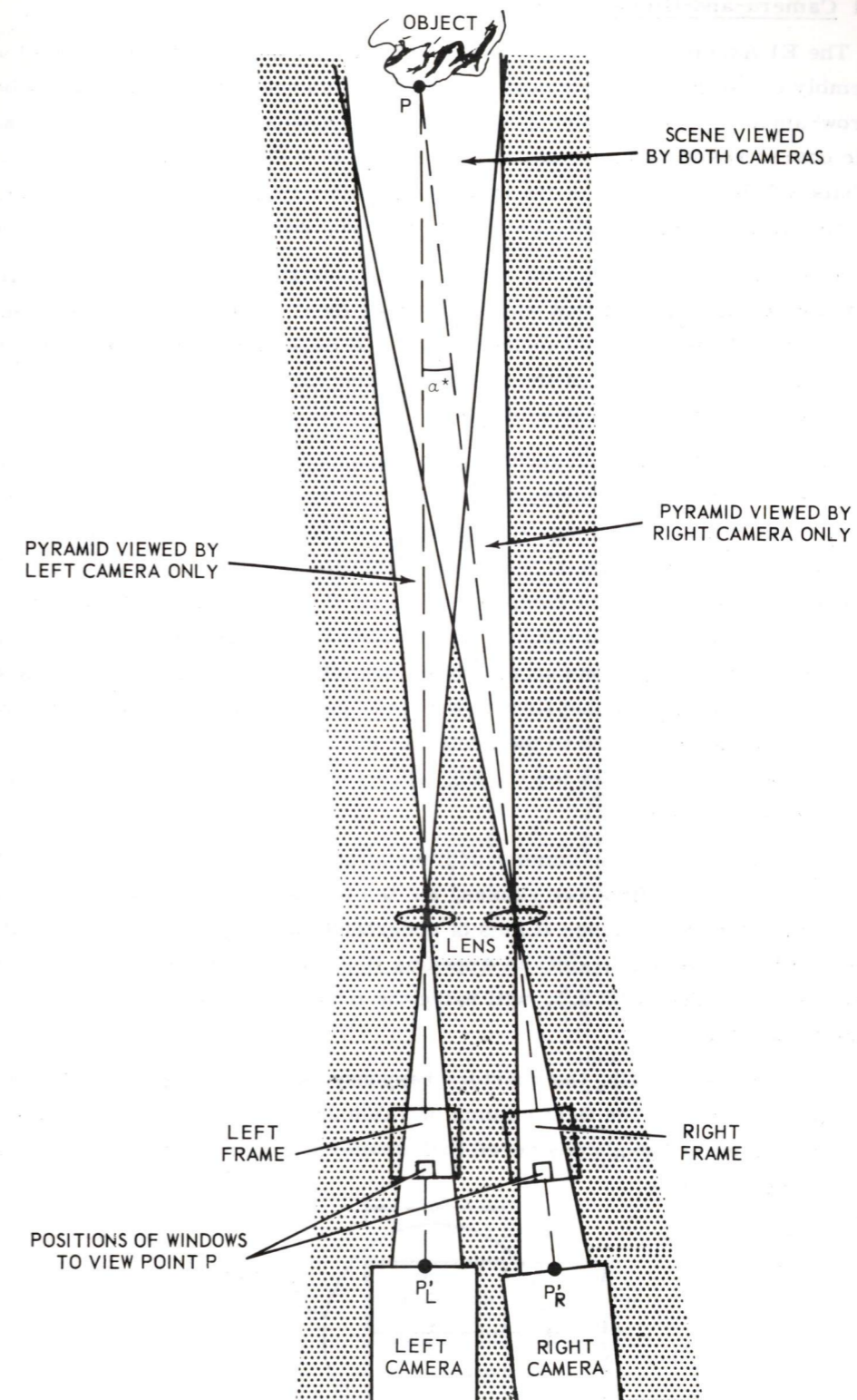


Fig. 3-7. Narrow-angle cameras converging on the same object.

3.11 Camera-and-Gimbal-Assembly E1

The E1 Assembly is being developed to determine the size and shape of a stereo assembly composed of two cameras, each of which receives on one vidicon both the narrow-angle high-resolution field shown at the bottom of Fig. 3-8 and the wider-angle coarse-resolution field shown at the top. It is to be carried by a vehicle. Its gimbals will level the axis of the stereo box each time the robot stops to look, at the same time permitting the cameras to look up, down, sideways, or backwards.

In addition to providing both wide- and narrow-fields of view, the optical train of camera E1 is folded so that while its longer focal length is 400 mm (16 inches) its front-to-back length is 6-1/2 inches. This will be achieved by mounting the vidicon at a right angle to the axis of the entrance to the optical train. (See Figs. 3-9 and 3-10.) Details of the optical train are given in Appendix D.

The wide-field optical train of the E1 Assembly will provide range finding performance with more uncertainty than that of the D1 Assembly while the narrow field optical train will provide range finding performance with less uncertainty. Figure 3-11 shows the bounds of maximum uncertainty for the three optical trains, D-D representing that of camera computer chain D1-3, E_1-E_1 and E_2-E_2 representing those of the wide-angle and narrow-angle trains respectively in camera-computer chain E1. The causes of the uncertainties are tabulated in Table 3-2. Note that the angular uncertainties due first to the finite resolution of the vidicon and geometry of the optics and second to the centering of the electron beam are ten times as great for the E1 train as for the E2, while the uncertainties in positioning the optic axes are the same.

On the other hand, the number of lines of uncertainty are the same in both optical trains for the first two causes, but ten times as great in the second optical train for the third cause. (Lines of uncertainty are represented by s in the equations for range uncertainty in Appendix B. 2.) Doubling the aperture response, as proposed in subsection 3.5, will halve the first two causes of uncertainty but not affect the third.

The effect of doubling the aperture response is most striking in the C configuration reported in the first column of Table 3-2. Accordingly, a possible C2 stereo optics is being considered that would be an improvement over the C1 in other ways as well.

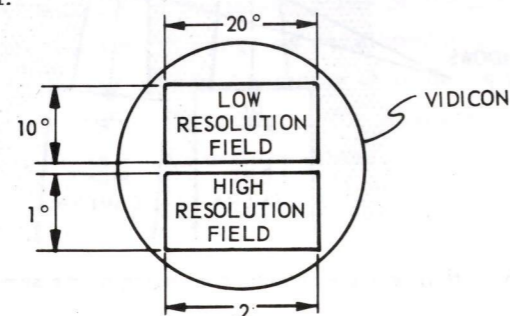


Fig. 3-8. Diagram of composite field of view at vidicon of an E1 camera.

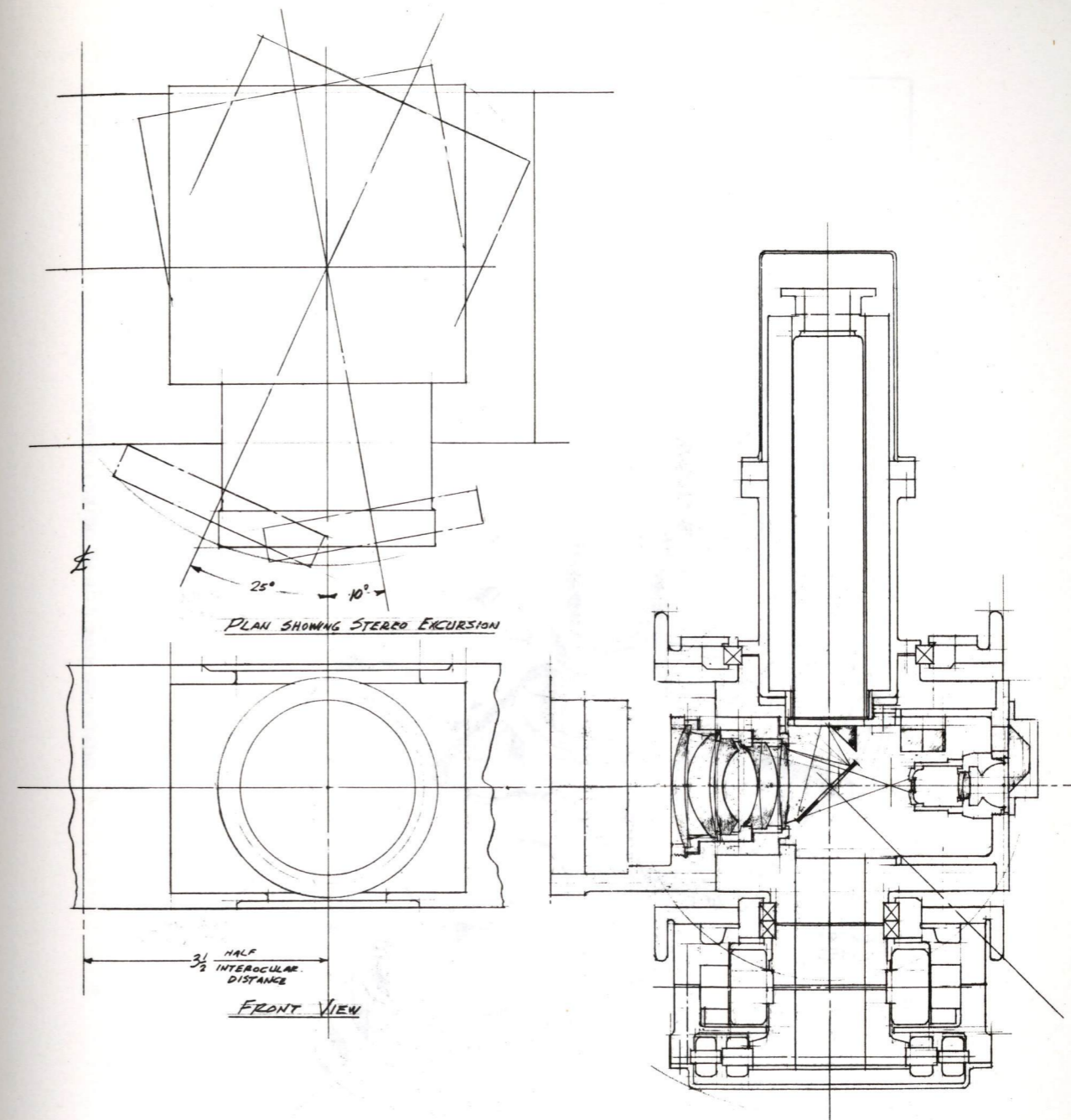


Fig. 3-9. Camera E1. Not shown are the shutter, means of focussing the lens and means of installing and changing filters.

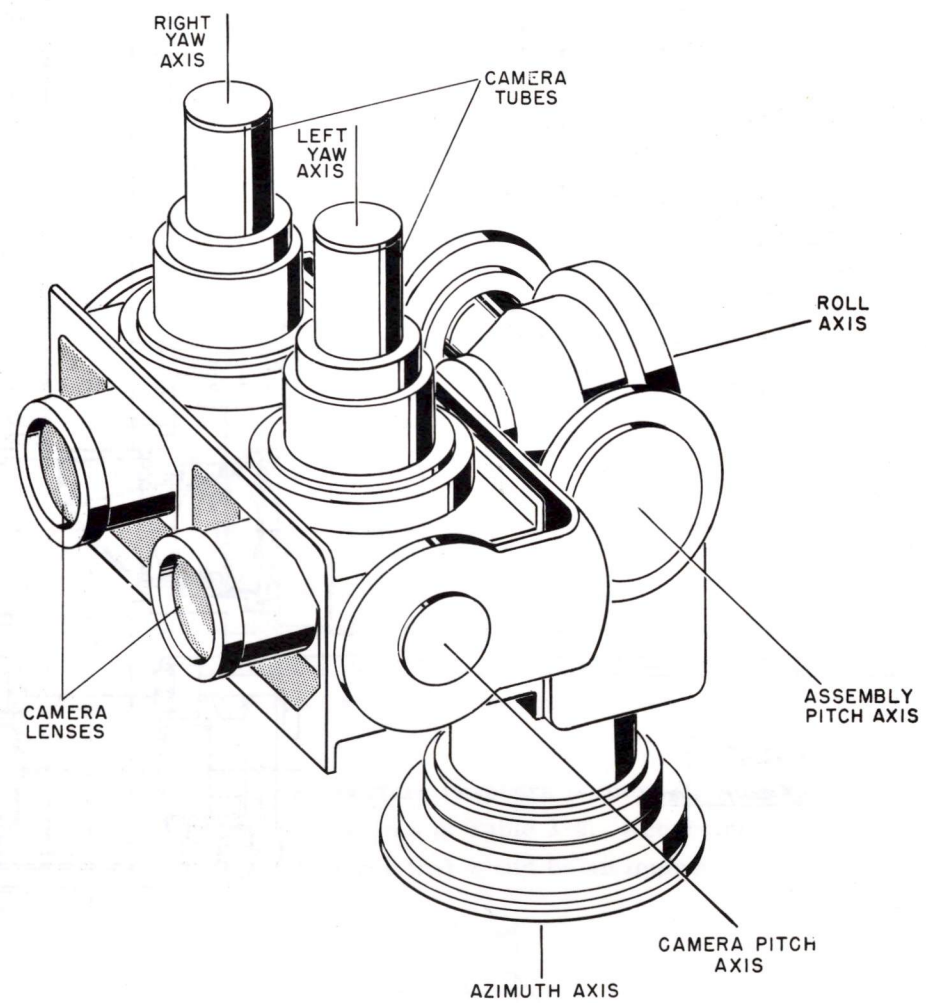


Fig. 3-10. Camera-and-gimbal assembly E1.

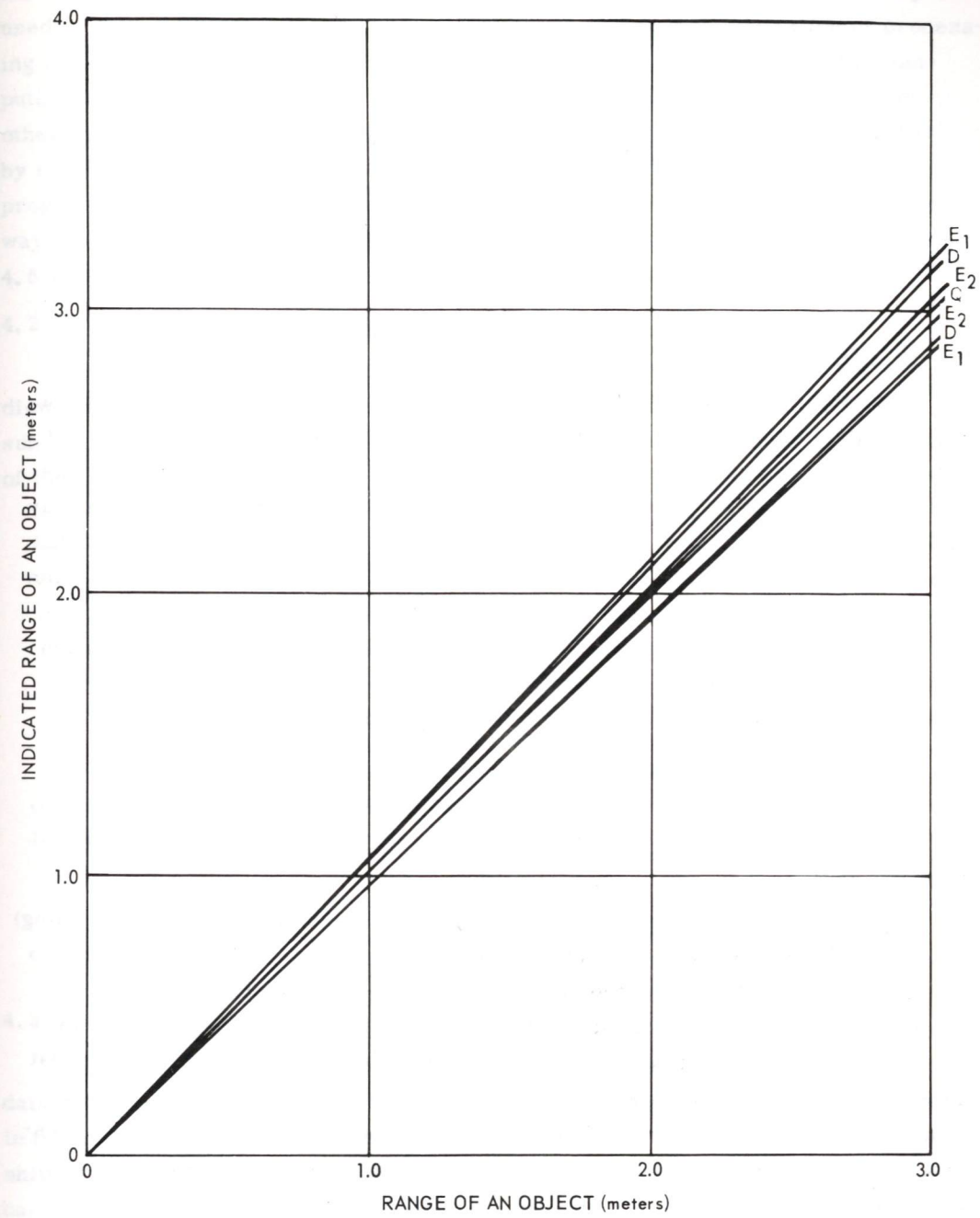


Fig. 3-11. Maximum uncertainties in indicated range for camera-computer chains D1-3 and E1-3. The curves D-D are the bounds of maximum uncertainty about the actual range q_0 . E_1-E_1 and E_2-E_2 are the corresponding bounds for the low and high resolution fields of the E1-3 chain respectively. (See Appendix B.2 for computation of the points plotted here.)

SECTION 4

VISUAL DATA PROCESSING

by

Roberto Moreno-Diaz and James Bever

4.1 General

This section is concerned with methods of processing visual data in such a way that significant features can be enhanced and abstracted. In particular, it considers processing of visual data for the generation of a "point model" of the environment. What is meant by a point model can be expressed in a simple way as follows: imagine television records of the scene in front of the camera, that have been divided into small areas or portions, each containing many resolution elements. The position of each small area is given by a pair of numbers (x, y). By means of a range finding process, a third number could be assigned to each small area indicating how far away the part of scene there is imaged. The point model consists of the set of "points" (x, y, z) describing the scene.

The point model represents the scene by giving a "replica" of it without much abstraction. The realization of an internal model in a more abstract way, so that storage can be further reduced, is the aim of such investigations as that reported in Section 5.

The ranging process is based on finding pairs of homologous (corresponding) small areas – or windows – in a stereoscopic pair of views. To decide that two windows – one in the left and the other in the right – are homologous, the window in one view is kept fixed while being compared to a variety of them in the other view, until matching is obtained. The matching process requires that the edges appearing in both windows are well defined. In general, the initial visual data is not so, and some initial processing to enhance the edges is necessary. For this reason, we began by considering different processings of the data to enhance contrast.

A conventional technique to detect – or enhance – the edge in a picture is that of differentiation in two directions.^(11, 12) This method is, however, very susceptible to noise. By considering how animal retinae process the data (see Section 8) a more generalized method of contrast enhancement was arrived at,

that of lateral inhibition. Actually, similar procedures have been extensively used in optical filtering techniques.^(13, 14, 15) A natural generalization of processing of the type indicated in subsection 8.2 has resulted in the writing of a computer program able to handle not only contrast enhancement but also perform other kinds of "filtering" of the data – the word filtering having been suggested by similar techniques in optical processing and range finding. The general program is described in the following subsections, along with an account of the way it was arrived at. The method of finding homologous windows in subsection 4.5 was the first attempt to solve the range finding problem.

4.2 Characteristics of the Visual Processing Program

Visual processing must include such manipulations of data within a two-dimensional array as those performed by the set of routines associated with a sub-array called "windo." As illustrated by Fig. 4-1, the routines are capable of the following:

1. The "windo" array may be varied in size from 1×1 to $18 \times N$, to (some multiple of $18) \times N$ for special applications.
2. a. "Windo" can be loaded and unloaded from four directions. Hence, "windo" can be "moved" through a larger array;
b. Also, "windo" can be modified selectively.
3. A variety of (a) irregular and (b) regularly-shaped regions can be defined within "windo" which can be processed directly or mapped onto an array called "windo-within".
4. Arbitrary positions within "windo" can be processed together.
5. "Windo" can function as (a) a one- or (b) two-dimensional shift register. In particular, one "windo" can be superimposed upon and shifted across another.

4.3 Initial Approach

Our initial approach to processing visual data was to reduce a small array of data to some sort of line, representing a summary of grey level contours present in the array. Grey level contours, frequently termed "edges", occur as sudden shifts in grey level, extending typically along the boundary between a shape and a background of contrasting luminance. It should be noted that specks of noise also cause such level shifts.

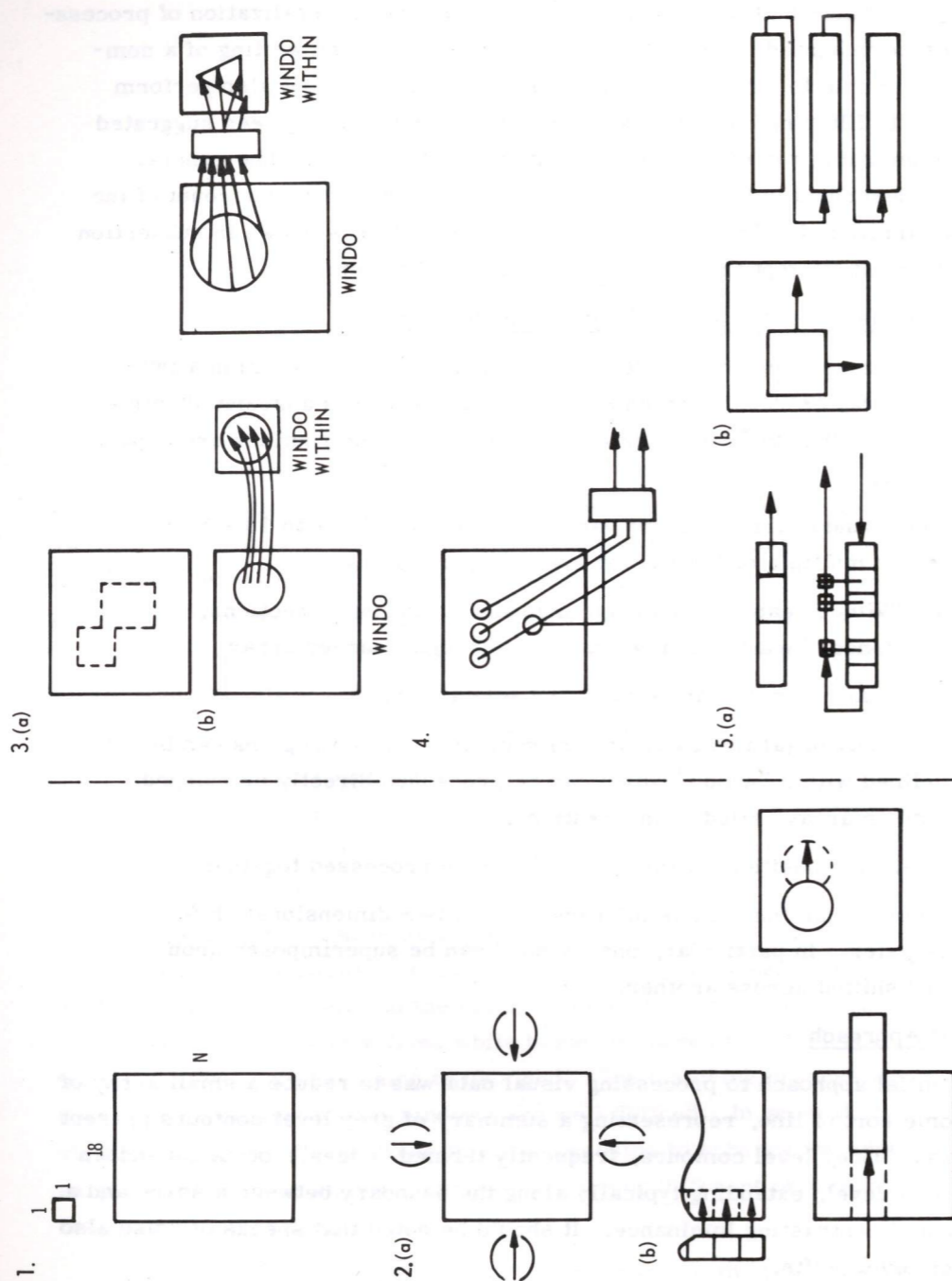


Fig. 4-1. Some properties of the information-processing space.

A computer program which reduces data to the line form first loads "windo" with data, then sweeps across this data searching for abrupt shifts in grey level. The program marks such a shift by placing a cross, consisting of five binary one's, in an array called "windo-within." A chain of binary-coded crosses may form a contour which will extend unbroken between two sides of "windo-within." A subsequent process reduces contours to straight lines, and counts them as components of a "vector."

A further process endeavors, by means of comparing addresses of data from which matching left and right "eye" pointers* have been computed, to establish the range of the data.

The line reduction scheme above cannot be used to interpret all the data configurations possible in "windo." It is especially ineffective where high resolution increases the complexity of the data.

4.4 A Second Approach

The line approach suggested that a form of correlation would be useful. A method of correlation was then developed whose first step is to differentiate the data, such that the presence of an edge is represented by a 1, the absence by 0. Afterwards, a frame is transformed into a string of raster lines taken in succession. Correlation consists of holding say, the string of the left frame fixed while passing the string of the right frame along it, one (bit) position at a time; at each position, corresponding elements are "anded" together; thereafter, a count of non-zero results is made and entered in a "correlation function." The latter is a block of registers whose length is the same as the number of bits in the longer string.

A "correlation function" can be searched as a list to determine locations of peaks and other features. In the case of a simple form of range-finding, the location in the "correlation function" of an absolute maximum determines, in principle, the range of a feature in the field scanned by the pair of cameras.

While such a correlation method has been found useful, difficulties, such as multiple peaking of the correlation function, arise when the strings are formed, since it destroys the two-dimensional character of the input. Also the presence of an absolute maximum value of the correlation function does not guarantee similarity between two forms being compared.

Input to the correlation process has been provided by a program which first differentiates data then recodes as binary the result of differentiation which consists of edges present in the data. Because of the sensitivity of this process, the results are often crowded with trivia, and do not characterize simply the original data.

* See subsection 2.3.

4.5 The Present Approach

Experience obtained in modelling animal visual systems^(16, 17) suggested the use of integral processing. Routines have been written called "Latin", short for "lateral inhibition", which is a particular case of the lateral interaction of afferent fibers along pathways leading from animal sensors.

"Latin" routines substitute, for a piece of data, a weighted average of data over a region surrounding it. Weighting functions which are positive value in the center and negative in the surround smooth noise and yield results consisting mainly of contours and other specifiable features. "Latin" is now being tested in the processing of photographs and will be tried in the processing of television images when camera-computer chains C and D go into operation.

4.6 Range Finding Planned for Camera-Computer Chains C and D

To find a pair of homologous windows in the left and right views, a window is first fixed in, say, the left view, according to some criterion such that it be the area of highest contrast or of highest luminance; then a "likely area" is defined in the other view (see Fig. 4-2). Since this area contains as many potential windows as horizontal resolution elements, the problem is to determine which of these corresponds to the one fixed in the other view. The simplest way is to compare, point by point, the fixed window to all those contained in the area of suspicion, generate some error function and use this as a criterion to decide when the matching is best attained. A point by point comparison is very susceptible to shifts of the position of the points and to changes in the gain of the video amplifiers. To lessen these difficulties, we may substitute the value of the output of the camera at a point by an average over a few positions around the one being compared, or, in other words, smooth the initial data from the cameras. This smoothing is similar to defocusing the original image.

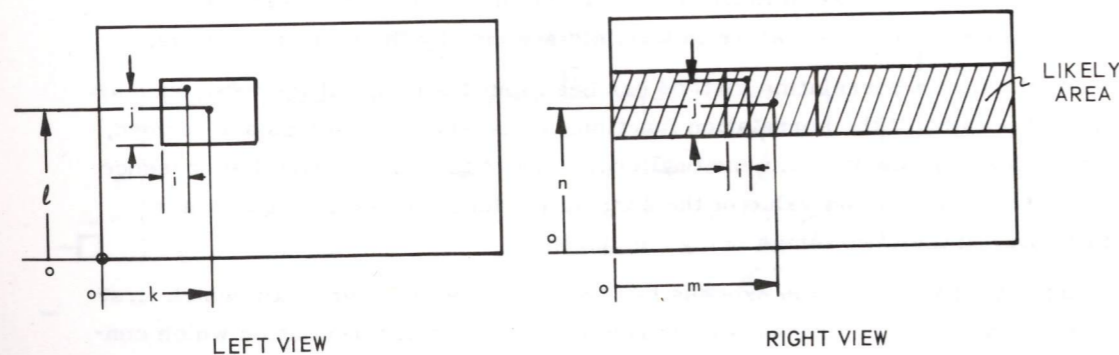


Fig. 4-2. Finding a pair of homologous windows.

Let us represent by $f_L(i, j, k, \ell)$ the value of the output of the camera at the point (i, j) of the window (k, ℓ) . (See left view of Fig. 4-2.) Let $f_R(i, j, m, n)$ be the value of the camera output at the point (i, j) of the window (m, n) in the right view. By smoothing, we obtain similar distributions $\Phi_L(i, j; k, \ell)$ and $\Phi_R(i, j; m, n)$. The problem is then to find the coordinates (m, n) of the window in the right view that correspond to the window (k, ℓ) in the left view. If the cameras are horizontal, $\ell = n$, and the width of the likely area equals the width of a window. For every window (m, ℓ) in the likely area, we compute the absolute value of the difference,

$$|\Phi_L(i, j; k, \ell) - \Phi_R(i, j; m, \ell)|$$

and integrate it over the size of the window, that is, we form the "error":

$$\epsilon(k, \ell; m, \ell) = \frac{1}{M \times N} \sum_{i=1}^{i=M} \sum_{j=1}^{j=N} |\Phi_L(i, j; k, \ell) - \Phi_R(i, j; m, \ell)|$$

where $M \times N$ are the dimensions of the windows (M and N are odd) and (17×17) are lines used for experimentation.*

Since, to start, k and ℓ are fixed, $(k, \ell; m, \ell)$ is computed for all values of m , that is, from $\frac{m-1}{2}$ to $F - \frac{m-1}{2}$, where F is the number of resolution elements in the horizontal dimension of the parts of the frames viewing overlapping areas. The errors so computed are stored in a list which is searched to find the absolute minimum value of error. To the absolute minimum error there corresponds a pair of addresses (k, ℓ) , (m, ℓ) of windows in the left and right view respectively. These windows are regarded as homologous.

For example, assume that the frames overlap horizontally in 200 resolution elements, that the window size is 17×17 , and that a left window has been fixed at $k = 30$, $\ell = 50$. For the likely area, $n = \ell = 50$, and the list of errors is formed by computing

$$\epsilon(30, 50; m, 50) = \frac{1}{17 \times 17} \sum_{i=1}^{i=17} \sum_{j=1}^{j=17} \Phi_L(i, j; 30, 50) - \Phi_R(i, j; m, 50)$$

for values of m between $m = \frac{17-1}{2} = 8$, and $m = 200 - 8 = 192$, that is, 184 values. Assume that their absolute minimum value is $\epsilon(30, 50; 35, 50)$. Windows in position $(30, 50)$ and $(35, 50)$ are considered homologous. The parallax is $35 - 30 = 5$ elements which, since the geometry of the camera is known, can be converted into distance (see Section 3).

*This method of generating errors is actually equivalent to template pattern recognition, where the template is provided by the fixed window.

Present studies indicate the need of further preprocessing the digitized video data before a point-by-point comparison is done to find range. In the approach described here, the preprocessing is only an averaging. When a complicated scene is viewed - for example, containing many edges - enhancement of the edges and reduction of the number of them prior to the range finding is necessary.

A variety of ways of achieving visual data reduction are being investigated with the computer program described in subsection 4.2. It is hoped that experimentation will lead to a family of weighting functions applicable to different types of scenes. This study will lead to specification of the characteristics of the computer in the camera-computer chain.

SECTION 5

TOWARD A METHOD OF MODELLING AN ENVIRONMENT

by

Louis Sutro

5.1 General

Since identification of life and geological forms is ultimately performed on Earth, the task of a robot in a planetary mission is to acquire pictorial data and reduce it both as a basis for decision there and for transmission to Earth. The reduced data must be reconstructed on Earth into a recognizable display.

To reduce the pictorial data we are here concerned first with the creation of a "point model" of the scene, then the determination of properties that are intrinsic, that is, invariant, say, to changes in illumination. Such a property is the reflectance of each triangular area formed by a group of three neighboring points. The point model would enable a robot to select a path to follow and, transmitted to Earth, provide a three-dimensional map of the terrain. The point-and-reflectance model, by outlining areas of different reflectance, should permit discriminating surface patterns.

The point model is considered in subsections 5.2 through 5.6, the point-and-reflectance model in subsections 5.7 through 5.10. The final subsection 5.11 considers the Earth station needed to display these models.

5.2 Proposed Method of Acquiring a Point Model

A necessary first step is to aim the cameras at an object of approximately known position, such as a flashing cross on the forefinger of one hand. The forefinger will then be moved until it touches an edge or other feature that can be identified. The feature will have to be found in two views by the method of subsection 4.6 since the cameras cannot be aimed at the point of the finger with great accuracy.

A way of finding a second point, described in subsection 2.3, is to move the windows along the edge used to identify the first point. Another way is to place the left window over any area with a sharp discontinuity in luminance and then search for the homologous point with the right window, while keeping the first point where it can be recovered.

When three points have been acquired, they can be stored in the computer memory as points of a triangle. Using two of these points as the base of a second triangle, a new third point can be acquired, and so on. As the number of points grows, the uncertainty of position also grows until it becomes desirable to aim again at a known point.

5.3 An Example

Assume that the robot has landed on Mars and has transmitted to Earth both a panoramic view of its environment and its approximate position as determined by two observations of the sun. The Earth operator has commanded the robot to proceed along azimuth 43° for 50 meters, if feasible, to map and describe the terrain as it proceeds, and to transmit this description together with sample television pictures.

Looking in azimuth 43° , the robot views a scene as that in Fig. 2.3. It needs to form a point model and from that determine a route to follow.

Let us assume that points are then determined on the surface of the three nearest mounds and on the ground among them. Figure 5-1 shows points on the surface of the mound at the left center of Fig. 2-3, drawn by the same artist who made that illustration. He, of course, knew where to place the points. Where the choice of points is made by the computer it is better to obtain more points so that the connections between them, if made, are not as significant. For example, triangle GEF is a shelf whose appearance is destroyed by connecting points E and J. For a point model, connections between points do not need to be made as explained in subsection 2.3. But for a point-and-reflectance model such lines serve as bounds of the planes for which reflectance is determined.

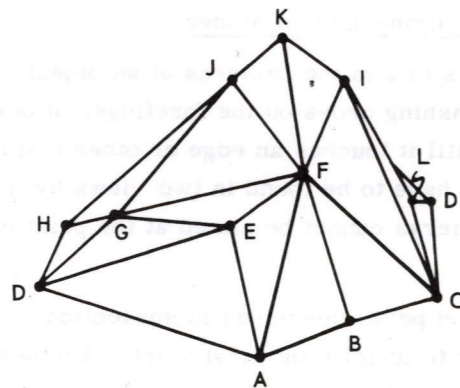


Fig. 5-1. Triangulated view of the mound in Fig. 2-3.

Table 5-1 presents the location of points F, G, and J in a coordinate system whose origin is in the robot, whose Y axis is a horizontal line through this origin in the direction the robot is headed, whose X axis is a horizontal line perpendicular to the Y axis and whose Z axis is perpendicular to the plane of the other two axes. For use in an extension of this example, the centroid of the triangle formed by these points is included.

Table 5-1. Coordinates of points F, G and J and their centroid.

	METERS		
	x	y	z
F	0.14	3.45	-0.09
G	-0.46	2.74	-0.25
J	-0.14	4.64	0.28
Centroid of ΔFGJ	-0.15	3.94	-0.02

The coordinates presented in Table 5-1 are computed from the data of Appendix E.1. They are located with an uncertainty given in the lower right corner of Table B. 1 of Appendix B.

5.4 Interrogation of a Point Model

To find a passage between mounds, the computer can form what a surveyor calls "profiles." A transverse profile will be defined as one in a plane perpendicular to the horizontal line of sight into the scene which will be the robot's Y axis if the robot is level. Figure 5-2 shows such a profile drawn on the imaginative illustration of Fig. 2-3. This profile does not have to be drawn; it can be stored as a set of points.

An algorithm to determine a profile can be developed as follows. A point p in the profile is $p(x,y,z)$ where y is constant. To find a point p on the profile for each value of x, a volume of the model such as that shown in Fig. 5-3 has to be searched, a task which may be facilitated by storage of the model in the form of a list indicating the adjoining point in each direction. Such a list can be searched for the highest point in each elementary volume shown in Fig. 5-3.

When the profile of Fig. 5-2 has been formed, it can be searched to determine whether it contains an expanse wide and level enough to permit the robot to cross. When such an expanse has been found, more transverse profiles can be "drawn" nearer and further away. The apertures found in each case can then be connected by longitudinal profiles, and these can be examined to determine if their pitch permits the robot to move ahead.

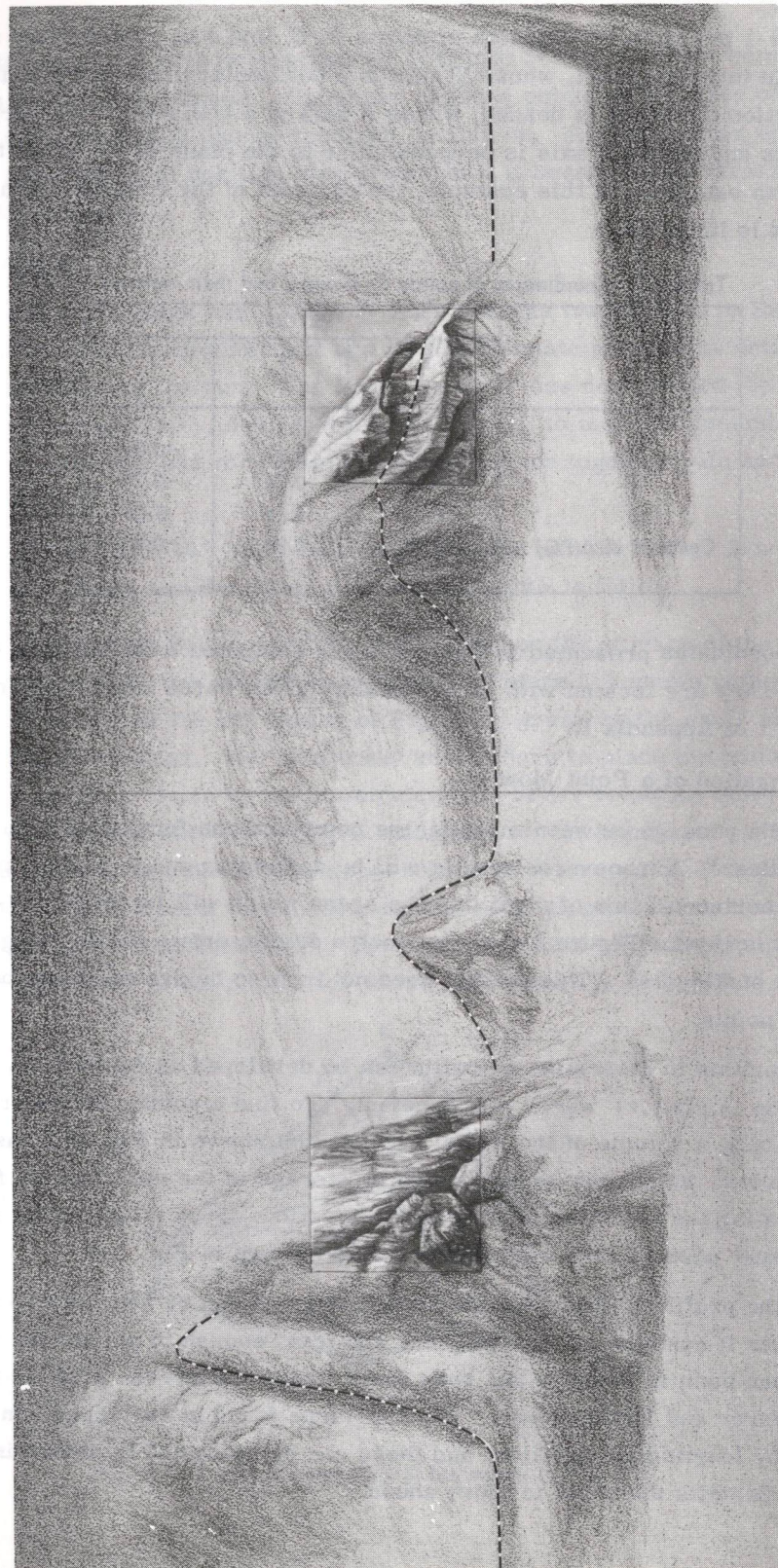


Fig. 5-2. Supposed views of the surface of Mars with a profile drawn in.

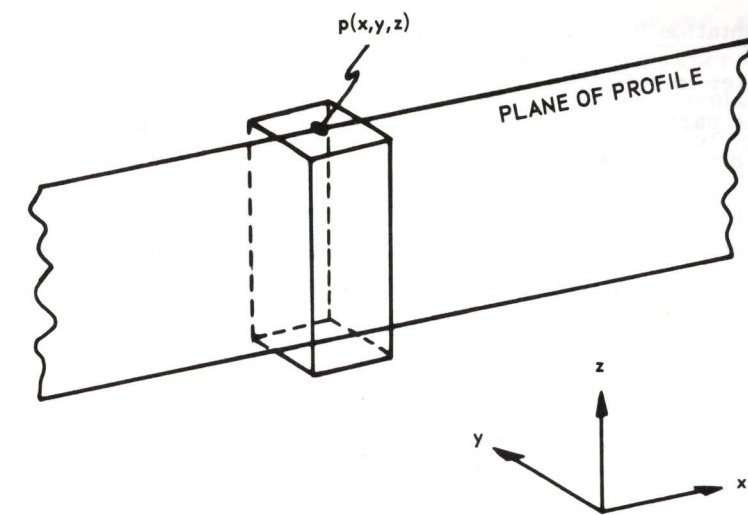


Fig. 5-3. Volume surrounding point p in the plane of a profile.

Another form of interrogation that can be performed on the model is generalizing from it both to locate the peaks, bases and approximate curvature of the mound as well as to conclude that except for the mounds all of the surface shown in Fig. 2-3 is an approximate plane.

5.5 Interrogation a Point Model in the Search for Evidence of Life

A profile may be "drawn" of a free-standing object, obtained in a manner similar to the making of a transverse profile of the scene. The profile can then be checked against a stored criterion such as: Is the object larger above than at its base? This is a characteristic of much of plant and animal life on Earth.

5.6 Mapping

Mapping of the surface of Mars can be carried out by the methods of a surveyor. From a starting station he runs a traverse of straight line segments, measuring the length of each and the angle between the forward and backward segments. The intersection of each pair of segments is called a "station" and is marked on the ground.

A Mars station can be defined as a place where the robot stops to look, and the distance between stations may be determined by an odometer. Enough stations should be marked to permit the robot to retrace its path. A suggested form of marking is a transponder which can be excited by a radio on the robot and homed in on.

5.7 Representation by Points and Reflectances

A next step in refining the model of the environment is to record a property of the surface particularly one that is invariant under changes in illumination. Such a property is reflectance.

The minimum area for which reflectance needs to be determined is a triangle formed by three neighboring points. The amount of computation required to determine the reflectance will, of course, be weighed against the data reduction achieved. The principles to be employed follow.

5.8 Computation of Reflectance^(18, 19, 20)

A television camera detects a mosaic of values of what used to be called brightness but is now called luminance, L . The term brightness is now limited to subjective response while luminance is used where the response can be measured. The unit of luminance in the United States is the footlambert. Except when it is the property of a light source, luminance is the product of illumination onto a surface and the intrinsic property of that surface called reflectance. Let us consider each of these factors separately, then together.

Illumination is a vector whose unit in the United States is the footcandle, whose representation is \bar{E} (for éclairage) and whose component normal to a surface is \bar{E}_N . The latter two measures are related by the Lambert cosine law:

$$E = E_N \cos \theta_i \quad (5-1)$$

where θ_i is the angle of incidence (See Fig. 5-4).

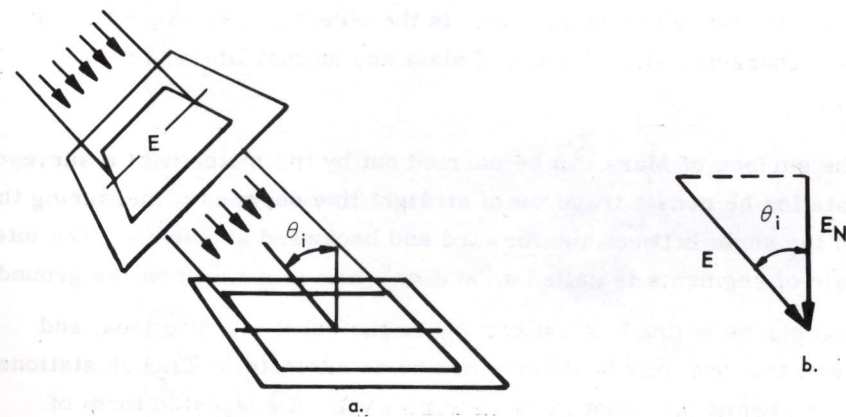


Fig. 5-4. a. The Lambert cosine law showing that light flux striking a surface at angles other than normal is distributed over a greater area.
b. Representation of illumination E as a vector and its normal component as a smaller vector of magnitude $E \cos \theta_i$.

Three forms of reflectance are identified: diffuse, specular and, spread. (See Fig. 5-5.) Diffuse reflectance ρ_D is that such as blotting paper; that is, it yields the same luminance from whatever angle it is viewed. Specular reflectance ρ_{SR} is that of a polished surface; that is, it reflects light at an angle from the normal equal to the angle of incidence. Spread reflectance ρ_{SD} is that of a roughly polished surface, reflecting so that the angle of reflectance only roughly equals the angle of incidence. Other forms of reflectance are combinations of these.

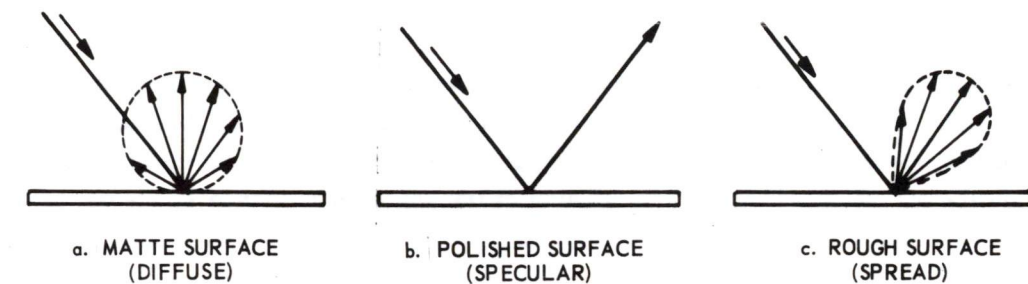


Fig. 5-5. The type of reflection varies with different surfaces:
(a) matte surface (diffuse or Lambert);
(b) polished surface (specular);
(c) rough surface (spread).

The product of normally incident illumination E_N onto a surface, expressed in footcandles, and the diffuse reflectance of the surface ρ_D is luminance expressed in footlamberts,

$$\rho_D E_N = L \quad (5-2)$$

Stated another way, the diffuse reflectance of the surface is the ratio of the luminance from it to the illumination upon it. Figure 5-6⁽³¹⁾ shows the use of this equation in determining the reflectance of a diffuse or non-glossy surface such as a painted wall. In Fig. 5-6(a), a General Electric type 213 light meter has been placed with its photosensitive surface against the wall and then drawn back two or three inches until no shadow falls on the wall. A typical reading might be 40 fL of luminance. In Fig. 5-6(b) the meter is held with its base against the wall. A typical reading in this position might be 65 fc of illumination. Substituting the two measurements into Eq. (5-2) yields

$$\rho_D = \frac{40 \text{ fL}}{65 \text{ fc}} = 60\%$$

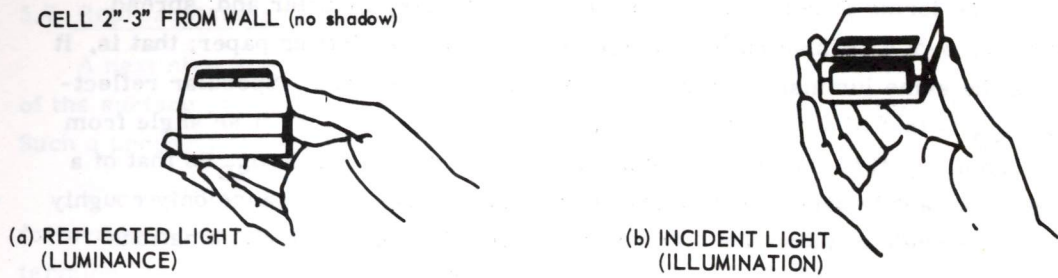


Fig. 5-6. Reflectance measurement with G.E. Type 213 light meter.

This is an approximate determination because the areas measured are large and the accuracy of this instrument is not high.*

A television camera can be used on Mars to determine diffuse reflectance in a similar manner. Aimed at the surface in question the camera measures luminance. Aimed at the sun with a suitable protecting filter it can measure illumination E . Multiplying E by the cosine of the angle of incidence θ_i yields the normally incident illumination E_N .

Figure 5-7 diagrams the measurement of diffuse reflectance by a television camera. The energy represented by the vectors E_N is reflected in an amount determined by the value of ρ_D and in a manner represented by the spherical clusters of vectors. The magnitude of each vector in a cluster bears a cosine, or Lambert, relation to the maximum vector in the cluster. It can be seen that if the camera is aimed at a steep angle it detects few long vectors, at a shallow angle many short vectors. The luminance detected in both cases is the same.

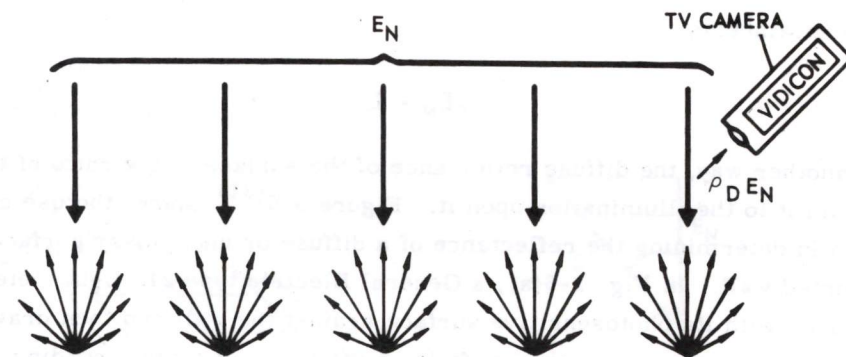


Fig. 5-7. Conversion of illumination to luminance when the surface has reflectance ρ_D is large and is uniformly illuminated.

*An instrument that measures luminance in a narrow angle of view at high sensitivity is the Spectra Brightness Photometer.

To measure specular and spread reflectance requires the ability to move around an object and measure vectors whose magnitude is different in different directions as in Fig. 5-5. When specular and spread reflectance are present luminance is determined by the polynomial:

$$L = \rho_D E_N + \rho_{SR} E + \rho_{SD} E \quad (5-3)$$

where

$$\rho_{SR} = \text{specular reflectance}$$

$$\rho_{SD} = \text{spread reflectance}$$

Note that the illumination E from the light source to the reflecting surface is used in the second and third terms. To measure specular reflectance it is necessary to move the camera to a point such that the light source, the measured surface, and the camera are in the same plane, with the angle equal to the angle of reflection. To measure spread reflectance, the same conditions must be followed and, in addition, the angle of reflection must be varied about the value equal to the angle of incidence to measure the "spread" of the light.

In most illuminating engineering work, luminance is computed from only the first term in the above equation. There may, however, be three such terms, one for each light primary. These are referred to as "tri-chromatic coefficients"⁽¹⁹⁾. Color measurement, which this implies, will be considered later.

The point model can be acquired under any illumination sufficient for edges, differences in shape, or differences in reflectance to be detected as contrast. The point-and-reflectance model, on the other hand, requires measurement of the illumination. To acquire the second kind of model, therefore, the robot must first determine the illumination upon the surface or, better yet, of the illumination upon the scene.

From its observations of the sun at two times in one day the robot can predict at any other time, the azimuth and declination of the sun and the resulting illumination normal to any surface of known position. Because there is almost no atmosphere on Mars, it is expected that there will be almost no sky light, except during dust storms. Then the lenses of the cameras will probably have to be capped to protect them from injury. The only sources of light are expected to be the sun and light reflected from the surface of Mars. In the example that follows only sunlight is considered.

The method presented here is similar to that of a painter who sketches a scene so that he can return to his studio to paint it. He observes the direction of the sunlight and, if he is acquainted with the effect of sunlight, keeps it in the

back of his mind as he concentrates on shapes and their reflecting properties. In doing this he is recording a scene invariant to illumination. Returning to his studio he can re-introduce sunlight as he paints the highlights, shades and shadows.

No painter that we know is aware of the mathematical methods of this task. Such methods are common in illuminating engineering, for example, where to achieve a desired appearance of a room, bulbs, wattages and reflectors as well as the reflectances of walls, ceiling and floors are considered.

5.9 Computation of the Diffuse Reflectance of ΔFGJ

The example of subsection 5.3 will now be extended by determining

1. The direction components of the normal to triangle FGJ , computed in Appendix E.2,
2. The average luminance L of this triangle, measured by the left camera, as 1300 footlamberts,
3. The illumination E from the sun, measured also by the left camera, as 4000 footcandles,
4. The direction components of sunlight, determined from the position of the sun stored in the computer memory,
5. The angle between the direction of sunlight and the normal to ΔFGJ computed in Appendix E.3 to be $\cos^{-1} 0.71$,
6. The illumination normal to the triangle, employing Eq. (5-1),

$$E_N = 4000 \times 0.71 = 2840 \text{ footcandles}$$

7. The average diffuse reflectance of ΔFGJ , employing Eq. (5-2),

$$\rho_D = \frac{1300}{2840} = 0.46$$

Measurement of the luminance of the remaining triangles, together with continued application of the above equations, will yield the average diffuse reflectance of every surface of the model.

5.10 Interrogation of a Point-and-Reflectance Model

A point-and-reflectance model can be interrogated for properties that can not be extracted from the point model. For example, in a point-and-reflectance model of a scene containing dark rock and light sand, the computer could determine their boundary and report it to Earth; or, on command from Earth, the robot could explore one area and ignore the other.

5.11 The Earth Station

Two ways of transmitting pictorial data to Earth are being planned. One is transmission of television pictures, either monocular or stereoscopic, the other transmission of the model of the surface of Mars employed by the robot in making decisions. The model can be presented on the same screens (Fig. 5-8) as the television pictures, provided a computer is employed on Earth to store it and arrange it for display. The arranging process can be such as to give the viewer the impression of moving through the scene.

If the model is composed of points and reflectances then the Earth operator will command under what illumination he views it. Figure 5-9 illustrates a possible display whose luminance to the eye of the operator is computed from the reflectances of the surfaces and from illumination that he has commanded to come from the upper right of the scene.

5.12 Acknowledgement

The concepts presented above and corollaries yet to be described evolved with the help of

Prof. John T. Rule, former Head, Dept. of Graphics, MIT

Dr. Adelbert Ames, former Director, Dartmouth Eye Institute

Dr. S. Howard Bartley, former Research Associate, Dartmouth Eye Institute

Nickerson Rogers, former Research Fellow and stereoscopic photographer, Dartmouth Eye Institute

Paul Sample, former Artist-in-Residence, Dartmouth College

Willard Allphin, Illumination Research, Sylvania Electric Products Co., Salem, Mass.

Dr. Parry Moon, Associate Professor of Electrical Engineering, emeritus, MIT

Dr. Domina Eberle Spencer, Professor of Mathematics, University of Connecticut, Storrs, Conn.

Robert N. Davis, Research Staff, Lincoln Laboratory, MIT

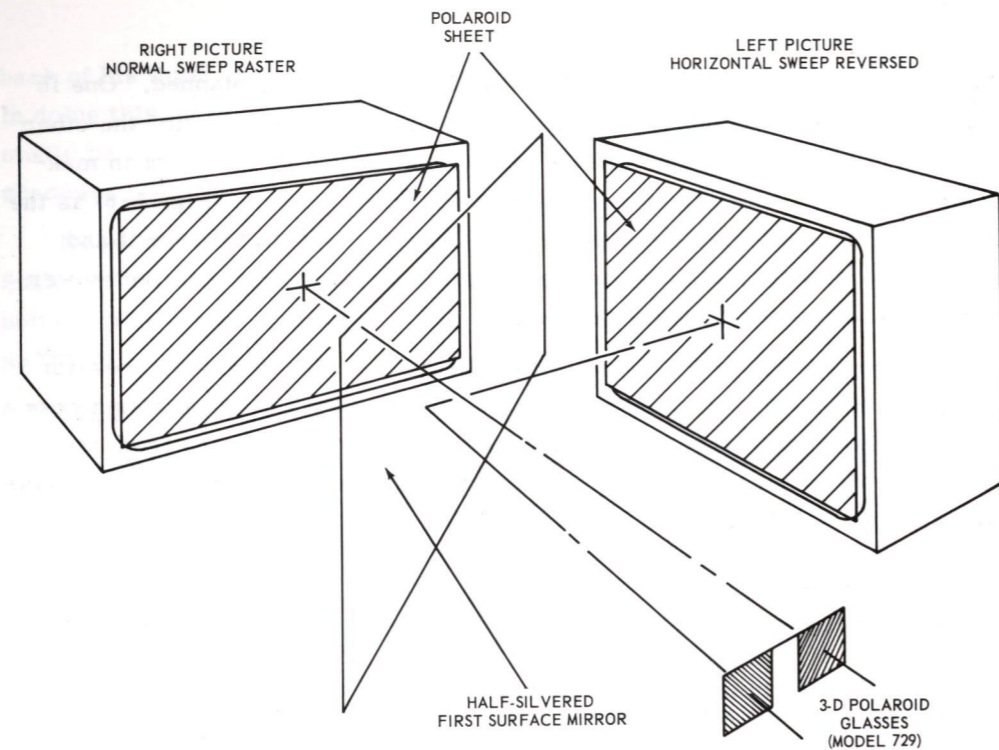


Fig. 5-8. Stereo picture presentation using two 23-inch monitors and Polaroid filters. Orientation of each filter in glasses corresponds to orientation of filter before each monitor.

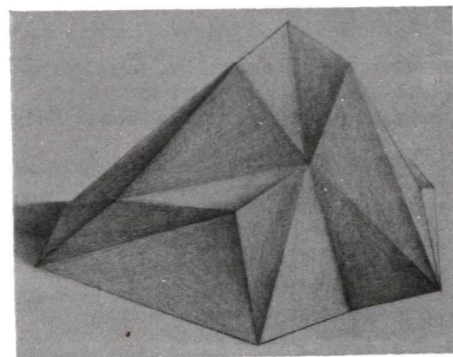


Fig. 5-9. Triangularized view of the mound in Fig. 2-3 shaded as though by an Earth computer.

SECTION 6

ADDITION OF A LEARNING ABILITY TO THE DECISION-MAKING SUBSYSTEM

by

William L. Kilmer, Jay Blum and Warren McCulloch

6.1 Introduction

In a temporally changing environment, a static decision-making computer such as the model described in Refs. 3 and 4 may be at a significant disadvantage. Consequently, that model is being extended from that which has been S-RETIC (where S denotes a static environment) to a model of a decision computer capable of several forms of conditioning and learning. The new model is called STL-RETIC, where T denotes a temporally changing environment, L a conditioning and learning ability of the kind described in Section 9.

As simulated on the Instrumentation Laboratory general purpose computer, the static S-RETIC model consists of twelve modules (M_i), inputs (γ 's) to the modules from the simulated environment (Σ), and outputs (P) to threshold elements. The decision computer commands a mode of activity only when a threshold has been exceeded.

To make learning possible, the model is being modified to include logic and memory elements reinforceable as a function of experience. The logic elements will operate in the A-part of each module, while memory elements will store past inputs from the environment, messages from other A-parts, and past A-part computations. Figure 6-1 shows each γ input to an M_{i_a} paralleled by a memory element and each connection (ω) from another M_{i_a} , similarly paralleled. The number of γ inputs from each M_{i_a} has been increased from five to seven. Note that the design continues to be one that can be built of hardware as well as simulated.

6.2 Requirements

The augmented STL-RETIC structure ought to satisfy the following requirements:

First, if the environment is rich in stimuli and if the responses which follow these stimuli are at least partially unknown, it is not possible to establish an input-output correspondence table defining for the decision computer which acts should follow which stimuli. The computer must be trained.

Second, the decision mechanism should respond differently to different overall environments; it should be able to adjust to a new input-output specification, so that it need not be reprogrammed.

Third, if one or more of the robot's sensory or action subsystems fails, the decision computer should compensate by changing its responses to input stimuli.

6.3 Nature of the Problem

A difficulty in augmenting S-RETIC to be able to learn is that each module appreciates only part of the input-output correspondence. How can all of the modules function harmoniously in the same direction, so that the effect is cooperative and not chaotic? Generally, each module is most sensitive to different input stimuli; for example, module 5 may discriminate among overall (Σ) input sets 67, 68, and 69, whereas module 4 may discriminate among inputs 23, 24, and 25. Moreover, neither may discriminate at all among other input sets. If one module makes necessary distinctions between two cases which call for different responses, it has to be able to notify neighboring modules of these distinctions as they occur. This would enable the other modules to act appropriately.

6.4 Approach to the Problem

The problem mentioned in subsection 6.3 may be limited by requiring that the decision-making computer be effective only over several hundred successive stimuli. Each call for a change in response would be stored in memory elements so that, when several hundred had been accumulated, actual alteration of the response would occur. Naturally, the smaller the number of stimuli required, the better the model will be able to yield up-to-date responses. Thus, the benefit of reduced complexity achieved through storage of desired responses must be weighed against the delays intrinsic to such storage.

6.5 Implementation by Operations on P_{AB} Vectors

Each kind of learning will be implemented by operations on the P_{AB} vector which goes from the A to the B part of each module. These operations will be similar to those described in subsection 5.6 of Ref. 4: flatten vectors, elevate vectors on certain modes, depress them on certain modes, measure the peakedness of vectors in order to decide what to do with them, and permit one vector to change into another systematically, smoothly, and gradually.

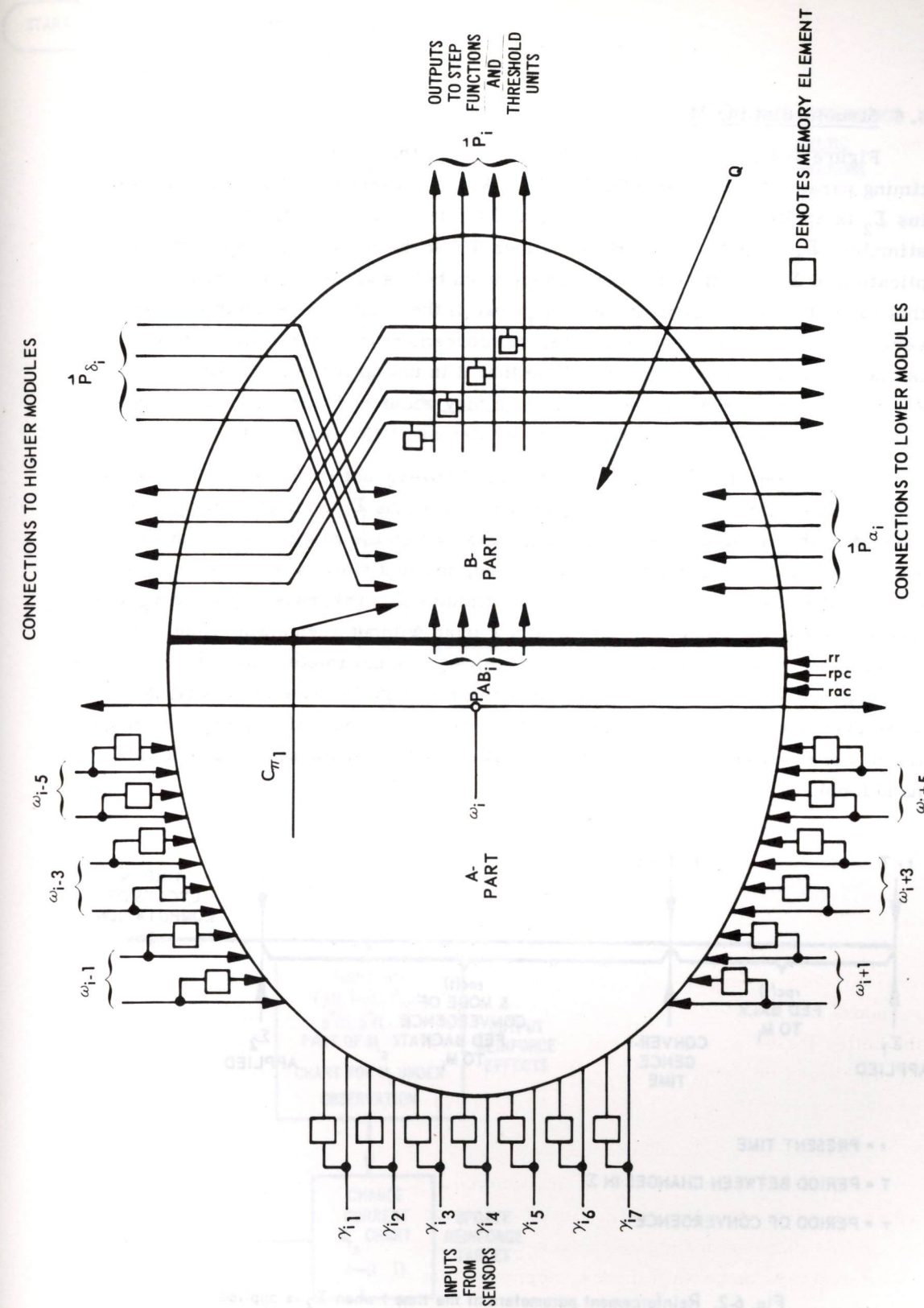


Fig. 6-1. Input and output connections to parts A and B of module M_i .

6.6 Structure of the Model

Figure 6-2 is a schematic representation of the relation between the three timing parameters used in STL-RETIC. At the present time t at which a stimulus Σ_2 is applied, the model looks back to the time of the application of the last stimulus, Σ_1 - the time difference is then T . If the model converged after application of Σ_1 , the time to convergence from $t-T$ is called τ , with $rpc(t)$, in this case, the reinforcement due to Σ_1 , $rac(t)$ the reinforcement due to convergence. If no convergence occurs, reinforcement over the entire period T is that due to Σ_1 , namely $rpc(t)$. Note that it is not until the application of Σ_2 that reinforcement due to convergence may occur and that consequently the reinforcement signals rpc and rac are written as $rpc(t)$ and $rac(t)$.

A condensed flow chart of the stimulus-convergence-reinforcement process is given in Fig. 6-3. After getting the first stimulus Σ , computations based on receipt of the stimulus are carried out, after which the model checks for convergence. If convergence has not occurred, present time is incremented one unit and the model looks for a new input stimulus Σ . Otherwise, if convergence has occurred, the model checks for a new input without augmenting time. If, in either case, a new input has not been received, the parameters used to determine convergence are recomputed. Note that these change in value from what they were previously only if time has been incremented. Thus, if convergence occurs, the model remains converged until receipt of another stimulus. (Stays in left-hand loop).

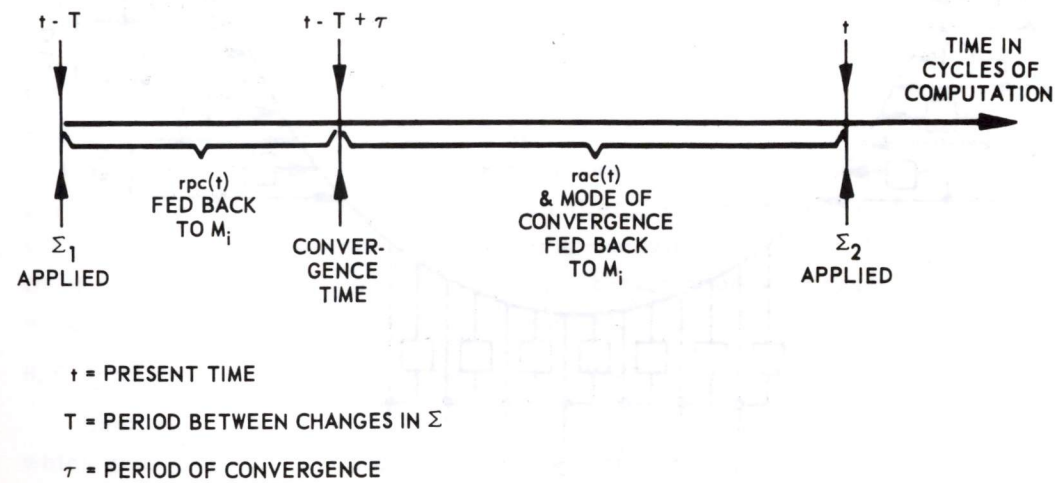


Fig. 6-2. Reinforcement parameters at the time t when Σ_2 is applied.

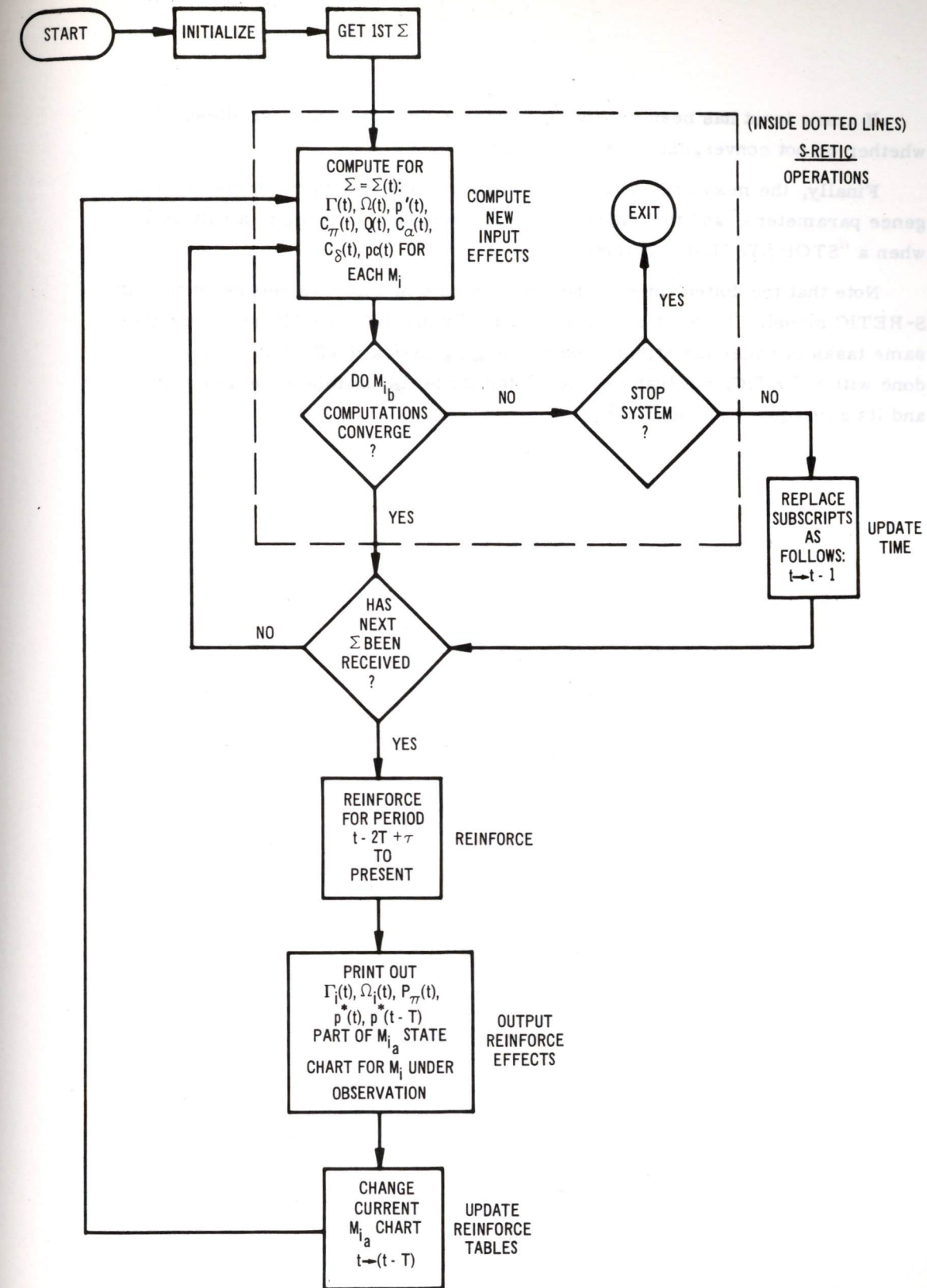
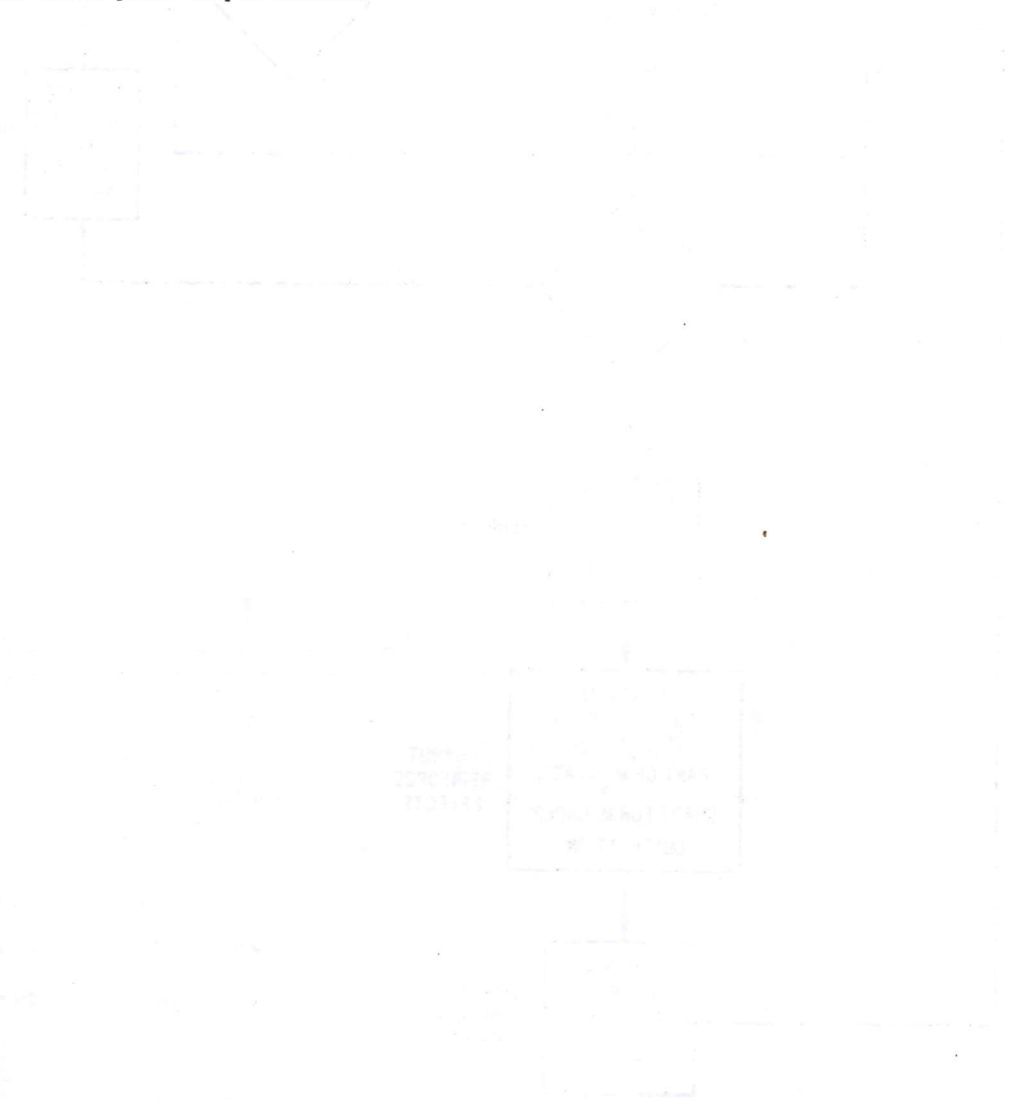


Fig. 6-3. STL-RETIC flow chart.

If a new input has been received, reinforcement occurs, regardless of whether or not convergence has been attained.

Finally, the model recomputes the effects of the new input on the convergence parameters, and the entire process is repeated, coming to a halt only when a "STOP SYSTEM" command is given to it.

Note that the dotted lines in the upper middle of Fig. 6-3 represent the old S-RETIC model. Thus, it can be seen not only that STL-RETIC performs the same tasks of receiving stimuli and converging upon a mode of activity as were done with S-RETIC, but that the augmented model is capable of reinforcement and its consequent output effects.



PART II

RESEARCH INTO THE NATURE OF ANIMAL SENSORY,
DECISION AND CONTROL SYSTEMS

SECTION 7

ANIMAL COMPUTATION

by

Louis Sutro and David Lamport

with the technical assistance of Warren McCulloch

7.1 Background to a Computer Model of a Vertebrate Brain

From his life-long study of the human nervous system, Dr. Warren McCulloch has concluded that the essential properties of human computation must serve as a basis for the electronic acquisition and storage of human knowledge. Although as a neurologist, psychologist, and physiologist he is aware of the dangers involved in embodying mental functions in physical devices, he has nevertheless developed a simplified model of a vertebrate brain. His intention is merely to suggest an organizational hierarchy necessary for efficient robot performance.

Figure 7-1 outlines his model of the vertebrate nervous system, identifying what he feels are five principal computational areas and their functional connections. At the left is the retina, consisting of three layers of cells, two of which are engaged in computation. The eye is shown as representative of the senses because its computational capacity qualifies it as a principal computer; it is the foremost data source to the primate brain, providing two million of its three million inputs.

At the upper left is the cerebrum, which Dr. McCulloch calls the "great computer" and in which computation is carried out in many layers; each of which, in man if unfolded, would be about the size of a large newspaper.

The computer which controls all others is shown at the center right. It is the reticular core of the central nervous system and extends from the base of the brain through the spinal cord. It makes the major decisions, or rather, shifts the focus of attention so as to determine what is to be done from moment to moment. By deciding which command function is to take over, it thus controls all other computers and, through them, the whole organism. A model of this computer was presented in Refs. 3 and 4 and further developed in Section 6 of this report.

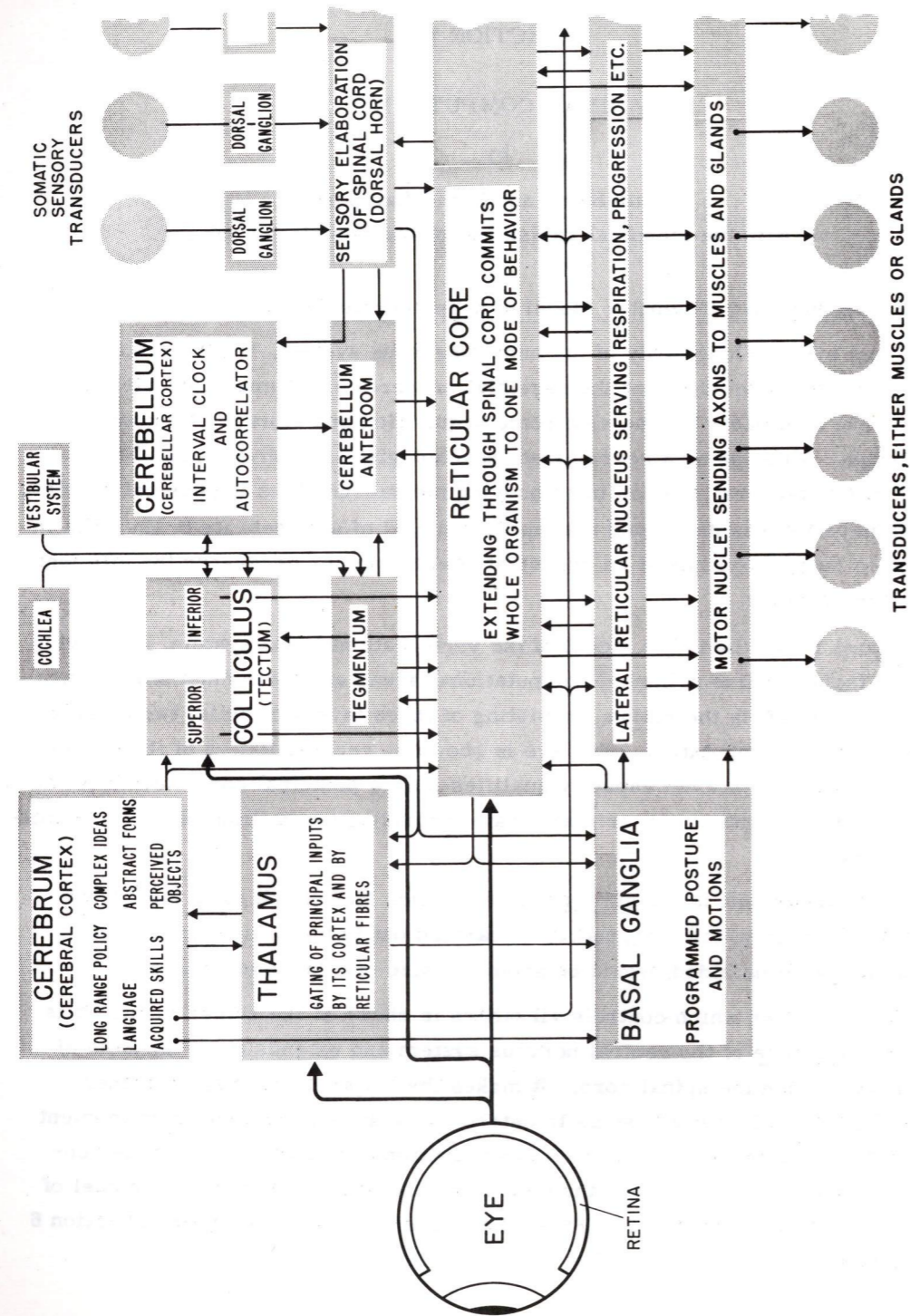


Fig. 7-1. Block diagram of generalized vertebrate nervous system.

Clusters of nerve cells at the base of the cerebrum comprise the basal ganglia, a computer shown at the lower left of the figure. Here are programmed all innate or learned movements, such as feeding, walking or throwing a ball. Additional programs may be acquired through the growth of connections to the motor control nerve cells, shown along the bottom of the illustration.

Completing the list of principal computational areas is the cerebellum, shown at the top of Fig. 7-1. It programs the completion of a movement, such as reaching to touch an object, and requires inputs both from the vestibular system, to detect tilt and acceleration of the head, and from skin and muscle sense cells, to detect posture and information as to what is being touched.

Interconnected with the principal computers are switching structures, such as the thalamus, colliculus, and cerebellum anteroom. In lower vertebrates (fish, amphibians, birds), the colliculus perceives form and movement; in higher animals, it is subdivided into a superior colliculus, which determines the direction of gaze under the general control of the cerebral cortex, and an inferior colliculus, which is probably concerned with auditory and vestibular inputs as well as the somatic body image. The base around the colliculus is the tegmentum.

Around the reticular core are specialized structures that could also be called computers, such as the nucleus of nerve cells that control respiration and other routine bodily functions, and the dorsal horn of the spinal cord, through which pass inputs from sense cells. Note that the reticular core acts on all other computers and that they report to it. It reaches decisions with the aid of raw data from sensory systems as well as processed data from other computers, the latter of which comprising its primary input source.

The computers of Fig. 7-1 are shown as arranged in animals with horizontal spines. Monkeys and man have the same computers in approximately the same relation, but the arrangement is tilted, with the cerebrum, now very much larger, at the top.

All these computers have a common ancestry. All evolved from the central computer, the reticular core, and in so doing have established only those interconnections necessary for efficient communication with "retic." In this manner the reticular core has been constantly able to maintain the complexity necessary to meet the aggregate demands of the entire system.

7.2 A Possible Engineering Equivalent

Figure 7-2 is a diagram equivalent to Fig. 7-1 but labelled with engineering terms to suggest how the system can be simulated. In place of the retinae are the cameras and the visual first-stage computer.

First-stage computers receive inputs from all of the senses – auditory, vestibular and somatic sensory. Each is called a computer rather than a pre-computer or preprocessor to indicate that it receives feedback from the central computers.

Other substitutions are as follows:

decision computer	for	reticular core
associative computer	for	cerebral cortex
timing, coordinating and autocorrelating computer	for	cerebellum
computer of effector sequences	for	basal ganglia (nucleii)
computer of specialized controls	for	lateral reticular nuclei

7.3 Memory in Biological Computers

If models of the computers described above are to be built, what will be the form of the memory?

Computer designers recognize three principal forms, but there is a fourth more important for our present work. The first form is storage on tape where data is found by searching linearly. The second form is random access where data is found by its address. In the third form, data is found by searching for part or all of its content. This has unfortunately come to be called "associative" memory.

The fourth form is the true associative memory employed in animal brains. Here data is found by a relation. For example, if you are looking for a friend in a crowd, you may find him by recognizing the proportions of his face or the time of day that he passes by. These are relations. You acquire this form of memory by constantly updating a model of the world in your head. A change in the model is considered to be the formation of a trace embodying the relation. (The connection for the trace may have existed but have had a different threshold than that required.) Conditioning, which is the most elementary method of forming trace, is described in Section 9. This form of memory will be simulated, at first, by the other three. Hardware (or chemistry) more appropriate to a true associative memory may evolve.

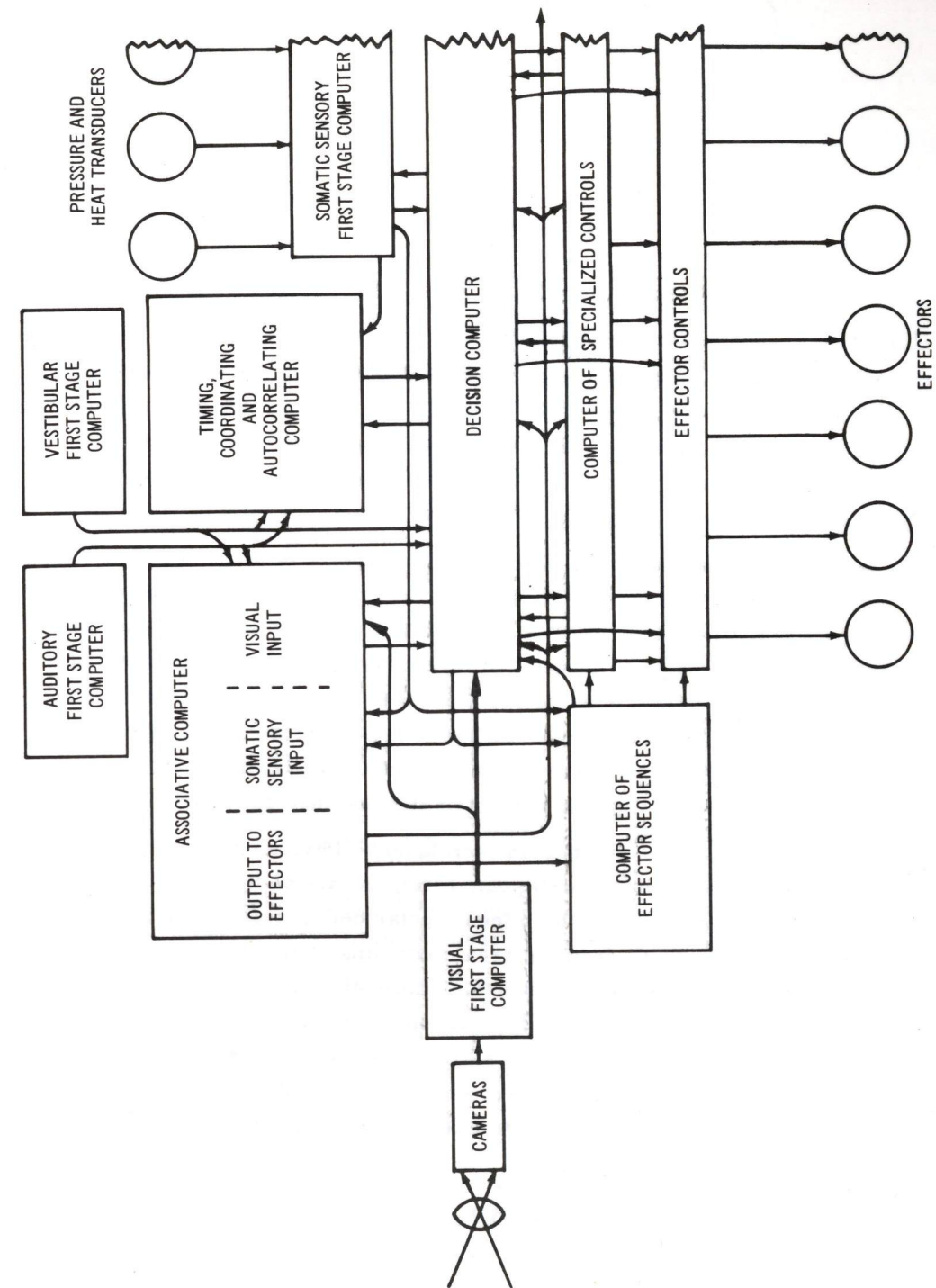


Fig. 7-2. Block diagram employing functional engineering nomenclature.

SECTION 8

VISUAL PROCESSING IN ANIMALS

by

Roberto Moreno-Diaz

8.1 Development of a Computational Model

Existing models of visual processing in the frog's retina and tectum (Refs. 17, 21, 22) suggest a concept which may be applied to retinas of higher animals as well. It is hoped that a model of this concept will soon be realized which, founded in known anatomy and physiology, may be further extended to apply to tasks such as those described in Section 4 of this report.

The concept of visual processing being developed may be regarded as a layered computation; that is, elements of a similar nature comprise a given layer, and computation occurs first through interconnections between elements in the same layer, then between those of different layers. In a gross sense, a succession of such layers will come to represent the higher animal retina. Although there are only three significant cell types in vertebrate retinæ (photoreceptor, bipolar, ganglion), the number of computational layers is larger, since interaction may occur between dendrites at more than one level.

What is the nature of computation in each layer? Physiology suggests the possibility, at the photoreceptor-bipolar level, of lateral inhibition. Assume a mosaic of photoreceptors which feed bipolar cells, such that a particular bipolar receives inputs from an area containing many photoreceptors by means of photoreceptor ends, or "fibers." Lateral inhibition may then be formulated as follows. Consider a typical bipolar cell, called "B" in Fig. 8-1. It receives both an excitatory signal from the photoreceptor - or photoreceptors - situated directly above it as well as inhibitory signals from the surroundings. If the effect of the inhibition is subtractive, the total signal arriving at the bipolar B would be

$$B = \sum_{A_1} F_i - \sum_{A_2} F_j \quad (8-1)$$

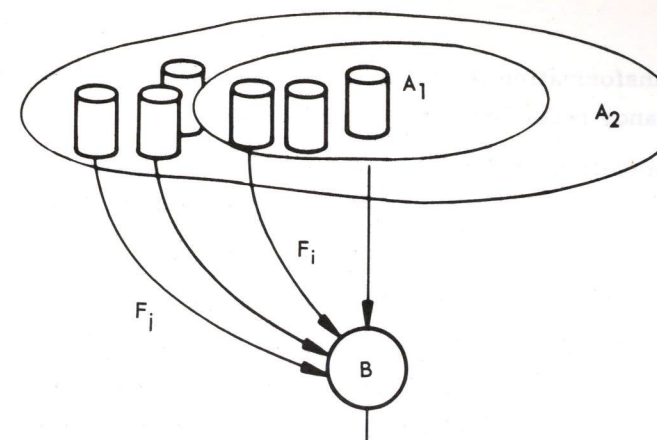


Fig. 8-1. Dendritic level, defined by areas A₁ and A₂, with dendrites leading to cell body B.

where F_i is the output of a photoreceptor transmitting an excitatory signal, F_j the output of an "inhibitory" photoreceptor, and A_1 and A_2 the areas of excitatory and inhibitory photoreceptors respectively.

Equation (8-1) refers to interaction only between the photoreceptor-bipolar levels. Regardless of the location of such interaction in the retina, it is possible to postulate a corresponding layer which produces the same effect. Let $F(\alpha, \beta)$ be the output of an ideal photoreceptor situated at the point (α, β) (in a suitable coordinate frame superimposed over the retina). This corresponds to the input of an idealized computational layer, called a layer of lateral inhibition. Let $B(x, y)$ be the output of that layer corresponding to the point (x, y) . The transformation rule that defines the computation at that layer is then

$$B(x, y) = \frac{1}{N_1} \sum_{(\alpha, \beta) \in A_1} F(\alpha, \beta) - \frac{1}{N_2} \sum_{(\alpha, \beta) \in A_2} F(\alpha, \beta) \quad (8-2)$$

where summation is over all points (α, β) in areas A_1 and A_2 , and N_1 and N_2 correspond, respectively, to the number of elements in A_1 and the number in A_2 .

The values of the coordinate variables x , y , α , and β , according to the discrete nature of the retinal mosaic, are first considered discontinuous, taking, for example, values 1, 2, 3, If the area A_1 is reduced to a single element, that at position (x, y) , Eq. (8-2) becomes

$$B(x, y) = F(x, y) - \frac{1}{N_2} \sum_{A_2} F(\alpha, \beta) \quad (8-3)$$

If the elements of the retina are sufficiently small and numerous, the summation may be replaced by an integral, to give

$$B(x, y) = F(x, y) - \frac{1}{A_2} \iint_{A_2} F(\alpha, \beta) d\alpha d\beta \quad (8-4)$$

The effect of this transformation is known to produce contrast enhancement (Ref. 23) and redundancy reduction of the initial image.

In the above formulation, variables B and F are considered time independent. This restriction will be eliminated in future modeling to take time into account.

8.2 Generalization of Interaction Coefficients

Equation (8-2) defines computation over specific areas, giving equal weight to each photoreceptor in a field. That equation may be generalized by introducing a set of coefficients which reflect the relative strength of a particular input fiber connected to the output fiber under consideration. Let $C(x, y, \alpha, \beta)$ be the strength of the input in position (α, β) over the output in position (x, y) . Equation (8-2) then becomes:

$$B(x, y) = \sum_R C(x, y, \alpha, \beta) F(\alpha, \beta) \quad (8-5)$$

where the summation is over the entire retina.

That Eqs. (8-2 and 8-3) are particular cases of Eq. (8-5) is easily verified. To obtain Eq. (8-2), let $C(x, y, \alpha, \beta)$ be $+1/N_1$ over the area A_1 , $-1/N_2$ over A_2 , and zero elsewhere; Eq. (8-2) is then the consequence. Let $C(x, y, \alpha, \beta)$ be $+1$ for $C(x, y, x, y)$, $-1/N_2$ over A_2 , and zero elsewhere; the result is Eq. (8-3).

Equation (8-5) is called the general expression for lateral interaction, of which lateral inhibition is a specific case. In the continuous retina, Eq. (8-5) becomes

$$B(x, y) = \iint_R C(x, y, \alpha, \beta) F(\alpha, \beta) d\alpha d\beta \quad (8-6)$$

Theoretical work continues at present to investigate various forms of the weighting coefficients to obtain different processes. It has been found that Eq. (8-6) explains several properties of the visual system - such as reliability, in the sense that the destruction of a fiber $B(x, y)$ or a group of them, produces not blind spots but merely a general loss of resolution. Such points will be examined in detail in future reports.

Computations as expressed in the above equations are not easy to carry out analytically in any but the most simple forms of $F(\alpha, \beta)$ - which are not the most interesting ones, but clues about the behavior of the transformation have been obtained analytically through treating the simple one-dimensional case. For a real image in the retina, however, the form of $F(\alpha, \beta)$ is extraordinarily complicated and requires a simulation of the kind presented in Section 4.

8.3 Examples of Lateral Interaction

For the one-dimensional case, Eq. (8-4) becomes

$$B(x) = F(x) - \frac{1}{2s} \int_{x-s}^{x+s} F(x) dx$$

where $2s$ is the size of the inhibitory area, which in this case is a line. Statisticians call the second term a "moving average". $B(x)$ may be regarded as the result of a transformation T_s upon $F(x)$. The transformations of some simple functions are:

a) When $F(x) = Ag(x)$, with $A = \text{constant}$, then the transformation is:

$$T_s(Ag(x)) = AT_s(g(x)).$$

Proof:

$$\begin{aligned} T_s(Ag(x)) &= Ag(x) - \frac{1}{2s} \int_{x-s}^{x+s} Ag(x) dx \\ &= A \left[g(x) - \frac{1}{2s} \int_{x-s}^{x+s} g(x) dx \right] \\ &= AT_s(g(x)) \end{aligned}$$

b) When $F(x) = Ag(x) + Bh(x)$, with A, B constant, then $T_s(F(x))$ is:

$$T_s(Ag(x) + Bh(x)) = AT_s(g(x)) + BT_s(h(x)) \text{ as suggested by the above proof}$$

c) For $F(x) = A$, with $A = \text{constant}$,

$$T_s(A) = 0$$

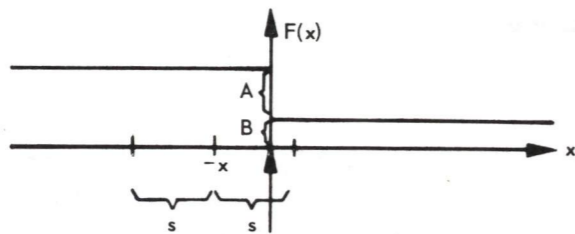
d) For $F(x) = Ax$, with $A = \text{constant}$,

$$T_s(Ax) = 0$$

Proof of d:

$$\begin{aligned} T_s(Ax) &= Ax - \frac{1}{2s} \int_{s-x}^{s+x} Ax dx \\ &= Ax - \frac{1}{2s} \left[\frac{A(s+x)^2}{2} - \frac{A(s-x)^2}{2} \right] \\ &= Ax - \frac{A}{4s} [s^2 + 2sx + x^2 - s^2 + 2sx - x^2] \\ &= Ax - \frac{A}{4s} [4sx] \\ &= 0 \end{aligned}$$

e) If $F(x)$ has a step discontinuity:



For x negative, the transformation of the above function is:

$$T_s(F(x)) = A + B - \frac{1}{2s} \int_{x-s}^0 (A+B) dx - \frac{1}{2s} \int_0^{x+s} B dx$$

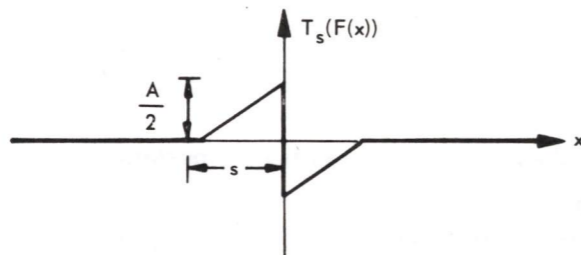
$$= \frac{A}{2} \left(1 + \frac{x}{s}\right) \quad -s \leq x < 0$$

For x positive:

$$T_s(F(x)) = B - \frac{1}{2s} \int_{x-s}^0 (A+B) dx - \frac{1}{2s} \int_0^{x+s} B dx$$

$$= \frac{A}{2} \frac{x}{s} - 1 \quad 0 < x \leq s$$

Plotting T_s over the entire domain of x ,

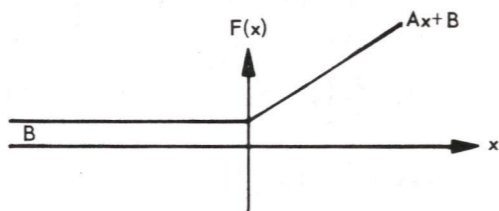


For $x < -s$ or $x > s$,

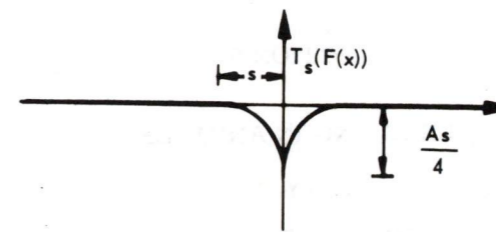
$$T_s(F(x)) = 0$$

because the function is a constant.

f) If $F(x)$ is a ramp



then T_s is



SECTION 9

LEARNING IN ANIMALS

by

Charles Sigwart

9.1 General

At the end of Section 7 memory traces in the nervous system were mentioned as the embodiment of relations comprising a model of the environment. The possibility of changing such a model with new environmental situations implies memory and learning. Attempts to define learning and memory have occupied the efforts of many neurophysiologists and psychologists for a long time. Thus there exists a body of literature concerning mathematical learning theory based on behavior and a range of hypotheses about the relation of learning and memory to structural, chemical and functional properties of the nervous system.

Learning is a term which encompasses a range of phenomena extending from simple changes in responsiveness to a stimulus to the very complex capacity for dealing with abstractions possible in man. If we can accept a broad generality, "We can define learning as that process which manifests itself by adaptive changes in individual behavior as a result of experience..."⁽²⁴⁾

Learning cannot be defined more precisely since mechanisms underlying many complex learning phenomena are not understood. The more elementary forms of learning such as habituation and conditioning have been subjected to neurophysiological study. In the following paragraphs of this section various concepts relevant and desirable in the behavior of a robot will be described. A more detailed discussion of recent research on learning may be found in Ref. 25 and Ref. 26.

9.2 Classes of Response

The decision computer's responses to stimuli may be classed as either momentary or predictive. Those which are momentary are a consequence only of the stimuli which they directly follow; those which are predictive occur when the decision computer receives a stimulus recognized to be a precursor of a later event and acts in accordance with that event. STL-Retic will implement

momentary response properties such as habituation and classical conditioning as well as predictive responses by incorporating operant conditioning and generalization. Implicit in providing the capacity to change response patterns is the ability to modify generalizations when they become inappropriate.

9.3 Classical Conditioning

Classical conditioning (also called Pavlovian and Type I conditioning) involves the association of a reflex response with a new stimulus. For this to occur the new stimulus which is to become a conditioned stimulus (CS) must regularly precede the unconditioned stimulus (US), which normally produces the unconditioned response (UR). It is assumed that the 'stimulus' may include many factors in a vertebrate animal.

"Sensory signalling that takes place during conditioning involves at least two differentiable sensory channels: specific and nonspecific. The nonspecific channel apparently establishes background conditions for widespread reception and retention of particular sensory and motor signals which may enter by way of specific and nonspecific channels... The net result of all this is an alteration in distribution and influence of incoming and outgoing signals which increases the likelihood that certain responses will occur in relation to certain previously experienced stimuli."⁽²⁷⁾

A working definition for classical conditioning is as follows:

"A stimulus class (CS) that before the conditioning procedure does not produce a response, or produces it with low incidence will, after the conditioning procedure, produce the response with a frequency above some specified criterion. The conditioning procedure consists in pairing of the CS with another stimulus (US) that ordinarily produces an unconditioned response (UR). The paired presentations will be in a certain order (CS - US) with a CS-US interval with specified limits both with respect to mean duration and variability. The CR must meet criteria of resemblance to the UR."⁽²⁸⁾

The production of new relations between a stimulus situation and a response may be considered as a memory allowing the building up of a class of stimulus situations to which a response is elicited.

9.4 Habituation

Habituation is "...the relatively persistent waning of a response as a result of repeated stimulation which is not followed by any kind of reinforcement. It

is specific to the stimulus and relatively enduring... distinct from fatigue and sensory adaption. ⁽²⁹⁾ It is assumed to be a central phenomenon, exclusive of receptor or effector fatigue.

The main function of habituation is to eliminate insignificant responses by reducing sensitivity to a specific stimulus which is repetitive. The man who hears a jack hammer going in the street stops paying attention to it after a while. Similarly the decision computer should ignore stimuli which it finds unimportant. When the robot does something new in its environment, or when that environment changes, such stimuli will occur and the decision computer should eradicate their traces as they enter over sensory subsystems.

9.5 Discriminative Learning

The capacity to generalize, given a set of stimuli appropriate to a particular response, so as to include in that set related stimuli, is known as discriminative, or perceptual, learning. It is directly involved with conditioning and habituation, in that effective response to an appropriate stimulus requires correlated sensory data.

Since changes resulting from generalization are reversible, those that are inappropriate should be unlearned as soon as a contradiction arises between results of behavior and expectation. Thus, the capacity to generalize depends on perceptual discrimination as well as information from the senses.

9.6 Instrumental Conditioning

Instrumental conditioning (also known as operant or Type II conditioning or trial and error learning) involves the modification of a response pattern dependent on a reinforcement. This is in contrast to classical conditioning, where reinforcement shapes the selection of the appropriate stimulus. Instrumental conditioning and its variants involve modification of a response that was initiated by the animal. In essence the animal is changing the probability of a particular response to a given situation consequent upon the reinforcing result, whether painful or pleasurable. Thus this type of conditioning includes avoidance conditioning, where a potential hazard is recognized and elicits a response to avoid it.

It is clearly important for the robot to eliminate undesirable responses to hazards. A task of the decision computer must be to seek other responses to lessen hazards in situations it encounters.

SECTION 10

SUMMARY AND CONCLUSION

The search for evidence of life and geological changes on Mars we interpret as a problem in reducing the data collected initially to a smaller amount of significant data to be sent to Earth, and of changing the mode of operation of the data-acquiring device (robot). Reducing the amount of visual data is called visual processing and modelling; changing the mode of operation, decision. We aim by data reduction to employ the limited transmitted information more efficiently than by conventional TV. By permitting the robot to change its mode of operation we aim to enable it to decide what to do next.

Components have been described which, when tied together, will comprise the visual subsystem of a robot. The computer-controlled television camera is sufficiently open in its design to permit analysis and resynthesis as demands upon it change. A succession of stereo-optics designs, from C1 through D1 to E1, has led back to a type with fixed convergence, called C2, as the one most likely to be developed next. Means have been devised of processing stereo pairs of pictures to extract range. Two structures have been planned to store a model of the environment, one employing points, the other points and reflectances. Data reduction is achieved first in the formation of a model and second in the selection of parts of that model (interrogation) for transmission to earth. One purpose of this interrogation is to search for evidence of life and geological changes. Another is to find paths that the robot can follow.

The contact subsystem is a hand-and-arm and a means of modelling the environment contacted. The hand is being planned at present both to feel the shape of objects and to provide a signal, such as a flashing cross, on which the cameras can converge at the start of the exploration of a scene.

The decision subsystem, previously devised to respond to a sequence of static environments, is now being augmented by logical and storage elements so that it can learn. The structure of this subsystem is such that while it is now in the form of a computer program it can later be implemented in special-purpose hardware.

The mobility subsystem is at present, a vehicle, wheel drives and controls, originally designed for a lunar landing by the General Motors Defense Research Laboratory and since adapted by them for a Martian landing. The synthesis of visual, contact, decision and mobility subsystems we call a robot. Three configurations of this robot have been pictured, the preferred one having the camera assembly directly above the front axle of the vehicle. All are capable of mapping by measuring distances with an odometer and angles with a radio direction finder aimed at transponders previously dropped by the robot. Accuracy can be improved by sighting landmarks.

While this program is expected to be primarily one of computer development, only commercially-available computers are used at present. The purpose of this is to determine by simulation on existing machines the specifications of the desired machine.

The first part of this report concerns the developments just summarized. The second part describes research into the organization of animal nervous systems, into the nature of visual processing and the nature of learning in animals. The organization of animal nervous systems is that found by Warren McCulloch and presented here in computer language. Visual processing is that found by physiologists and presented here analytically by Roberto Moreno-Diaz. "Learning in animals" is a summary of this subject as presently understood in psychology and physiology. These investigations will be the basis of further development of the kind described in the first part of the report.

APPENDIX A
MEASUREMENT OF VIDICON RESOLUTION

by

David Tweed

The method employed in the measurement of vidicon resolution is that described in Ref. 25. Figures A-2 and A-3 in that reference are replaced respectively by a photograph of an oscilloscope presentation of the output of the C1-1 camera chain and a graph derived from this presentation. Quoting from Ref. 25:

"A new type of test pattern was devised (Fig. A-1) which consists of nine line-groups; each of these groups contains a black and white bar to represent 100% response factor, followed by ten sets of four black and three white lines which represent 100 to 1000 TV lines (across the width of the chart), in 100-line increments.

"This new chart makes it possible to obtain data for a complete square-wave aperture response curve with one oscilloscope presentation. Figure A-2 is typical of the type of presentation that is obtained from an oscilloscope that is fed the video signal and set to select one horizontal scan line when an image from the line selector pattern is being transduced.

"The procedure for obtaining the aperture response curve is as follows:

1. The tube is set up under the desired test conditions, in this case optimum overall performance.
2. The test pattern is the "line selector" pattern of Fig. A-1.
3. The video signal from a selected horizontal line is presented on an oscilloscope and a photograph is taken. See Fig. A-2.
4. The photograph is measured for the relative response of the line-sets to that of the black and white bars, and a curve is plotted. See Fig. A-3."

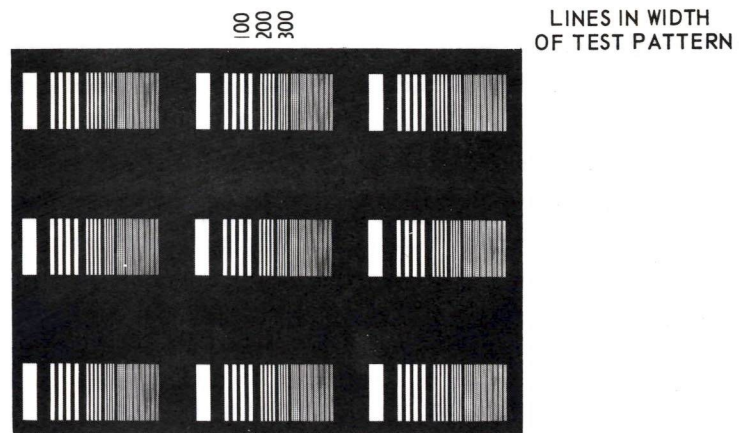


Fig. A-1. "Line Selector" test pattern (Westinghouse resolution chart ET-1332 purchased from Tele-Measurements Inc., 145 Main Ave., Clifton, N.J.), reproduced 1/3 full size.

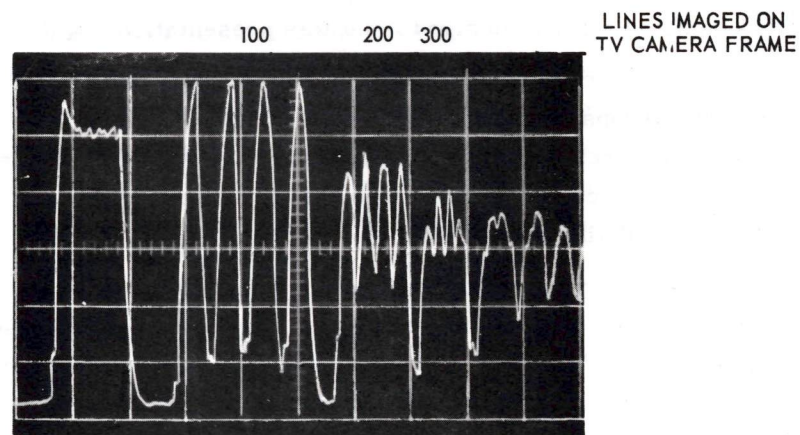


Fig. A-2. Tracing of oscilloscope presentation of video signal from part of single scan line of Fig. A-1, employing the amplifier described in subsection 3.5.

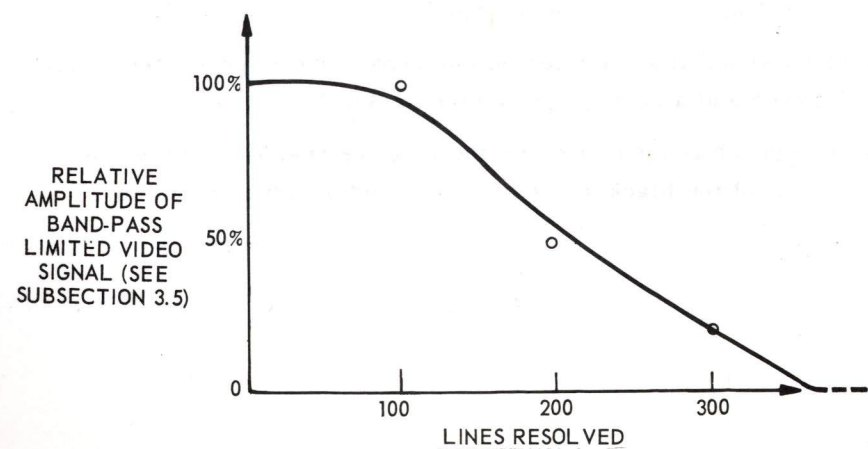


Fig. A-3. Square-wave aperture response curve derived from Fig. A-2.

Note to Section 3

APPENDIX B

RANGING AND RANGE ERROR

by

Charles Sigwart

B.1 General

To obtain the range of objects in the environment it is possible to use a stereo optical system and in essence triangulate. Thus the range z to point P is determined by measuring α_1 and α_2 , knowing that the base separation of the two optical systems A_1 and A_2 is $2b$. The measurement of range is uncertain to the extent that the measurement of convergence angles α_1 and α_2 is uncertain. In general there will necessarily be some small angular uncertainty $\pm\delta$ in measuring the orientation of the optic axis. Thus the localization of P by the method is in a volume bounded by the intersection of the two cones whose axes intersect at P . The cones have a half angle δ as shown in the plane of the axes in Fig. B-1. The intersection of the plane of the optic axes with this volume produces the quadrilateral of uncertainty surrounding P .

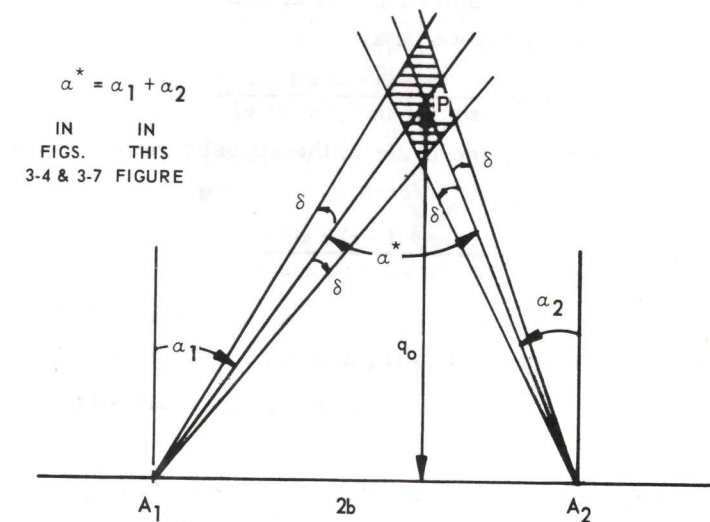


Fig. B-1. Quadrilateral of uncertainty surrounding point P when viewed by cameras at A_1 and A_2 .

This zone of uncertainty varies with α_1 and α_2 as indicated in Fig. B-2. Thus it is apparent that ranging on objects along the line q_0 away from the z axis results in an increasing uncertainty of measurement and that the range error is minimized by restricting the convergence angles such that $\alpha_1 = \alpha_2$. The effect shown in Fig. B-2 is of course amplified by the large angular uncertainty δ shown, but the general result is that range uncertainty increases significantly with deviation from the condition of equal convergence angles. Thus the following discussions will deal with range uncertainties for points on the center line.

B.2 Range and Uncertainty for Points on the Q Axis

To keep the range uncertainty minimal as indicated above, we will consider only the range of points on the Q axis, employing equations derived by Roberto Moreno-Diaz and David Tweed. The range of an object at P shown as OP in Fig. B-3 is determined from

$$q = \frac{b}{\tan \alpha} \quad (B-1)$$

The range is measured from the midpoint of a line of length $2b$ connecting the yaw axes of the imaging system. Each yaw axis intersects an optic axis at point A distance x from the image plane. The convergence angle α is measured from the forward direction (parallel to Q). N is the nodal point of the lens.

Considering the similar triangles in Fig. B-3, $\triangle MCN \sim \triangle SEN$, we see that $MC/CN + ES/EN$ or

$$\frac{q_0 + \Delta q - (f-x) \cos \alpha}{b - (f-x) \sin \alpha} = \frac{f \cos \alpha + s \sin \alpha}{f \sin \alpha - s \cos \alpha} \quad (B-2)$$

The range uncertainty in the direction of increasing q is denoted $+\Delta q$, shown as PQ. Thus from Eqs. (B-1 and B-2) we obtain

$$+\Delta q = \frac{bs - (f-x) s \sin \alpha}{\sin \alpha (f \sin \alpha - s \cos \alpha)} \quad (B-3)$$

Similarly a displacement in the image plane in the opposite sense, $-s$, gives the uncertainty in range in the $-Q$ direction from P.

$$-\Delta q = \frac{-bs + (f-x) s \sin \alpha}{\sin \alpha (f \sin \alpha + s \cos \alpha)} \quad (B-4)$$

The second term of the numerator is minimal when the yaw axis is through the point on the optic axis a distance $x = f$ in front of the image plane, then $(f-x) = 0$.

It can be shown that range uncertainty increases approximately with the square of the range.

Substitutions into Eqs. B-3 and B-4, noting that $\alpha = \tan^{-1} b/q_0$, using values of q_0 (from 0 to 3 meters at intervals of 0.5 meter), b and f from the text and Δs from Table 3-2, leads to the range uncertainty of camera assemblies D1 and E1 tabulated in Table B.1 and plotted in Figs. 3-5 and 3-11.

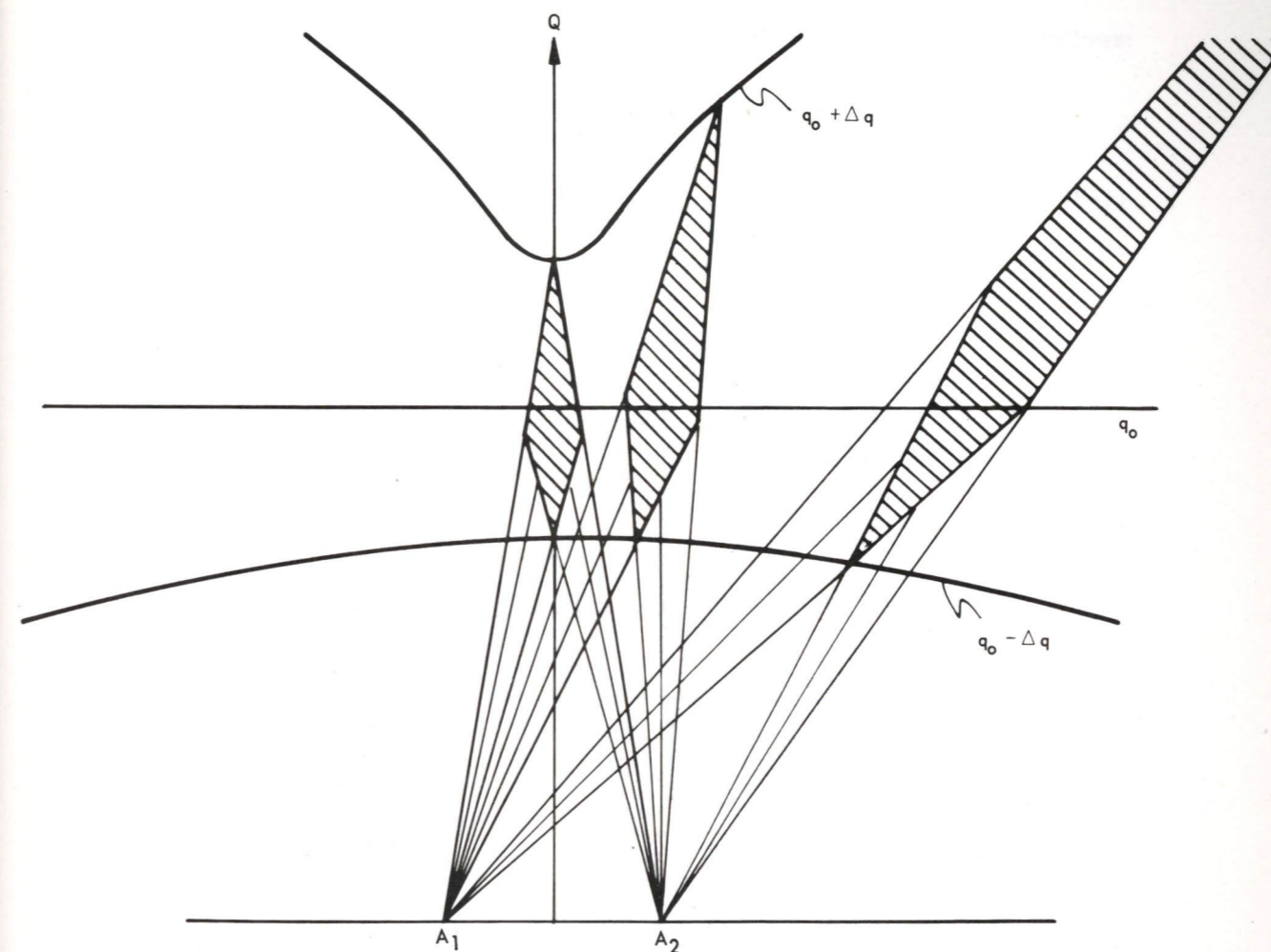
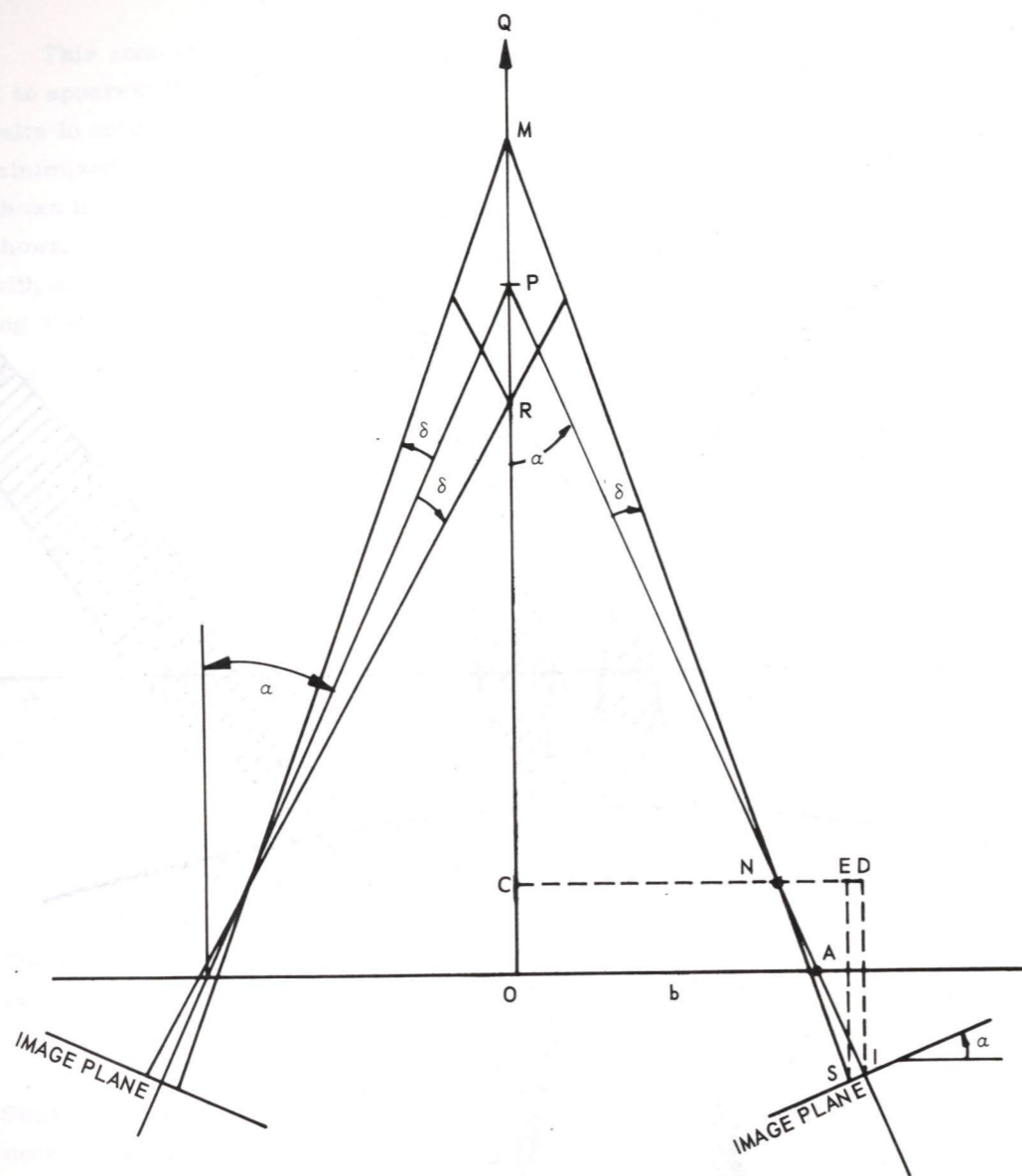


Fig. B-2. Quadrilaterals of uncertainty for points on the line q_0 .



SYMBOL DEFINITIONS

- | | |
|---|--|
| A = POSITION OF YAW AXIS OF CAMERA | IS = s, A DISPLACEMENT ON THE IMAGE PLANE EQUIVALENT TO ANGLE δ |
| AI = x, SEPARATION OF IMAGE PLANE AND YAW AXIS ALONG THE OPTICAL AXIS | PO = q |
| AO = b, ONE HALF OF SEPARATION OF YAW AXES | PM = $+\Delta q$ |
| I = POINT ON IMAGE PLANE PIERCED BY OPTIC AXIS | PR = $-\Delta q$ |
| IN = f, FOCAL LENGTH OF LENS SYSTEM | α = CONVERGENCE ANGLE |
| IP = OPTIC AXIS | δ = UNCERTAINTY OF ANGLE α |

Fig. B-3. Geometry of variable convergence optics for range finding.

B.3 Systems with Parallel Optic Axes

When the optic axes are held parallel to the Q axis, as in camera assembly C1, $\alpha = 0$, and an object at a point P on the Q axis will be represented by corresponding points S in the image planes. Uncertainties in the position angle $\pm \delta$ will produce equivalent displacements $\pm \Delta s$ in the image plane.

The range q of a point P, as seen in Fig. B-4, may be determined as follows:

Since $\Delta MOS \sim \Delta NAS$, $\frac{f}{s} = \frac{q}{b+s}$

$$q = \frac{f(b+s)}{s} \tag{B-5}$$

The range $q + \Delta q$ of a point M is

$$q + \Delta q = \frac{f[b+(s-\Delta s)]}{s-\Delta s} \tag{B-6}$$

Subtracting Eq. (B-6) from (B-5) we obtain the range uncertainty Δq due to uncertainty in a position S on the image plane (or equivalently an angular uncertainty δ in determining β):

$$\Delta q = \frac{fb \Delta s}{s(s-\Delta s)}$$

and from Eq. (B-5)

$$s = \frac{fb}{q-f}$$

so that

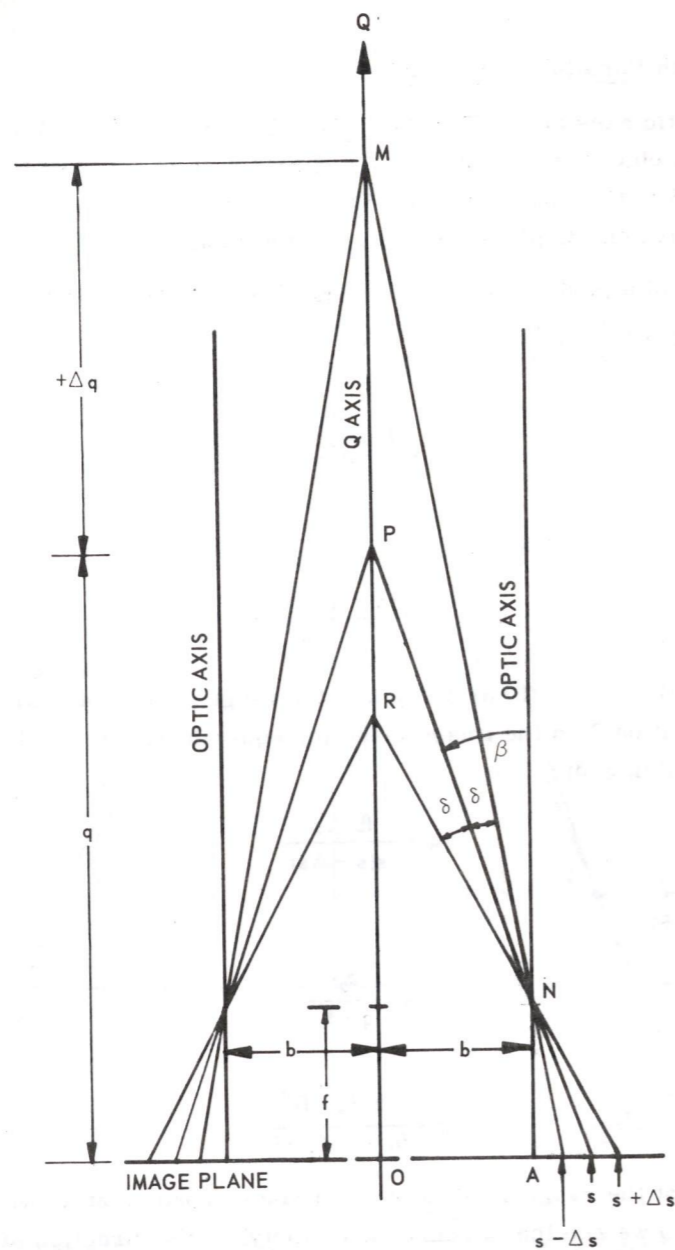
$$\Delta q = \frac{\Delta s(q-f)^2}{fb - \Delta s(q-f)} \tag{B-7}$$

Thus it is seen that the range uncertainty increases approximately with the square of the range. Also we see that a range uncertainty, in the direction of increasing range, corresponds to an uncertainty in s towards the optic axis.

Substitution into Eq. (B-7) of values of q (from 0 to 3 meters at intervals of 0.5 meter), b and f from the text and Δs from Table 3-2, leads to the range uncertainty of camera assembly C1 tabulated in Table B. 1 and plotted in Fig. 3-5.

B.4 Stereo Optics with Fixed Convergent Axes

In a system with fixed convergence the field of view of the lenses will in general be limited. As shown in Fig. B-5, there will be some near point NP on



SYMBOL DEFINITIONS

- q = RANGE OF A POINT P FROM IMAGE PLANE
- A = POINT WHERE OPTIC AXIS PIERCES IMAGE PLANE
- s = DISTANCE ON THE IMAGE PLANE FROM POINT A TO IMAGE OF POINT P
- Δs = UNCERTAINTY IN THE POSITION OF POINT s CORRESPONDING TO THE UNCERTAINTY ANGLE δ
- $+\Delta q = PM$
- $-\Delta q = PR$

Fig. B-4. Geometry of fixed optics for range finding.

Table B-1. Range and range uncertainty for three camera-computer chains (meters).

q_0	C1-1		D1-3		E1 ₁ -3		E1 ₂ -3	
	$+\Delta q$	$-\Delta q$	$+\Delta q$	$-\Delta q$	$+\Delta q$	$-\Delta q$	$+\Delta q$	$-\Delta q$
0.5	.0417	.0366	.0032	.0031	.0048	.0048	.0011	.0011
1.0	.118	.0948	.0137	.0133	.0243	.0231	.0046	.0045
1.5	.384	.2041	.0320	.0306	.0553	.0515	.0104	.0102
2.0	.540	.350	.0578	.0548	.0997	.0908	.0209	.0205
2.5	.898	.494	.0922	.0845	.1321	.1195	.0288	.0280
3.0	1.431	.734	.1346	.1235	.1813	.1617	.0377	.0367
5.0	9.16	1.96	.3743	.3440	.6741	.5901	.1321	.1275

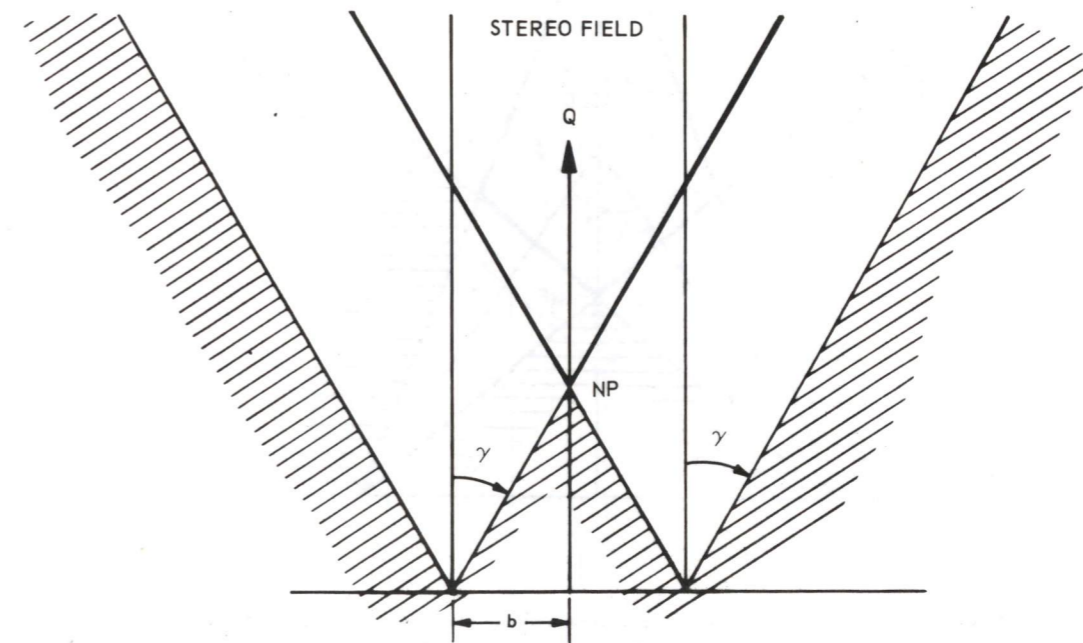


Fig. B-5. Field of view of parallel optical systems. Shaded area is outside view of either system, heavy line bounds area viewed by both.

the Q axis which is the minimum range at which a stereo image can occur. Thus there is a minimum distance at which the range of an object can be determined. This is at the point

$$q_{NP} = \frac{b}{\tan \gamma}$$

where γ is the half angle of the visible field of the lens system. When the optic axes are not parallel ($\alpha \neq 0$) but have fixed convergence there may also be a maximum point at which stereo images can be obtained, FP in Fig. B-6.

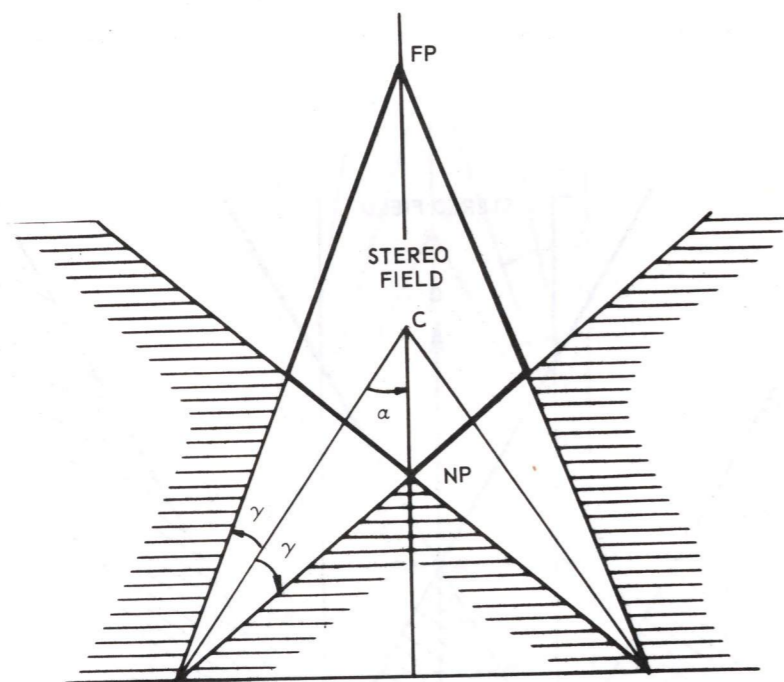


Fig. B-6. Fields of view of optical systems whose axes converge at C. Since $(\alpha + \gamma) < 90^\circ$ there is a point FP which is the maximum point for a stereo image. Area outside view of either system is shaded. Stereo field is bounded by heavy line.

APPENDIX C

OTHER STEREO CAMERA CONFIGURATIONS

by
Joseph Convers

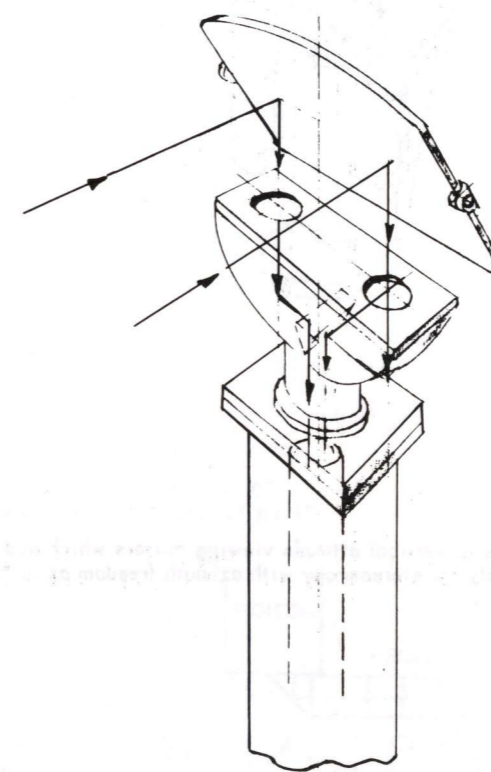


Fig. C-1. Single vidicon camera in vertical attitude receiving information from a single nodding mirror and giving 1/2 of surface to each stereo axis.

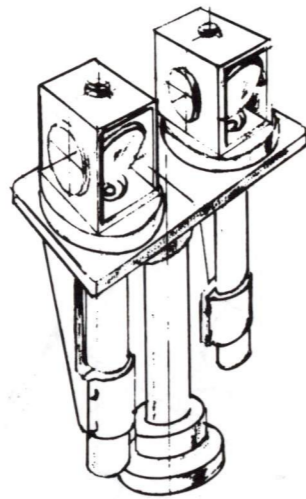


Fig. C-2. Two vidicon cameras in vertical attitude viewing mirrors which nod together and rotate individually for stereoscopy with azimuth freedom as in Figs. 3-6 and 3-10.

APPENDIX D

PROPOSED OPTICAL TRAIN FOR CAMERA E1

by

Robert Magee

The proposed optical train (Fig. D-1) is a dual system combining a wide-field camera with a high resolution but smaller-field camera. Both systems share the same objective lens by means of a beam splitting element. The physical size of the composite focal plane fits within a vidicon tube format size of approximately 0.5 inches by 0.5 inches.

The wide angle objective lens consists of a scaled version of the Minolta MC-Rokkor PF lens. This lens is designed for a speed of $f/1.4$ and is a form of the Zeiss Biotar. The focal length has been reduced from 58 mm in the Minolta design to 40 mm for Camera E1. One particular advantage of the lens is its relatively long back focal length. This lens-to-focal-plane distance is necessary in the single lens reflex camera with which the Rokkor PF is used. Since the optical arrangement of Camera E1 poses design problems similar to those encountered with a single lens reflex, a logical choice for investigation was the form of lens used in the reflex camera.

WIDE ANGLE FIELD IS 20° BY 10°
HIGH RESOLUTION FIELD IS 2° BY 1°

DRAWING: TO SCALE

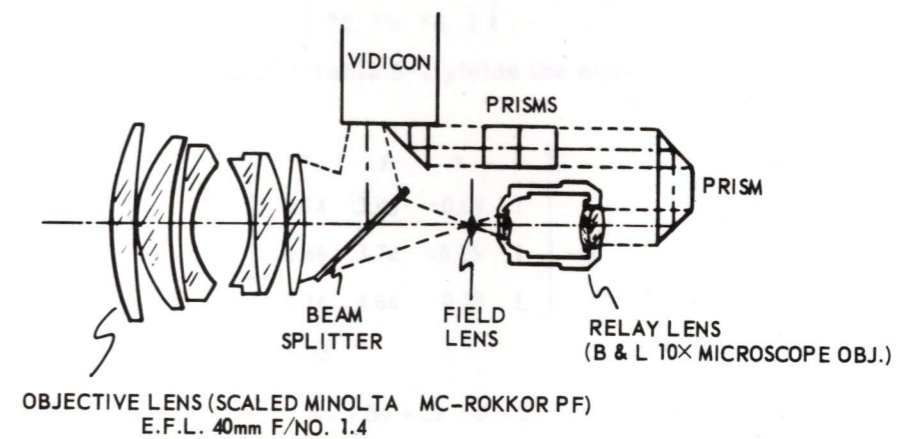


Fig. D-1. Proposed optical train for camera E1.

Light rays forming the wide angle image at the vidicon reflect from a beam splitter located to the right of the objective lens. Most of the light, however, is transmitted through the beam splitter and imaged at a field lens. A small central portion of this image is then relayed to the vidicon by means of a ten power microscope lens. This produces a ten to one scale relation between the two field images at the vidicon. The numerical aperture of the microscope lens is not adequate to handle all of the rays emanating from the $f/1.4$ objective lens and it appears that the latter should be stopped down to $f/2$. This lens would probably be used at $f/2$ for the high resolution field in any case because of the slight improvement in image quality.

A number of right angle prisms and mirrors are then used to conduct the light from relay lens to vidicon. The composite field of view at the vidicon then appears similar to that shown in Fig. 3-8. It is important to keep vidicon signal levels from both fields of view roughly equivalent. To accomplish this, the beam splitter is left uncoated. A very thin, or pellicle, form of beam splitter appears advisable to minimize reflection interference from the second beam-splitter surface.

APPENDIX E

CALCULATION OF THE DIRECTION COMPONENTS
OF THE NORMAL TO TRIANGLE FGJ AND THE
ANGLE BETWEEN THIS NORMAL AND SUNLIGHT.

by
Louis Sutro

E.1 Location of Points and Triangles

Figure E-1 is a plan view of the robot aimed at azimuth 43° and viewing the mound pictured in Figs. 2-3 and 5-1. Two reference lines have been drawn in the north and east directions from which measurements were made to determine the relative latitude and longitude of points F, G and J in Table 5-1. Figure E-2 is a side view employed by the artist to generate the perspective projections in Fig. 5-1. It is presented in lieu of a stereoscopic view which could communicate the spatial relations of all of these figures.

E.2 Calculation of the Direction Components of the Normal to Triangle FGJ

The equation of the plane through any three points (x_1, y_1, z_1) , (x_2, y_2, z_2) and (x_3, y_3, z_3) is, in determinant form:

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0 \quad (\text{E-1})$$

Substitution of the data of Table 5-1 yields the equation of the plane through points F, J, and K:

$$\begin{vmatrix} x & y & z & 1 \\ 0.14 & 3.45 & -0.09 & 1 \\ -0.46 & 3.74 & -0.25 & 1 \\ -0.14 & 4.64 & 0.28 & 1 \end{vmatrix} = 0 \quad (\text{E-2})$$

This equation is of the form

$$Ax + By + Cz + D = 0 \quad (\text{E-3})$$

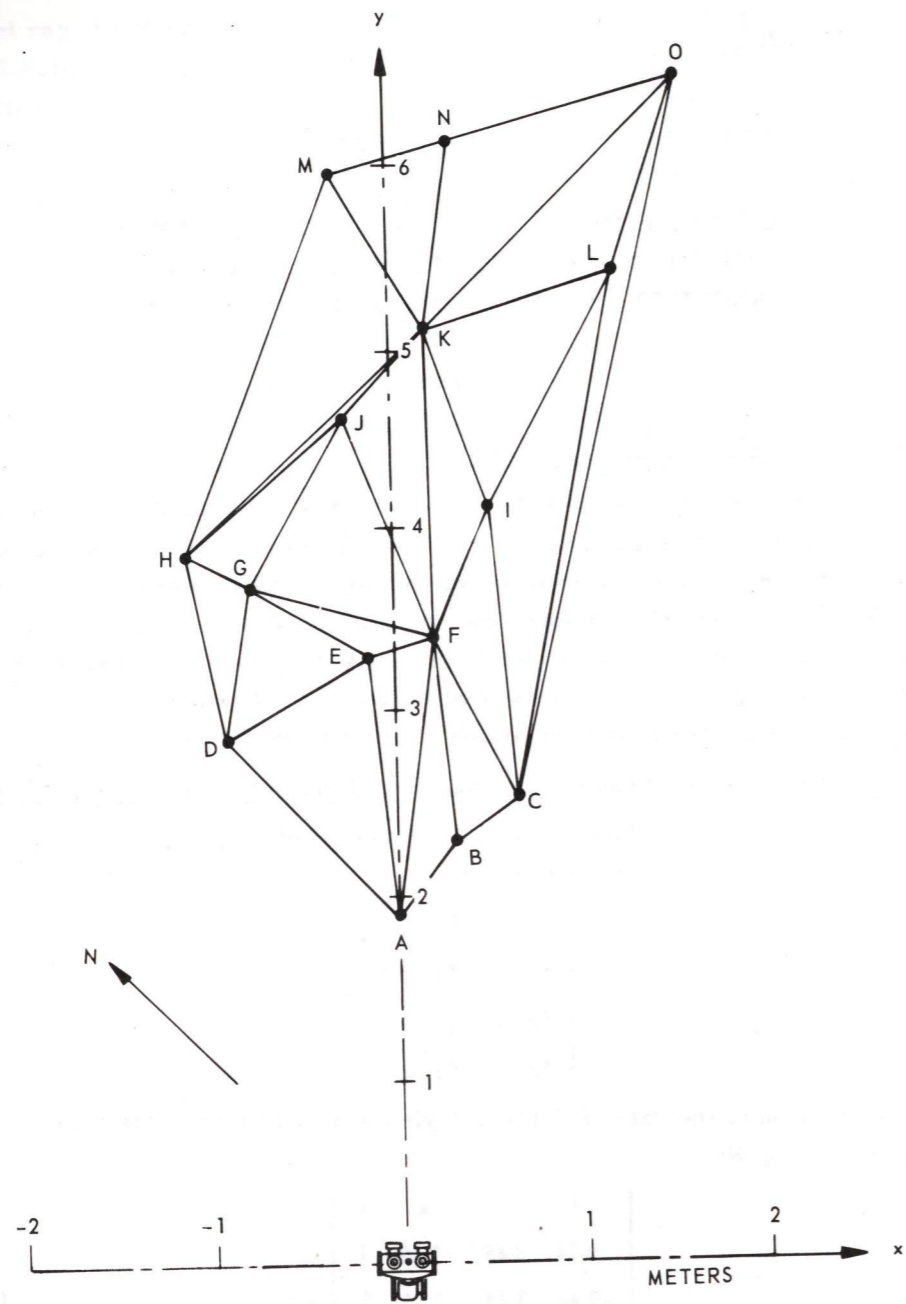


Fig. E-1. Plan view of robot and the mound shown near the center of Fig. 2-3.

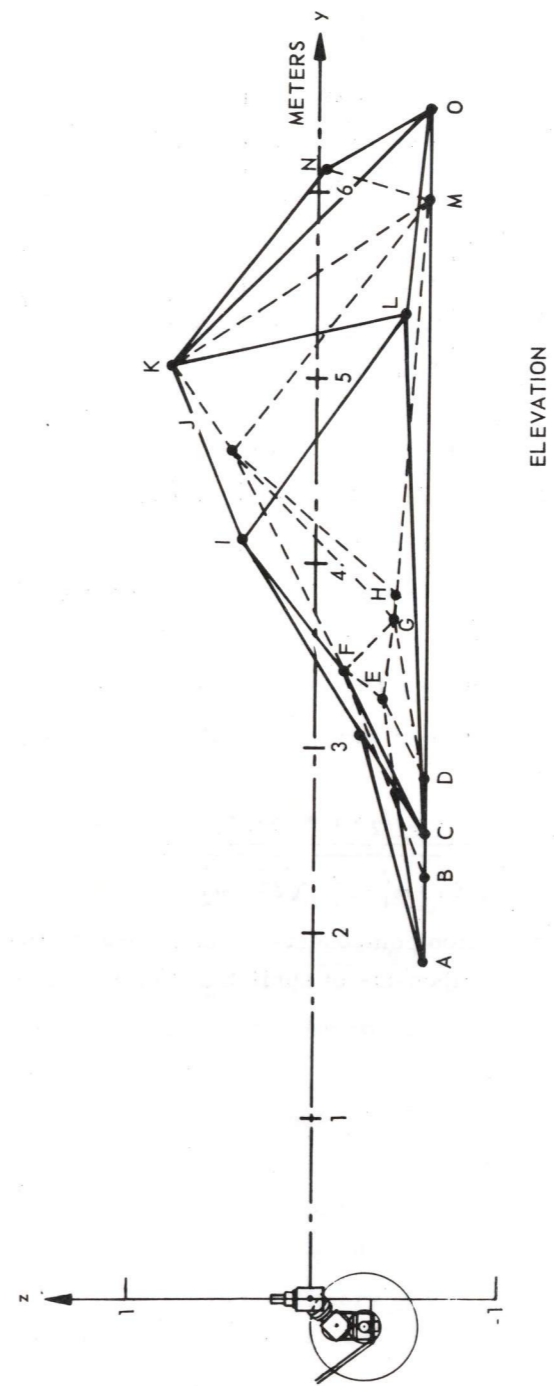


Fig. E-2. Side view corresponding to the plan view of Fig. E-1.

where

$$A = \begin{vmatrix} 3.45 & -0.09 & 1 \\ 3.74 & -0.25 & 1 \\ 4.64 & 0.28 & 1 \end{vmatrix} = -0.2977 \quad (E-4)$$

$$B = - \begin{vmatrix} 0.14 & -0.09 & 1 \\ -0.46 & -0.25 & 1 \\ -0.14 & 0.28 & 1 \end{vmatrix} = 0.2668 \quad (E-5)$$

$$C = \begin{vmatrix} 0.14 & 3.45 & 1 \\ -0.46 & 3.74 & 1 \\ -0.14 & 4.64 & 1 \end{vmatrix} = -0.6328 \quad (E-6)$$

The components of a vector perpendicular to the above plane are proportional to A, B, and C.

E. 3 Calculation of the Angle Between This Normal and Sunlight

The angle between the direction of sunlight and the normal to the plane is computed from the equation

$$\cos \theta = \pm \frac{l_1 l_2 + m_1 m_2 + n_1 n_2}{\sqrt{l_1^2 + m_1^2 + n_1^2} \sqrt{l_2^2 + m_2^2 + n_2^2}} \quad (E-7)$$

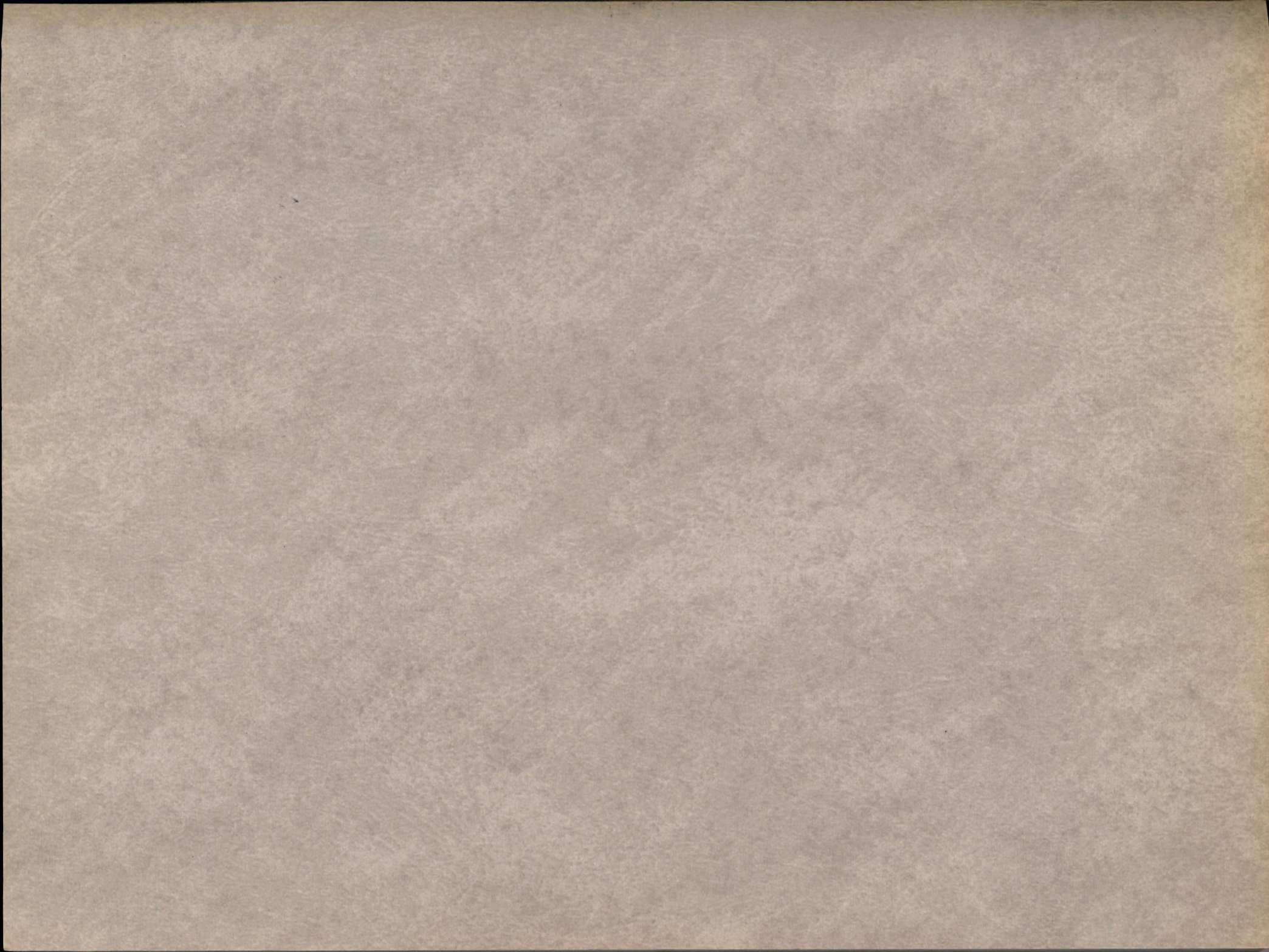
where l_1, m_1, n_1 are the direction components of the normal to the surface and l_2, m_2, n_2 are the direction components of sunlight. $\cos \theta = 0.71$.

REFERENCES

1. Sutro, L. L., Information Processing and Data Compression for Exobiology Missions, R-545, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge 39, Massachusetts, June, 1966.
2. Sutro, L. L., et al., Sensory, Decision and Control Systems, April to July 1966, R-548, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge 39, Massachusetts, November, 1966.
3. Sutro, L. L. (June 1966), op. cit. pp. 21-28.
4. Kilmer, W., Blum, J. and Peterson, D., The Decision Subsystem, Section 5 of Ref. 2, pp. 19-29.
5. Gregory, R. L., Eye and Brain, World University Library, McGraw Hill Book Co., New York, 1966, pp. 50-59.
6. Tompkins, D. N., Lunar Surface Panoramic Photography, Philco-Ford Corp., Newport Beach, California, Presented to Convention of American Society of Photogrammetry, Dayton, Ohio, September 23, 1965.
7. Tompkins, D. N., The Mars Facsimile Camera, Philco-Ford Corp., Newport Beach, California, Presented to American Institute of Aeronautics and Astronautics and American Astronautical Society, Baltimore, Maryland, March 28, 1966.
8. Sutro, L. L., et al. (Nov., 1966), op. cit., pp. 14-18.
9. Catchpole, R. J., Detection of Motion and Luminance Variation Using Vidicon Camera, T-470, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge 39, Massachusetts, September, 1966.
10. Beers, Y., Introduction to the Theory of Error, Addison-Wesley Publishing Co., Reading, Massachusetts, 1957.
11. Graham, D. N., Image Transmission by Two-Dimensional Contour Coding, Proc. IEEE, March, 1967, p. 336.

12. Huang, T. S. and Tretiak, O. J., Research in Picture Processing, in Optical and Electro-Optical Processing, The M.I.T. Press, Cambridge, Massachusetts, 1965.
13. Vander Lugt, A., Potz, F. B. and Klooster, A., Jr., Character Reading by Optical Spatial Filtering, The M.I.T. Press, Cambridge, Massachusetts, 1965.
14. Holmes, W. S., Babcock, T. R., Richmond, G. E., Pownall, L. A. and Vorie, G. C., Optical-Electronic Spatial Filtering for Pattern Recognition. The M.I.T. Press, Cambridge, Massachusetts, 1965.
15. Kulikowski, J. J., Adaptive Visual Signal Preprocessor with a Finite Number of States, IEEE Trans. on Systems Science and Cybernetics, December, 1966, p. 96.
16. Sutro, L. L., editor, Advanced Sensor and Control Systems Studies, R-519, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, January 1966, pp. 16-36.
17. Moreno-Diaz, R., An Analytical Model of the Group 2 Ganglion Cell in the Frog's Retina, E-1858, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1965.
18. IES Lighting Handbook, Fourth Edition, published by the Illuminating Engineering Society, 345 East 47th Street, New York, 1966.
19. Moon, P., The Scientific Basis of Illumination Engineering. McGraw Hill Book Co., New York, 1936, and Dover Publications, New York, 1961.
20. Sutro, L. L., A Model of Visual Space, Biological Prototypes and Synthetic Systems, Volume 1, Plenum Press, New York, 1962.
21. Moreno-Diaz, R., Conceptual Model of the Frog's Tectum, Section 6 of Ref. 2, pp. 30-40.
22. Moreno-Diaz, R., Contrast Detectors, Quarterly Progress Report No. 85, Research Laboratory for Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, April 15, 1967, pp. 321-324.
23. Barlow, H. B., Three Points about Lateral Inhibition, Sensory Communications, E. Rosenblith, editor, M.I.T. Press, Cambridge, Massachusetts, 1961, pp. 782-786.
24. Thorpe, W. H., Learning and Instinct in Animals, Methuen and Co., Ltd., London, 1963, p. 55.

25. Neurosciences Research Program Bulletin, vol. 4, pp. 105-233, Simple Systems for the Study of Learning Mechanisms, T. H. Bullock, chairman, 280 Newton St., Brookline, Massachusetts, November 30, 1966.
26. Neurosciences Research Program Bulletin, vol. 4, pp. 235-347, Brain Mechanisms in Conditioning and Learning, R. B. Livingston, chairman, 280 Newton St., Brookline, Massachusetts, Dec. 30, 1966.
27. Neurosciences Research Program Bulletin, vol. 4, pp. 235-347, op. cit., p. 245.
28. Neurosciences Research Program Bulletin, vol. 4, pp. 105-233, op. cit., pp. 177-178.
29. Thorpe, W. H., op. cit., p. 61.
30. Doyle, R. J., The Electrostatic Vidicon and Methods of Evaluation, 1962 IRE International Convention Record, American Institute of Electrical and Electronic Engineers, New York.
31. Light Measurement and Control, General Electric Large Lamp Department, Nela Park, Cleveland, Ohio, p. 20.



R-565
DEVELOPMENT OF VISUAL, CONTACT
AND DECISION SUBSYSTEMS FOR A MARS ROVER
July 1966 to January 1967

by

Louis L. Suto Roberto Moreno-Diaz
William L. Kilmer Warren S. McCulloch, et al.

July 1967

INSTRUMENTATION
LABORATORY ●

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Cambridge 39, Mass.

R-565

DEVELOPMENT OF VISUAL, CONTACT
AND DECISION SUBSYSTEMS
FOR A MARS ROVER

July 1966 to January 1967

by

Louis L. Sutro	Joseph J. Convers
William L. Kilmer	David Lampert
Roberto Moreno-Diaz	Robert J. Magee
Warren S. McCulloch	Charles D. Sigwart
James G. Bever	David G. Tweed
Jay Blum	

July 1967

A report of work performed from
July 16, 1966 to January 15, 1967 under
NASA Contract NSR 22-009-138,
and from May 1 to December 31, 1966
under NASA Grant NGR 22-009-140.

INSTRUMENTATION LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

*Prepared for publication by Jackson & Moreland
Division of United Engineers & Constructors, Inc.*

Approved


Professor of Aeronautical
and Astronautical Engineering

Approved


Associate Director

Approved


Deputy Director

ACKNOWLEDGEMENT

Section 1, part of Section 2, Sections 3, 4 and 5, part of 7 and Section 8 were prepared under the auspices of DSR Project 55-257, sponsored by the National Aeronautics and Space Administration, Headquarters, Office of Space Science and Applications, through Contract NSR 22-009-138.

Part of Section 2, Section 6, part of Section 7 and Section 9 were prepared under the auspices of DSR Project 76370, sponsored by the National Aeronautics and Space Administration, Electronics Research Center, Cambridge, Massachusetts, through Grant NSR 22-009-140.

All of the authors are either members of, or consultants to, the Longspur Section of the Instrumentation Laboratory. Two additional consultants are

Dr. Carl Sagan of Harvard University and the Smithsonian Astrophysical Observatory

Prof. Calvin Burnett of the Massachusetts College of Art who drew Figs. 2-3, 5-1, 5-2, 5-8, E-1 and E-2.

Further acknowledgement is given at the end of Section 5 to those who assisted in the preparation of that section.

TABLE OF CONTENTS

PART I DEVELOPMENT

<u>Section</u>		<u>Page</u>
1	INTRODUCTION	1
2	SUBSYSTEMS AND THEIR COORDINATION	3
	2.1 General	3
	2.2 Robot Subsystems	3
	2.3 The LOOK Mode of Operation	6
	2.4 Mounting a Stereo Pair of Cameras on a Vehicle	6
	2.5 Contact Subsystem	8
	2.6 Coordination of the Subsystems	12
	2.7 Digital Computers	12
3	CAMERA-COMPUTER CHAINS	14
	3.1 Objectives	14
	3.2 Camera-Computer Chain A	15
	3.3 Camera-Computer Chain B	17
	3.4 Designation of Stereo Optics and Camera Electronics	17
	3.5 Type 1 Camera Electronics	18
	3.6 Uncertainties Due to the Type 1 Camera Electronics	20
	3.7 Optics of Camera-and-Stereo-Attachment Assembly C1	20
	3.8 Anticipated Ranging Ability of Camera-Computer Chain C1-1	23
	3.9 Camera-and-Gimbal Assembly D1	25
	3.10 Anticipated Ranging Ability of Camera-Computer Chain D1-3	26
	3.11 Camera-and-Gimbal-Assembly E1	28
4	VISUAL DATA PROCESSING	32
	4.1 General	32
	4.2 Characteristics of the Visual Processing Program	33
	4.3 Initial Approach	33
	4.4 A Second Approach	35

<u>Section</u>		<u>Page</u>
4.5	The Present Approach	36
4.6	Range Finding Planned for Camera-Computer Chains C and D.	36
5	TOWARD A METHOD OF MODELLING AN ENVIRONMENT . .	39
5.1	General.	39
5.2	Proposed Method of Acquiring a Point Model	39
5.3	An Example.	40
5.4	Interrogation of a Point Model.	41
5.5	Interrogating a Point Model in the Search for Evidence of Life	43
5.6	Mapping.	43
5.7	Representation by Points and Reflectances	44
5.8	Computation of Reflectance	44
5.9	Computation of the Diffuse Reflectance of Δ FGJ	48
5.10	Interrogation of a Point-and-Reflectance Model.	48
5.11	The Earth Station.	49
5.12	Acknowledgement.	49
6	ADDITION OF A LEARNING ABILITY TO THE DECISION-MAKING SUBSYSTEM	51
6.1	Introduction.	51
6.2	Requirements.	51
6.3	Nature of the Problem	52
6.4	Approach to the Problem.	52
6.5	Implementation by Operations on P_{AB} Vectors	52
6.6	Structure of the Model	54

PART II

RESEARCH INTO THE NATURE OF ANIMAL SENSORY,
DECISION AND CONTROL SYSTEMS

7	ANIMAL COMPUTATION.	57
7.1	Background to a Computer Model of a Vertebrate Brain	57
7.2	A Possible Engineering Equivalent	60
7.3	Memory in Biological Computers	60
8	VISUAL PROCESSING IN ANIMALS	62
8.1	Development of a Computational Model	62
8.2	Generalization of Interaction Coefficients.	64
8.3	Examples of Lateral Interaction.	65

<u>Section</u>		<u>Page</u>
9	LEARNING IN ANIMALS	68
9.1	General.	68
9.2	Classes of Response	68
9.3	Classical Conditioning	69
9.4	Habituation	69
9.5	Discriminative Learning.	70
9.6	Instrumental Conditioning	70
10	SUMMARY AND CONCLUSION.	71
<u>Appendix</u>		
A	MEASUREMENT OF VIDICON RESOLUTION	73
B	RANGING AND RANGE ERROR	75
C	OTHER STEREO CAMERA CONFIGURATIONS	83
D	PROPOSED OPTICAL TRAIN FOR CAMERA E1	85
E	CALCULATION OF THE DIRECTION COMPONENTS OF THE NORMAL TO THE TRIANGLE FGJ AND THE ANGLE BETWEEN THIS NORMAL AND SUNLIGHT.	87
LIST OF REFERENCES		91

PART I

DEVELOPMENT

SECTION 1
INTRODUCTION

by

Louis Sutro

The first of the two efforts described in this report is supported by the Bioscience Division of the Office of Space Sciences and Applications, NASA, and is concerned with communicating pictorial data from a remote location, such as Mars, to Earth. The second, supported by the Electronics Research Center of NASA, is to develop methods of making decisions at remote locations. Both are instrumentation efforts applicable to a Mars rover and both use biology as a source of ideas.

Earlier reports^(1, 2) indicated that the processing of television pictures at a remote location was desirable, and could be facilitated by employing a "rover"; that is, a pair of stereo television cameras on a mobile base. Since further on-board processing was required to position the cameras, the concept evolved of an assembly - to be called "robot" - of stereo cameras, vehicle, and computer. Included in the concept are inputs to the computer from senses other than the visual, namely, accelerometers to indicate the inclination of the cameras and their motion, and a tactile-kinesthetic sense of both the posture of the robot and its contact with the ground.

The word "computer" has been used in two ways in this and earlier reports. As illustrated above, it is used first to describe all equipment employed by the robot for processing or computation. For testing stationary equipment this computer will be the general-purpose machine described in subsection 2.7; for testing mobile equipment it will be special-purpose and transportable. This report presents those specifications presently available for the mobile system, while more detailed specifications are being prepared that will lead to size, weight and power estimates necessary to judge practicality.

The second use of the word "computer" is to describe the kinds of computation performed by animal nervous systems. Section 7 introduces this second use of the word.

Each of the two efforts is being approached from the points of view of both development and research. Although the first six sections of the report, in general describe development and Sections 7 through 9 describe research into the nature of animal sensory, decision and control systems, the line between research and development cannot be sharply drawn.

The sections on development consider first, in Section 2, the tying together of the subsystems. Sections 3, 4 and 5 describe three aspects of the visual subsystem: camera-computer chains, processing and representation of the environment. Section 6 describes additions being made to the decision subsystem.

The first of the research sections - Section 7 - describes Dr. Warren McCulloch's concept of the nervous system as interconnected computers. Section 8 develops mathematically a concept of visual processing, known as lateral interaction, which appears basic to the functioning of higher animal retinas. Finally, Section 9 describes the theory behind animal learning, a body of knowledge which must be understood if an engineering system such as that in Section 2 is to adapt to changes in its environment. Section 10 is a summary and conclusion.

SECTION 2

SUBSYSTEMS AND THEIR COORDINATION

by

Louis Sutro, William Kilmer and Joseph Convers

2.1 General

The system whose parts are described here includes the robot, means of transmitting pictorial and scientific data to Earth, and equipment for decoding it and displaying it to an Earth operator. Subsystems of the robot are outlined in subsection 2.2, after which follow a description of the LOOK mode of operation and proposed means of mounting the visual on the mobility subsystem. Coordination of subsystems is described in principle in subsection 2.6 and simulation within a general purpose computer in subsection 2.7.

2.2 Robot Subsystems

The visual subsystem includes a camera assembly of the kind described in Section 3, processing of the kind described in Section 4, and the making and interrogating of a model of visual and tactile space described in Section 5.

Equipment intended to contact the environment either to report on it or to affect it is called the contact subsystem - examples are the hand and arm assemblies shown at the left center of Fig. 2-1. Since vision, in many ways, is a distant method of feeling the shape of things, coordination is required of the visual and contact subsystems. This appears achievable either by a matching of models of the kind described in Section 5 or by the learning mechanisms described in Section 6.

The mobility subsystem considered during this report period is the vehicle developed by General Motors for a landing on the moon and now being redesigned for a landing on Mars. Following their practice, this subsystem includes the basic structural elements, wheels, wheel drives and steering actuators as illustrated by Fig. 2-2.

The fourth component, the decision subsystem, determines the mode in which the robot operates. For initial tests, only the following four modes will be possible:

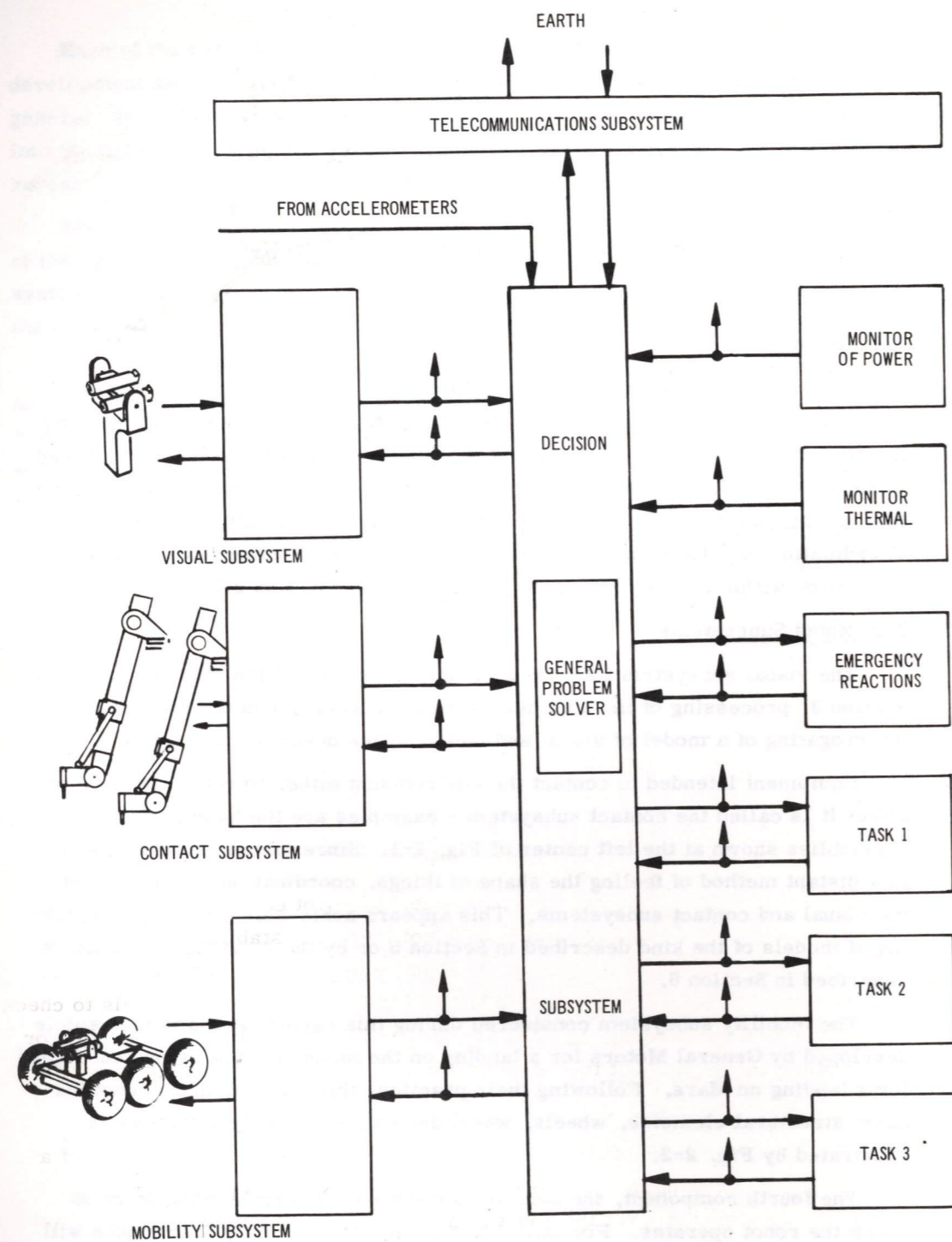


Fig. 2-1. Subsystems of the robot proposed for a Mars landing.

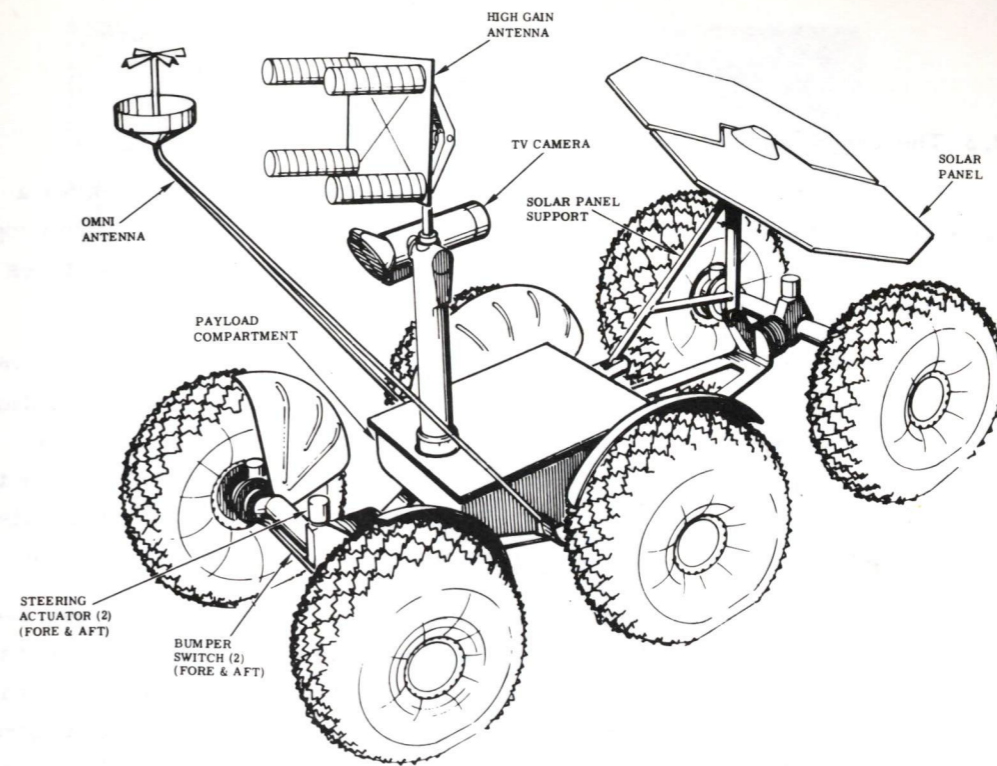


Fig. 2-2. Typical vehicle developed for lunar and planetary exploration by the Defense Research Laboratory, AC Electronics Division, General Motors. Length is 72 in., width 36 in., height of wheels 18 in.

- Look
- Move
- Rest
- Maintenance

These are mutually exclusive. For example, the robot cannot LOOK and MOVE at the same time because the cameras are not inertially stabilized during vehicle motion. The REST mode is planned for night, when lack of light makes looking and therefore movement difficult. MAINTENANCE could use the wheels to check for vehicular trouble or use the cameras either to check their own operation or check other equipment, but the intent will not be the same as the LOOK or MOVE modes.

The robot will have other modes of operation: one to try to get it out of a difficulty such as wedging a wheel between rocks, others to perform specified tasks or experiments. Yet the robot is not planned to be autonomous. The top command remains on earth.

The upward pointing arrows in Fig. 2-1 represent connections that can be made to the telecommunications subsystem, which in a test system being planned would be a microwave link to a nearby control station. A high-gain directional antenna and an omni-directional antenna, both intended for this purpose, are shown in Fig. 2-2.

2.3 The LOOK Mode of Operation

The LOOK mode of operation may take one of these forms: to look for a passage ahead for the robot to follow, to look for objects having specified properties of scientific interest, or to look at the scene or object to record pictures for transmission to Earth.

Figure 2-3 indicates how a scene might appear on Mars in two successive looks (the two are separated by the vertical line at the center of the illustration). At the center of each look is a high-resolution central field of the kind to be achieved by camera assembly E, described in subsection 3.10. Around it is the low-resolution peripheral field. In this illustration the ratio of the central field to the peripheral is made 1/5, whereas in camera E it is planned to be 1/10.

The LOOK mode begins with the levelling of the platform shown in Fig. 2-4. Possible next steps are the pitching of the cameras to scan the lowest line of the nearest rectangle; then turning of the cameras one position at a time, so that their point of convergence steps along a line the width of the rectangle. The illustration shows the cameras in the position where they receive images from the same point on the rock. If this image is an edge, as here, it can be considered a pointer in the direction to be explored next. Following such pointers, the cameras trace the edge of the rock. Then, converging further away, they can sweep the second rectangle to find the hole in the ground. Converging still further away, they can sweep the farthest rectangle to find the edge of the cliff. Since this process may be difficult to realize if images of the scene are crowded, other approaches are being studied.

Points as well as lines may define a surface. Experiments conducted at Lincoln Laboratory, MIT, indicate that white dots against a black background, located in both the left and right views of a stereoscopic picture, can perform the same function as a set of lines. Such dots could also be black against a light background or merely contrast with the adjoining area. Thus in the display on Earth a surface such as Fig. 2-4 could be defined by points without connecting lines.

2.4 Mounting a Stereo Pair of Cameras on a Vehicle

Figure 2-2 shows the proposed General Motors vehicle to which equipment is planned to be attached. A vehicle of similar but earlier design was delivered to Jet Propulsion Laboratory and is operating there.

Two approaches have been made to the mounting of a stereo pair of cameras on such a vehicle. One is to support the stereo pair, as in Fig. 2-5, on a boom, which can be raised to permit looking in the distance or lowered to permit looking immediately in front of the vehicle. The other approach is to mount the pair



Fig. 2-3. Supposed views of the surface to be explored, showing two high resolution looks surrounded by low resolution looks.

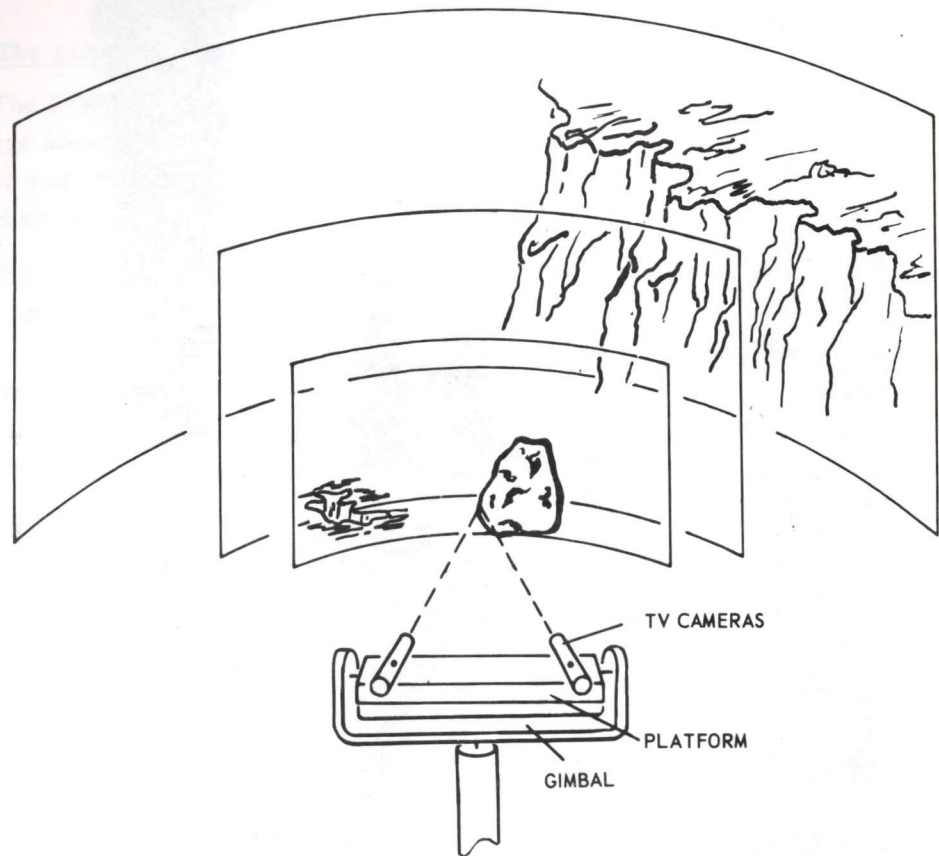


Fig. 2-4. Discrimination of objects by a stereoscopic pair of TV cameras and a computer.

as close to the ground as possible, as in Fig. 2-6. For both approaches it is desirable, after vehicle movement, to level the camera platform and then maintain the pitch axis of the cameras horizontal as they look up or down, sideward, or backward.

The first approach required more mechanical complexity than appeared wise at the start. The second placed the weight of the camera assembly too far forward. Figure 2-7 is a compromise, placing the camera assembly above the axle.

2.5 Contact Subsystem

If recognition of objects is to be performed as it is in animals, more than one sensory input is required. While information provided by a single sense is, in general, inadequate for recognition, the addition of a second sense generates what is necessary to validate the original evidence. With the integration of the contact and the visual subsystems, a reliable recognition capability should be achieved.

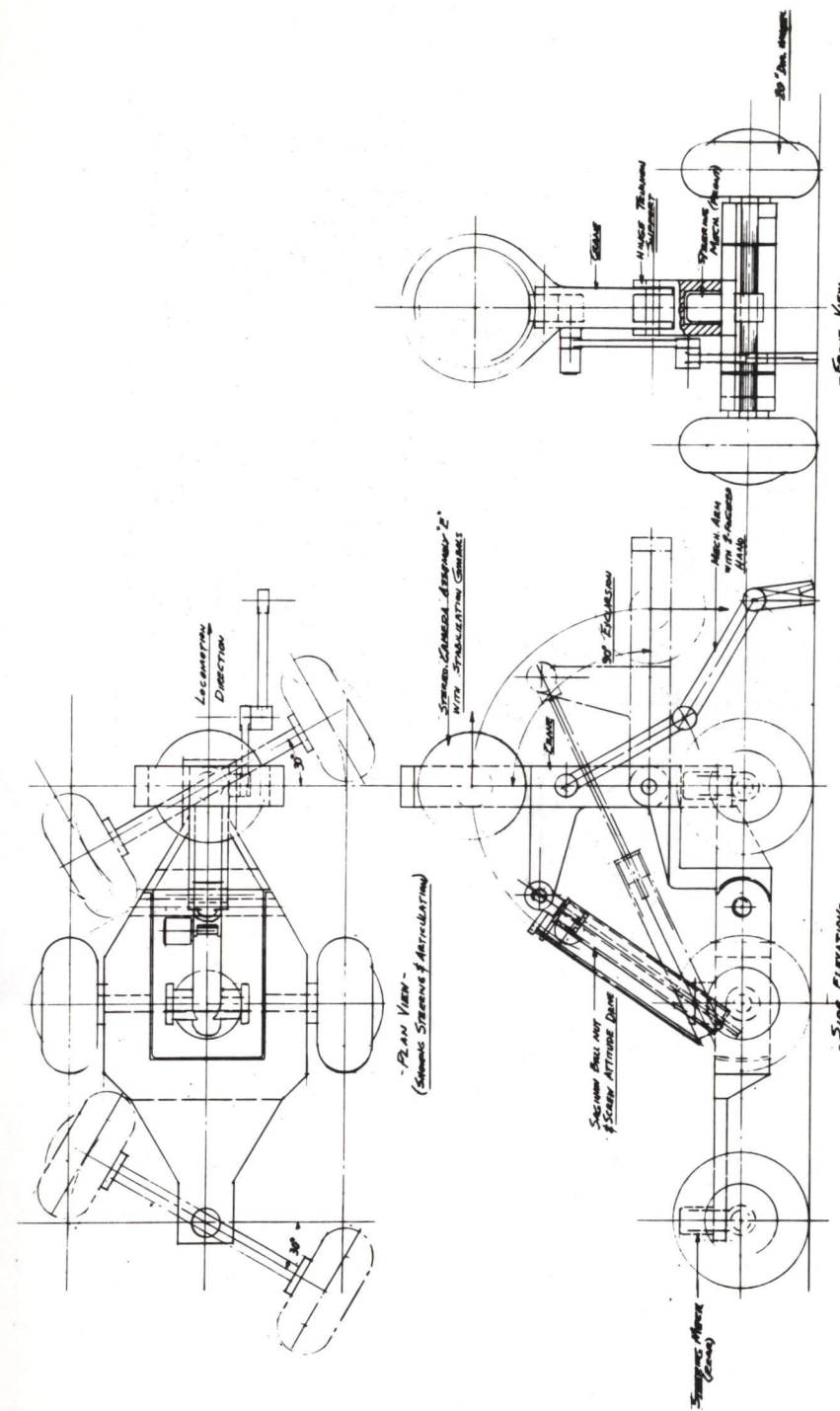


Fig. 2-5. Use of a boom (crane) to elevate stereo pair of cameras.

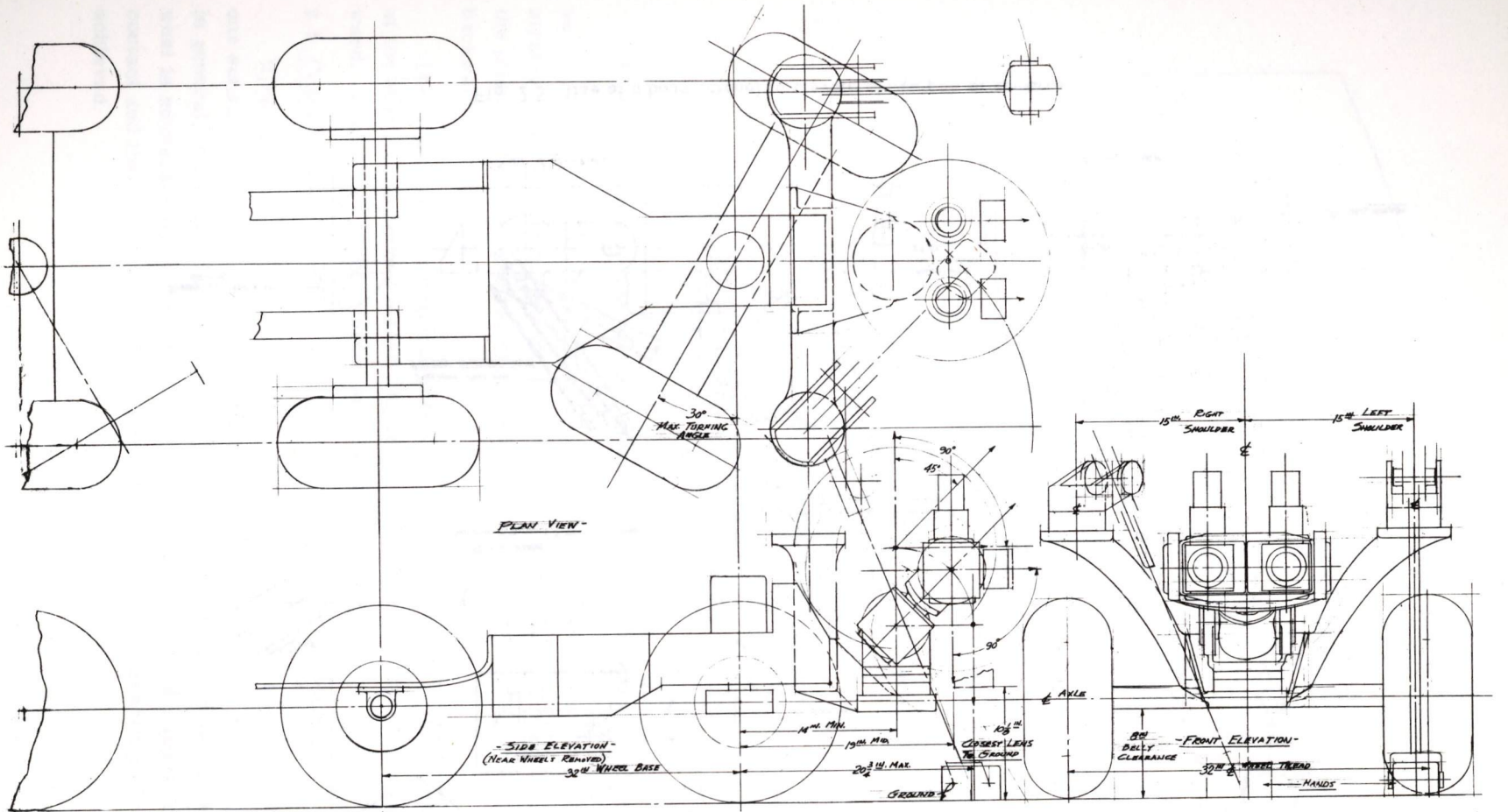


Fig. 2-6. Camera assembly E mounted in front of forward axle of General Motors six-wheeled vehicle.

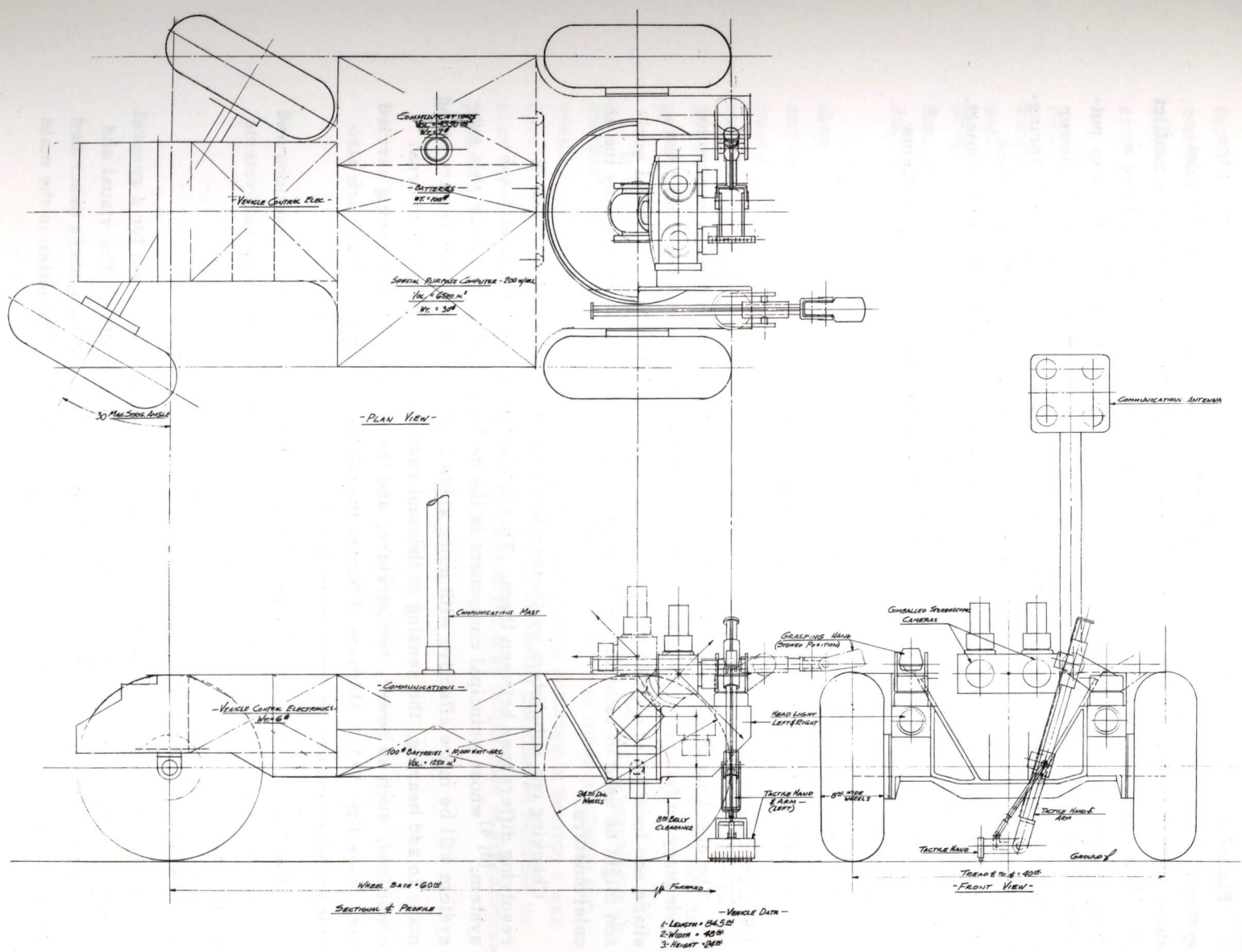


Fig. 2-7. Stereo pair of cameras mounted on front axle.

Figure 2-6 is the first attempt to include a hand-and-arm assembly in the system. Each hand is at the end of an arm which is supported through a joint above a front wheel. Two objections to this arrangement are the possible conflict of a front wheel with an arm and the mechanical difficulties of cantilevering a long arm with precision. The design of Fig. 2-7 lessens these difficulties by putting the wagon-type steering at the rear as in a ship, and the shoulders both lower and nearer the center. The hand, as presently designed, contains a line of spring-loaded pins with a shaft position encoder on each. From the line array can be determined the relative displacement of the pins and hence the contour of an object.

The following elements feed data to or receive data from other subsystems:

- a) Tactile input of the hands.
- b) Kinesthetic input of each joint.
- c) Motors which position the hand for touching.

2.6 Coordination of the Subsystems

The checking of one sensory subsystem by others should partially overcome inadequacy or failure. For example, the contact subsystem can reexamine the shape and location of objects detected by the visual subsystem. The visual horizon might be used to validate the inclination of the cameras as reported by the accelerometers.

Checking one subsystem by another would be of no use without means of resolving differences between them. This is the function of the decision subsystem,^(3, 4) whose principal component is the decision computer. Since this subsystem will be more effective with more kinds of sensory input, studies are being made to add hearing, the sensing of thermal radiation and sensors of internal conditions, such as power, temperature, and the progress of tasks being carried out. (See Fig. 2-1.) Methods of having the decision subsystem learn are also being studied and are reported in Section 6.

Thus Fig. 2-1 might be curled on itself so that the emergency reactions and tasks, shown at the right, relate to the visual, contact and mobility subsystems at the left.

2.7 Digital Computers

Each of the boxes in Fig. 2-1 is being developed as a program for a general purpose digital computer, although not for the same computer. The visual and contact subsystems are being developed in the PDP line of computers described below, the decision subsystem in the IBM 360 Model 75. The latter is the main

facility at the Instrumentation Laboratory and is provided with a very facile compiler - the M. I. T. Algebraic Compiler. However, pressure to use this machine is so great that attachment of external equipment such as cameras and arm-and-hands has not been permitted.

It is the intention of this effort to obtain first a combination of commercial machines in which all programs can be run and all equipment attached, then a single machine of equal facility but small and light enough to be carried on a vehicle such as that of Fig. 2-7. The combination of machines in which all programs can be run and all equipment attached is expected to be a Digital Equipment Corp. PDP-9 to be delivered to the Instrumentation Laboratory in July of this year, and a larger computer to be connected to it by telephone lines.

As presently planned, the larger computer would contain both the relatively slow formation of the model of the environment and its interrogation (Section 5) and the decision subsystem (Section 6). The PDP-9 would perform the faster interpretation of the output of the cameras (Section 4) and the commanding of their motion. The decision subsystem will continue to be developed in the IBM 360 Model 75 while a copy of it is programmed for the larger of the two tied-together computers.

In the interval between ordering and delivery of the PDP-9, a machine with similar instruction code, the PDP-7 in Project MAC, was made available for evening and off-hour daytime use. The programming reported in Section 4 was carried out in this machine. Finally, TX-0 is not being considered because its slow memory access rate is poorly matched to the high output rate of all cameras presently under development.

SECTION 3

CAMERA-COMPUTER CHAINS

by

Louis Sutro, David Tweed, Joseph Convers and Charles Sigwart

3.1 Objectives

While the primary task of the visual subsystem is to sense the environment, that is, to obtain pictures of the area surrounding the cameras, due to the time lag in communication and the low rate of transmission between the robot and Earth, it is also necessary that the robot construct a model of the environment, that is, an internal representation that can be interrogated to help in guidance and experimentation. As visual data flows from the cameras through successive transformations to the internal representation, it can be tapped at any point and transmitted to Earth. Thus, several choices will exist regarding not only what pictures are to be sent but also what will be their level of abstraction.

A prerequisite for dealing with visual data is a means of determining range, such as obtaining two views from two vantage points. Engineers call such an arrangement stereoscopic, while psychologists argue that stereoscopic (stereo) vision is not only range finding on one point at a time but determination of the relative range of every point in the entire field for which there is a disparity between the left and right views.⁽⁵⁾ Camera-computer chains for the purpose of finding the range of one point at a time are considered later in this section.

Beginning with subsection 3.2, the subjects considered are cameras A and B, both monocular. Then follow descriptions of three of the four types of stereo optics being considered for the visual subsystem, namely, the C, D and E. (A fourth type developed by Philco-Ford has been described elsewhere.^(6, 7)) This section then shows how the stereoscopic camera assembly and its electronics can be tied to a computer to interpret its output and direct it. An analysis of the errors inherent in each camera type is presented.

Since descriptions of hardware and software can be conveniently separated, this section will dwell on camera-computer chain A, operated in the summer of 1966, and chains C, D and E presently under development; while discussion of the programming to be employed in these chains is reserved for Section 4. A possible method of modelling the environment is presented in Section 5.

Nonetheless, the subject material of the three sections has been conceived as a single unit. For example, the window, employed for detailed examination of the image, is formed by counters that open or close gates at a window frame superimposed either on the scene or on the stored image. The counters will be either special-purpose electronics (hardware) or a routine within a general-purpose computer (software).

All of the cameras described in this section employ as their transducer the slow-scan vidicon (Type 1343-018) developed by the General Electrodynamics Corp. of Garland, Texas, for the Mariner and Surveyor spacecraft. It is being used to obtain the data reported here and may be replaced by solid-state transducer arrays as these become available.

3.2 Camera-Computer Chain A

Camera chain A was monocular; that is, it employed a single optical train. Built in 1964 and improved in the spring of 1966⁽⁸⁾ it was linked to the TX-0 computer in the summer of 1966 for experimentation in the detection of motion. (See Fig. 3-1.) The camera had been designed at Lincoln Laboratory, MIT, to be operated as fast as its output could be loaded on an analog tape (500,000 Hz). After loading, the tape was to be run at a slower speed, so that its contents could be converted from analog to digital form and stored on a second tape for computer processing.

In its use for the Instrumentation Laboratory, the camera was required to respond to four times as large a raster at the same resolution. Given the same dwell time per position in the raster, this would require that the video amplifier have four times as wide a passband as was required for the Lincoln Laboratory application.

In the actual test, due to the inavailability of high-speed analog tape, the camera was operated at the speed of digital tape, with the result that signals passed through the video amplifier at frequencies below the passband. Lowering

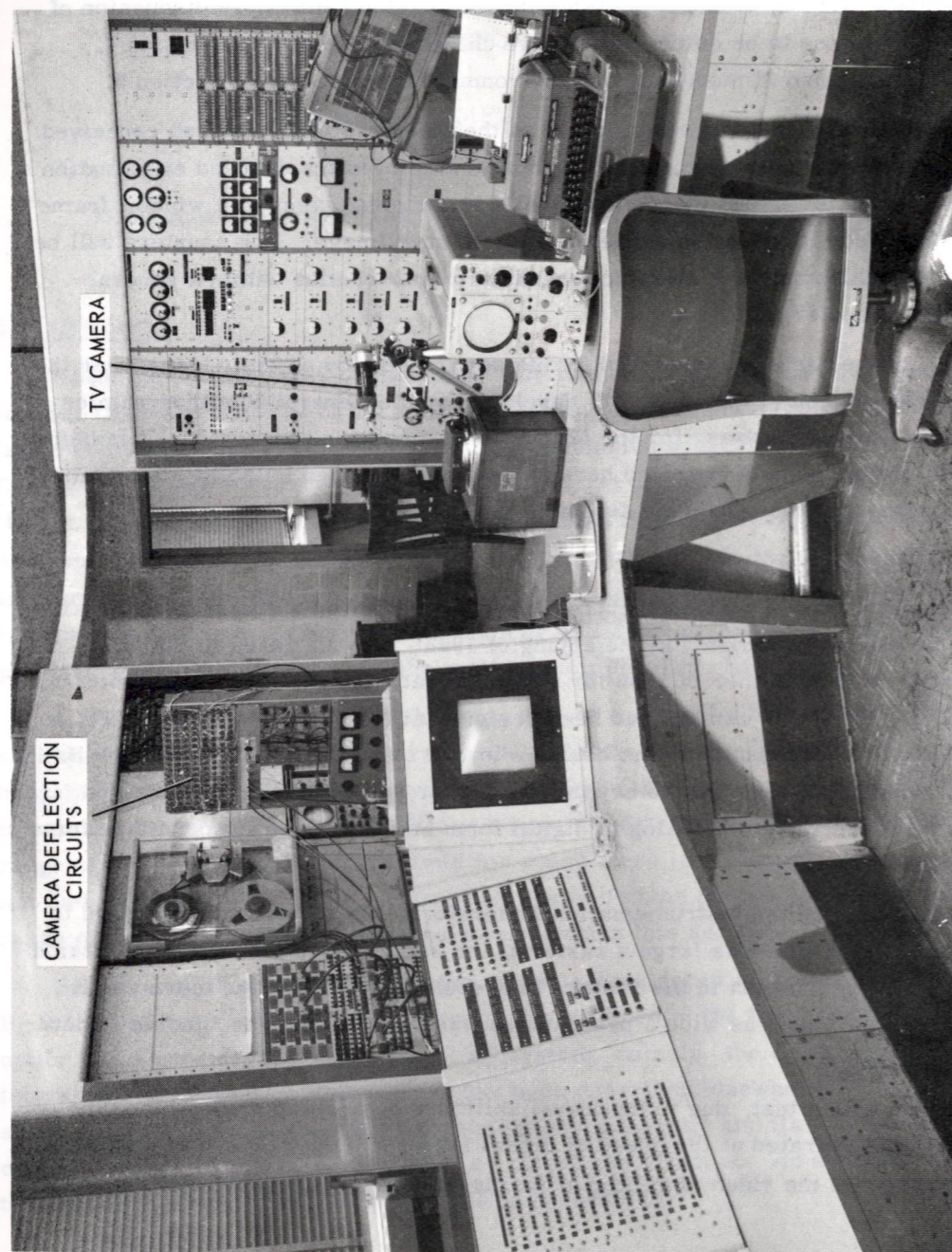


Fig. 3-1. Camera-computer chain A. The computer is the TX-0 in the Department of Electrical Engineering, MIT.

the bandpass through partial modification of the video amplifier could not prevent noise from appearing in the data, and construction of the amplifier on printed cards within a tubular case made further modification difficult. Since Camera C was already in development, further work on Camera A was dropped.

In spite of the difficulties camera-computer chain A was operated by a graduate student leading to a thesis on motion detection.⁽⁹⁾ The study consisted of recording in the TX-0 computer pairs of television frames, comparing corresponding positions in them and generating a "decision map" of changes in luminance. To lessen the effect of noise the computer was programmed to "reduce" the decision map by rejecting isolated changes.

Further experiments in motion detection have been deferred for the present. It is assumed that the method employed in camera-computer chain A could be employed also in succeeding designs.

3.3 Camera-Computer Chain B

Chain B was an assembly to implement motion detection in hardware. Un-developed beyond the sketching stage, it was to consist of two camera tubes, one of fast response, the other slow, both examining the same scene through a beam splitter. The outputs of the two tubes were to be compared to detect motion.

3.4 Designation of Stereo Optics and Camera Electronics

While camera computer chain A was being assembled, design began on stereo optics and continued on camera electronics. It became convenient to employ a different designation for each of these lines of development. Thus, the basic type of the stereo optics is represented by a letter, e.g., C, D, E; a variant of the basic type by a number and a letter, e.g., C1, C2, and the electronics by a second number separated from the letter and first number by a dash. (See Table 3-1.)

Type C stereo optics is an arrangement of fixed mirrors to bring two separated optical paths onto a single vidicon. Type C1 is an attachment for a Bolex 16 mm moving-picture camera that can also be attached to a television camera. As shown in Fig. 3-3, the axes of its two optical paths are parallel. A possible variant is a configuration with fixed convergent axes.

Stereo optics types D and E are characterized by both variable convergence and ability to mechanically scan a field of view larger than that of the lenses. For reasons given in subsection 3.8, it was decided to provide two cameras in each of these types and turn and tilt them to achieve the variable convergence.

Table 3-1. Components of camera-computer chains.

ITEM NO.	COMPONENTS	CAMERA-COMPUTER CHAIN			
		A	C1-1	D1-3	E1-3
1	Camera for A	X			
2	TX-0 Computer	X			
3	Opto-mechanical configuration Type C1		X		
4	Camera electronics Type 1		X		
5	PDP-9 Computer		X	X	X
6	Opto-mechanical configuration Type D1			X	
7	Camera electronics Type 3			X	X
8	Three gimbals, their torquers and resolvers				X
9	Six gimbals, their torquers and resolvers				X

and mechanical scanning. Type E stereo optics provides each camera with both a wide-angle coarse-resolution field and one of narrow-angle and high-resolution. In the Type E each camera is capable of the motions provided for the D, plus pitch, roll and yaw motions to erect the pair of cameras from a tilted base.

3.5 Type 1 Camera Electronics

Experience with Camera A led to the design of the Type 1 camera electronics which is planned to be used with the Type C1 stereo optics. The resulting camera chain is called C1-1. Type 2 camera electronics is a commercially available system, on order. Type 3 is the electronics of the two cameras of the D configuration, assumed here to have the characteristics of the Type 1. For purposes of comparison, Type 3 electronics is assumed to be employed also with camera assembly E1, although another type may be employed yielding twice the aperture response (resolution).

When the field of view is imaged on the face of the vidicon the brightness or luminance of each small angle in this view causes an appropriate change in the resistance of the photoconductive surface that modulates the video output waveform. The purposes of the electronics involved, as indicated in Fig. 3-2, are:

- To deflect the electron beam of the camera tube horizontally 512 positions and vertically 512 lines so as to describe a raster. Rate of advance of the beam should be a multiple of the computer clock rate.
- To amplify the video signal and convert its amplitude at each raster position into a digital word to be entered into the computer.

While the electron beam does not actually dwell at each position, it is useful to think of it as doing so.

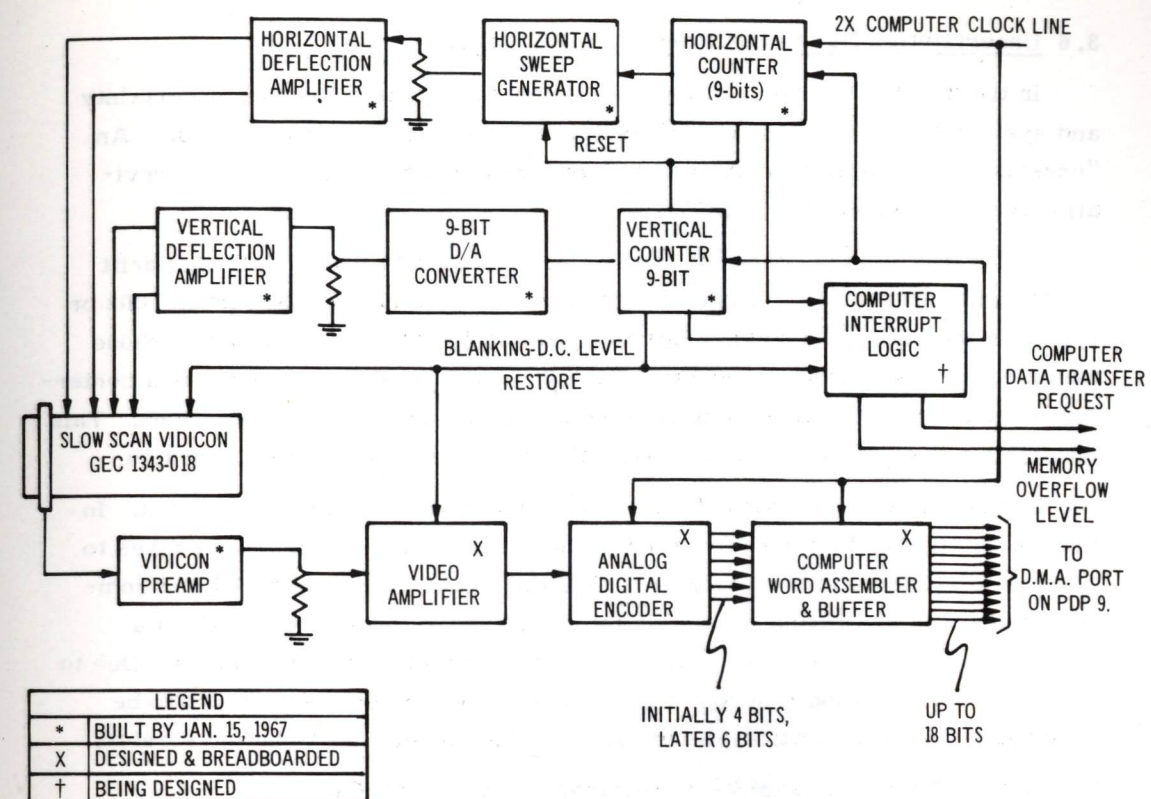


Fig. 3-2. Block diagram of Type 1 camera electronics

Deflection of the beam along one line of the raster is achieved by generating a ramp waveform in the horizontal sweep generator shown at the top center of Fig. 3-2. The beginning and end of the ramp are determined by the horizontal counter which counts from 0 to 511 at twice the computer clock rate. This multiple is chosen because two six-bit words can be fed into the computer memory in one clock time. (The remaining six bits are needed for flags, etc.). The computer clock rate is 1 MHz and the television camera frame rate is approximately 7 frames per sec.

The video signal is amplified and converted to digital words also at twice computer clock rate, i. e., 2MHz. While the sampling theorem suggested an amplifier bandpass of 1.0 MHz, it was possible to design it with a bandwidth greater than 1.5 MHz and a rapid rolloff after 2.0 MHz. Sampling distortion (aliasing) results but is moderate because the finite vidicon beam size causes the amplitude of the components of the video signal to roll off rapidly at high frequencies. While slightly reducing the S/N ratio of the amplifier, this extension of the video bandpass lessens distortion of these high frequency components.

3.6 Uncertainties Due to the Type 1 Camera Electronics

In the discussion that follows three terms will be used, error, uncertainty and systematic error. An "error" is a deviation from an expected value. An "uncertainty" is a random error. A "systematic error" is a consistent deviation from an expected value.⁽¹⁰⁾

Uncertainty in the centering of the electron beam is due to D-C bias drift which in the present design is estimated to be a maximum of one part in 1000 or 0.03 v in the bias of 30 volts. This deviation is 1/1000th of the raster, whose useful width, as reported in subsection 3.8, is 300 lines. Thus there is a centering uncertainty of 0.3 line along both the horizontal and vertical directions. This uncertainty can be reduced to one half this amount or less by careful design.

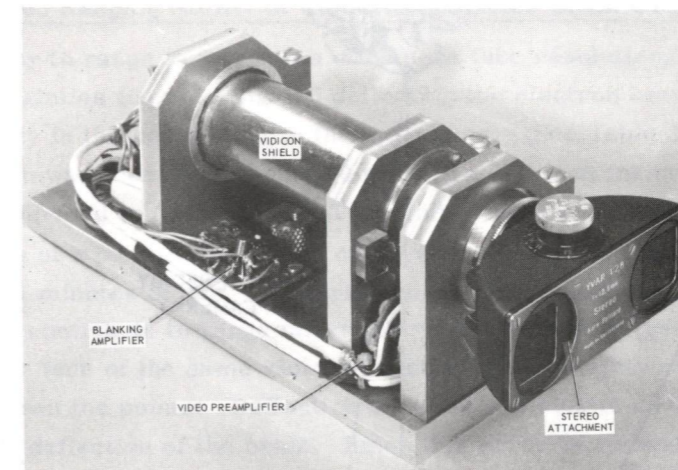
Systematic error in deflection is a function of the amount of deflection. In the present design this error is zero at the center of the raster, but it rises to 1/2% or 1-1/2 lines at the periphery. This error could be measured and compensated for in computing range. It can furthermore be reduced by about a factor of five through the use of high-gain feedback amplifier techniques. Due to lack of uniform response across the surface of the vidicon there may also be uncertainty in the indication of luminance if many levels are to be interpreted.

3.7 Optics of Camera-and-Stereo-Attachment Assembly C1

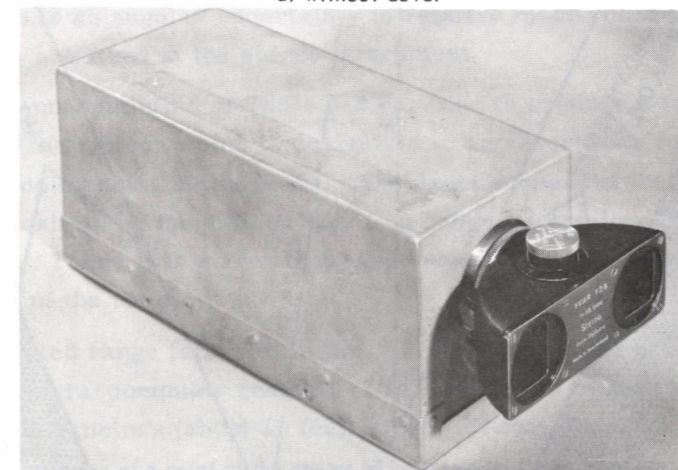
In this and succeeding subsections the word "assembly" is used to describe optics, camera case and supporting mechanisms, such as gimbals. The word "assembly" thus refers to the optical and mechanical component of a camera-computer chain, but not to its electronics.

Camera assembly C was designed to have the simplest available stereo optics.

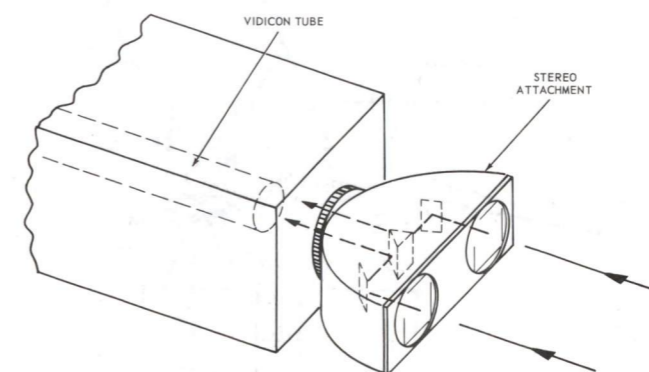
Since a vidicon camera may employ lenses intended for 16-mm movie cameras, it was possible to use the stereo attachment designed for a Bolex camera shown in Fig. 3-3. The stereo attachment receives two views of a scene, 63-mm (2-1/2 in.) apart, and forms images of the views side by side on the face of the vidicon. In considering the performance of this assembly it is convenient to straighten the two light paths and show the two views imaged as though on separate cameras as in Fig. 3-4. Each lens receives light from a rectangular pyramid in space subtending a minimum linear angle of 28°. The volume of space contained in the intersection of the pyramids is the scene that is viewed stereoscopically.



a) without cover



b) with cover



c) diagrammed to show reflecting surfaces

Fig. 3-3. Stereo camera assembly C1.

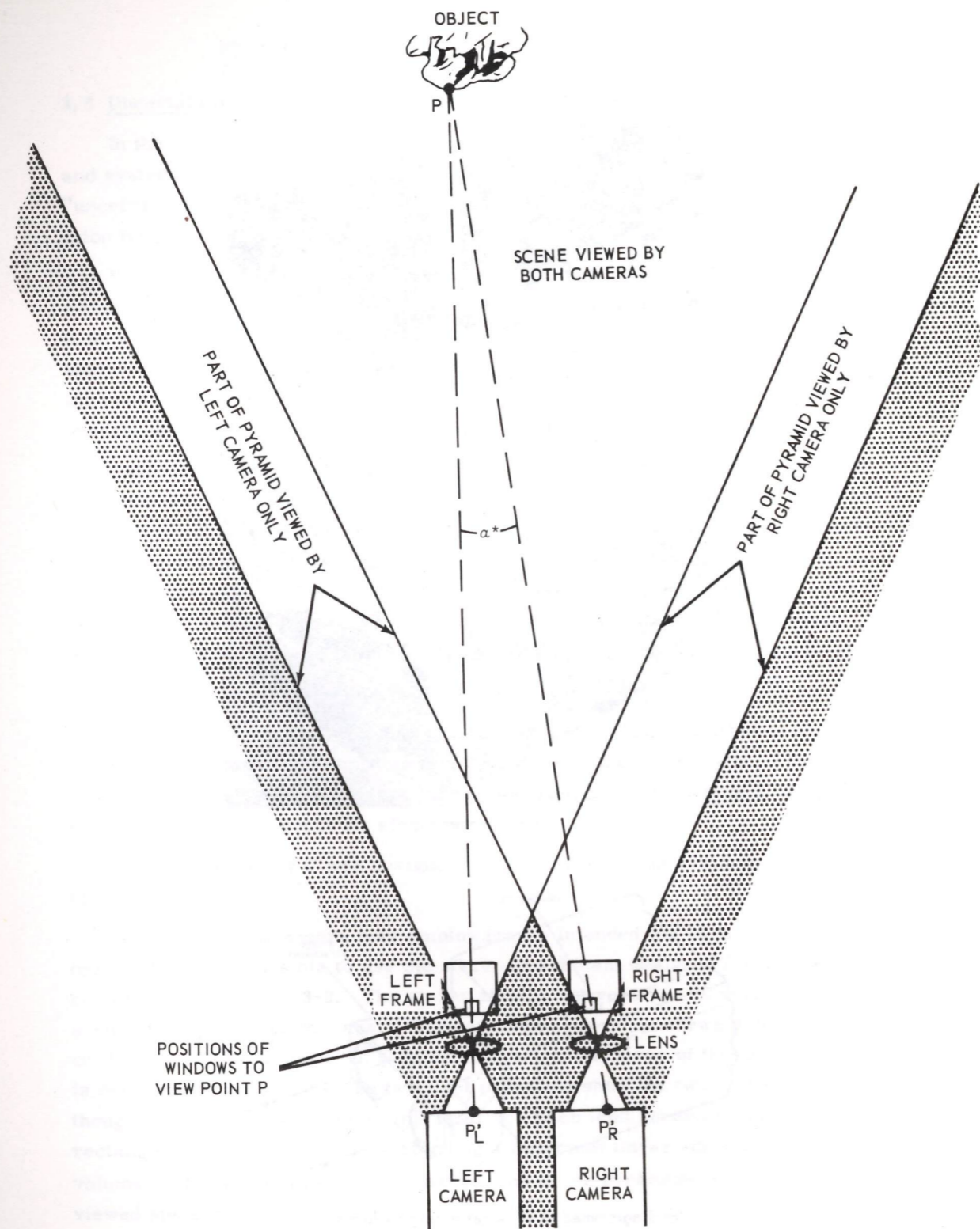


Fig. 3-4. Wide-angle parallel lenses viewing the same object.

3.8 Anticipated Ranging Ability of Camera-Computer Chain C1-1

The ability to range is a function of camera tube resolution, geometry of the optics, uncertainties in centering and deflecting the electron beam in the vidicon, and uncertainty in the positioning of the optic axes. (See Table 3-2.) Camera tube resolution was measured to be 300 lines in a 1/2 inch frame at 10% response factor, as described in Appendix A. This finite resolution produces an uncertainty in the position of a point in the image of ± 1.0 line which, in the geometry of the optics, is ± 24 minutes of arc. Centering uncertainty, described in subsection 3.6, does not contribute to range uncertainty. Since both images are on different portions of the face of the same vidicon, centering uncertainty cannot change the distance between the points. Deflection error is systematic and proportional to the amount of deflection of the beam. Since this error is consistent, measurable and can be compensated for in computing range or corrected by redesign, it is not listed. There is no angular uncertainty in relative mechanical positioning because the optic axes are fixed in the stereo attachment.

All of the uncertainties of Table 3-2 combine to produce a total uncertainty on the face of the vidicon which is equivalent to an uncertainty in angular position of a corresponding point in the scene. The exact expression relating the total uncertainty of position on the vidicon face to a range uncertainty is developed in Appendix B.3. There it is shown that range uncertainty increases approximately as the square of the range.

The indicated range falls within the curves of maximum uncertainty C-C in Fig. 3-5. Camera-computer chain C1-1 is expected to be tested in ranging on objects up to 3.0 meters (about 10 feet) away. At that distance

Table 3-2. Uncertainties of a point at the center of each image, that contribute to range uncertainty.*

	CAMERA-COMPUTER CHAIN							
	C1-1		D1-3		E1-3		E1-3	
	MIN. ARC	LINES	MIN. ARC	LINES	MIN. ARC	LINES	MIN. ARC	LINES
Uncertainty due to finite resolution of vidicon and geometry of optics	± 24	± 1.0	± 3.0	± 1.0	± 3.6	± 1.0	± 3.6	± 1.0
Uncertainty of centering of electron beam of the vidicon	0	0	± 1.8	± 0.6	± 2.2	± 0.6	± 0.22	± 0.6
Uncertainty in the positioning of the optic axes	0	0	± 1.0	± 0.34	± 1.0	± 0.28	± 1.0	± 2.75
Maximum total uncertainty	± 24	± 1.0	± 5.8	± 1.94	± 6.8	± 1.88	± 1.58	± 4.35
		min. lines		min. lines		min. lines		min. lines

*This table assumes systematic errors have been corrected

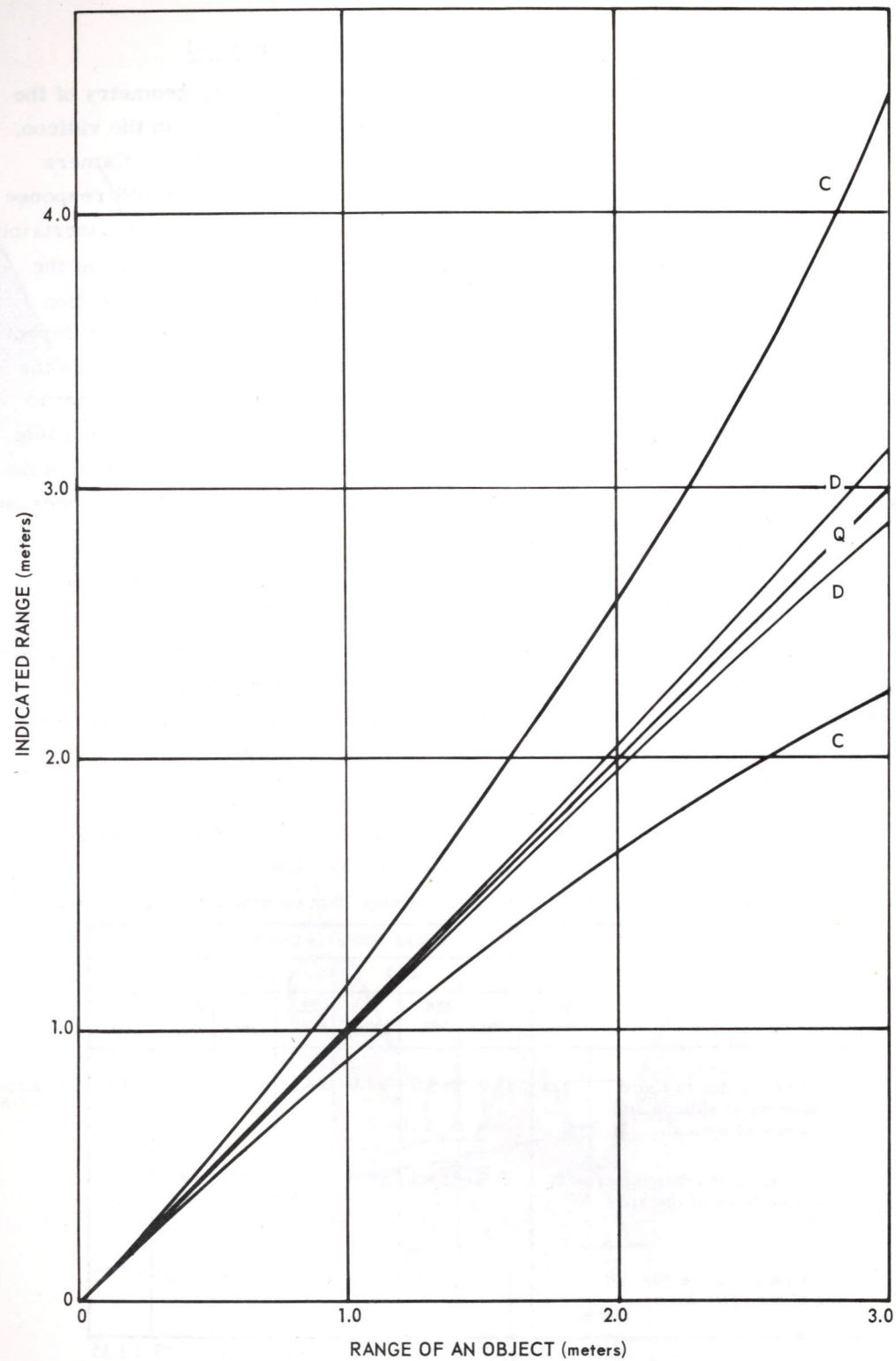


Fig. 3-5. Anticipated maximum uncertainties in camera-computer chains C1-1 and D1-3. (See Appendix B.2 and B.3 for computation of the points plotted here.)

the range uncertainty is about ± 1.0 meter. The range uncertainty may be reduced by increasing the separation of the optic axes, as is done in the D Assembly.

3.9 Camera-and-Gimbal Assembly D1

The D Assembly is a frame in which various combinations of transducers and lenses may be tested. The assembly shown in Fig. 3-6 provides for rotating each camera about its yaw axis and rotating both cameras as a unit about the pitch axis. The configuration chosen is appropriate for testing range finding with converging optic axes. The torquer-resolver platforms (gimbals) provided will make it possible to position each camera with an uncertainty of about 30 seconds of arc and position a pair with an uncertainty of about one minute of arc. (Other configurations which were considered are shown in Appendix C.)

An uncertainty of one minute of arc was also sought for each vidicon-lens assembly (camera).

$$\text{Vidicon resolvable line width} = \frac{0.5 \text{ in.}}{300 \text{ lines}} = 0.0017 \text{ in.}$$

$$\therefore \text{focal length, } f = \frac{\text{line width}}{\sin 1'} = 5.66 \text{ in.}$$

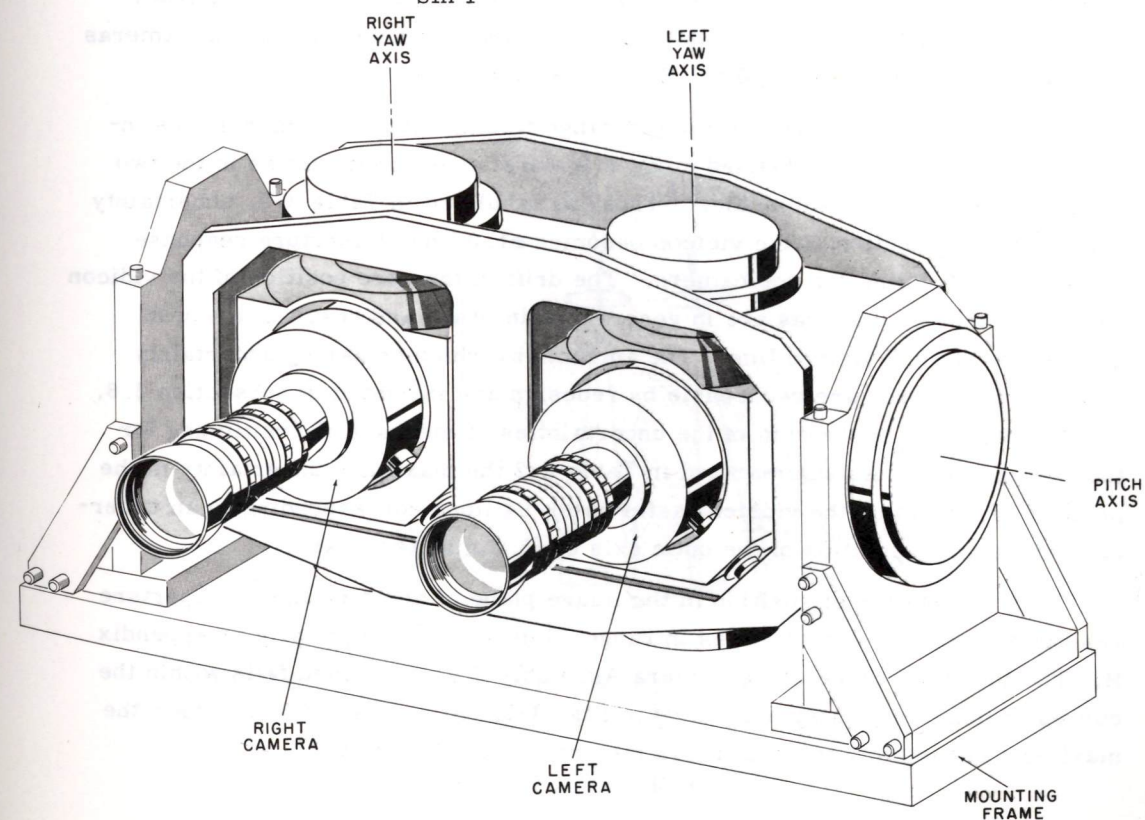


Fig. 3-6. Camera and gimbal assembly D1.

Such a lens is undesirable when considering a mechanical support of that length. The lens selected has a focal length of 3 inches and images a 5.67° square field on the raster of the vidicon. With this width of field and focal length the resolution limit is determined by the finite resolution of the vidicon.

The D1 Assembly is shown in Fig. 3-6 with 3-inch focal length lenses. The physical length of the lens is 3.9 inches and the overall length of the lens and vidicon assembly (camera) is 13.0 inches. Each camera is turned by a torque motor above it which measures 3 inches in diameter, while the position of the camera is sensed by a resolver below it measuring approximately 4 inches in diameter. The yaw axes of the cameras are separated by 6.250 inches. The cameras are mounted in a stereo box pitched by a torque motor, which is out of sight at the left. The pitch angle is sensed by the resolver at the right. Geometric relations in Camera-and-Gimbal Assembly D1 are shown in Fig. 3-7.

Motion of each camera is in increments of 1° , an amount being reduced by redesign.

3.10 Anticipated Ranging Ability of Camera-Computer Chain D1-3

Table 3-2 summarizes the uncertainties in ranging by Camera-Computer Chains C1-1 and D1-3. For the latter it is assumed that the axes of the cameras converge so that $\alpha_1 = \alpha_2$ (see Fig. B-1, Appendix B).

In determining the uncertainties in range finding with two cameras, the intrinsic uncertainties are doubled since in a worst case uncertainties in the two cameras do not cancel each other. Thus, as tabulated in Table 3-2, uncertainty due to finite resolution of the vidicon becomes ± 1.0 line of aperture response rather than ± 0.5 line for one camera. The drift of the zero position of the vidicon rasters of the two cameras are in general not in phase and result in a worst case error of ± 0.6 vidicon line. The amounts by which centering uncertainty and deflection error are reduceable by redesign are indicated in subsection 3.6. In addition, there is the sum of the uncertainties of mechanical positioning for the two cameras. As summarized in Table 3-2 the maximum uncertainty in the position of a point on the vidicon raster is ± 1.94 lines corresponding to an uncertainty in angular position of the optic axis of ± 5.8 minutes of arc.

This uncertainty of position in the image plane of about ± 2 lines of aperture response results in uncertainties in range as developed analytically in Appendix B. The indicated range using Camera Assembly D as described falls within the curves of maximum uncertainty D-D in Fig. 3-5. At a range of 3.0 meters the maximum uncertainty is about ± 0.13 meters (about ± 5 inches).

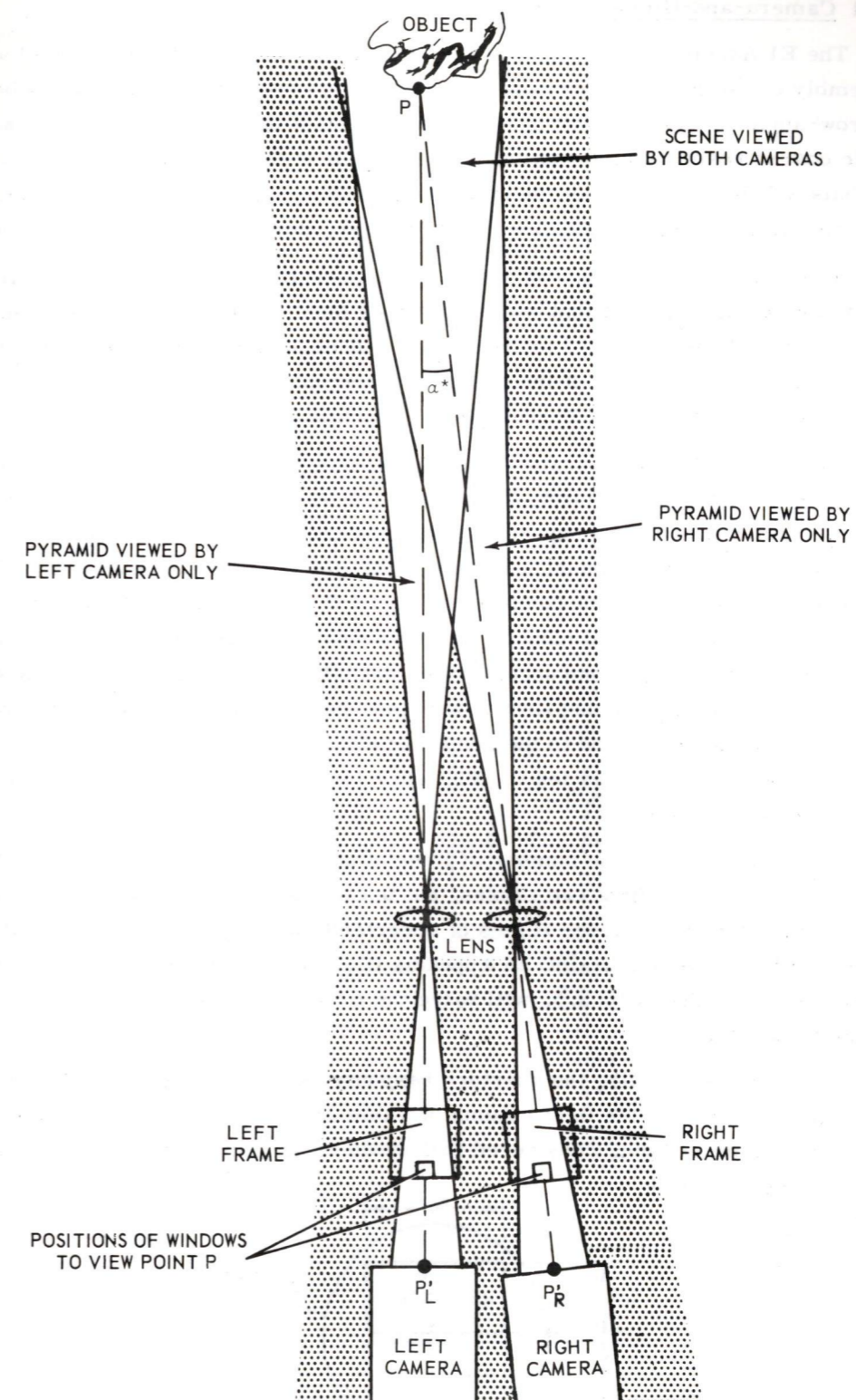


Fig. 3-7. Narrow-angle cameras converging on the same object.

3.11 Camera-and-Gimbal-Assembly E1

The E1 Assembly is being developed to determine the size and shape of a stereo assembly composed of two cameras, each of which receives on one vidicon both the narrow-angle high-resolution field shown at the bottom of Fig. 3-8 and the wider-angle coarse-resolution field shown at the top. It is to be carried by a vehicle. Its gimbals will level the axis of the stereo box each time the robot stops to look, at the same time permitting the cameras to look up, down, sideways, or backwards.

In addition to providing both wide- and narrow-fields of view, the optical train of camera E1 is folded so that while its longer focal length is 400 mm (16 inches) its front-to-back length is 6-1/2 inches. This will be achieved by mounting the vidicon at a right angle to the axis of the entrance to the optical train. (See Figs. 3-9 and 3-10.) Details of the optical train are given in Appendix D.

The wide-field optical train of the E1 Assembly will provide range finding performance with more uncertainty than that of the D1 Assembly while the narrow field optical train will provide range finding performance with less uncertainty. Figure 3-11 shows the bounds of maximum uncertainty for the three optical trains, D-D representing that of camera computer chain D1-3, E_1-E_1 and E_2-E_2 representing those of the wide-angle and narrow-angle trains respectively in camera-computer chain E1. The causes of the uncertainties are tabulated in Table 3-2. Note that the angular uncertainties due first to the finite resolution of the vidicon and geometry of the optics and second to the centering of the electron beam are ten times as great for the E1 train as for the E2, while the uncertainties in positioning the optic axes are the same.

On the other hand, the number of lines of uncertainty are the same in both optical trains for the first two causes, but ten times as great in the second optical train for the third cause. (Lines of uncertainty are represented by s in the equations for range uncertainty in Appendix B. 2.) Doubling the aperture response, as proposed in subsection 3.5, will halve the first two causes of uncertainty but not affect the third.

The effect of doubling the aperture response is most striking in the C configuration reported in the first column of Table 3-2. Accordingly, a possible C2 stereo optics is being considered that would be an improvement over the C1 in other ways as well.

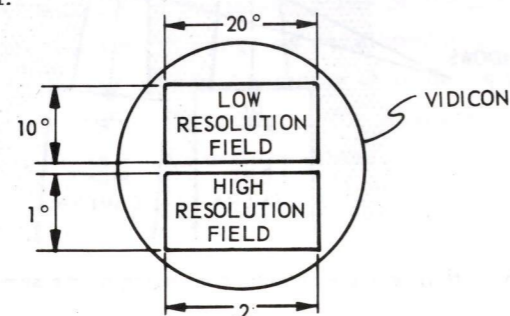


Fig. 3-8. Diagram of composite field of view at vidicon of an E1 camera.

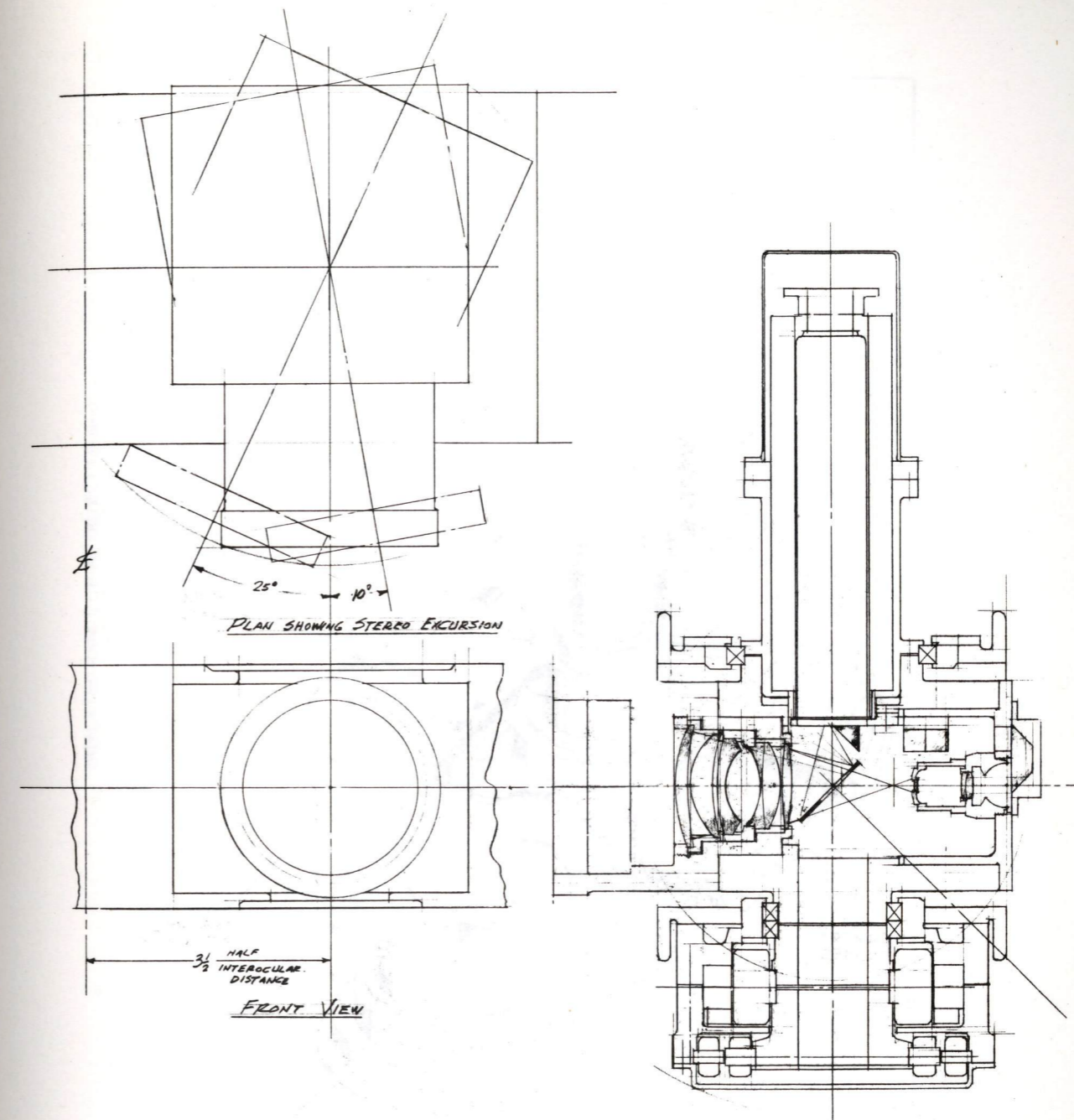


Fig. 3-9. Camera E1. Not shown are the shutter, means of focussing the lens and means of installing and changing filters.

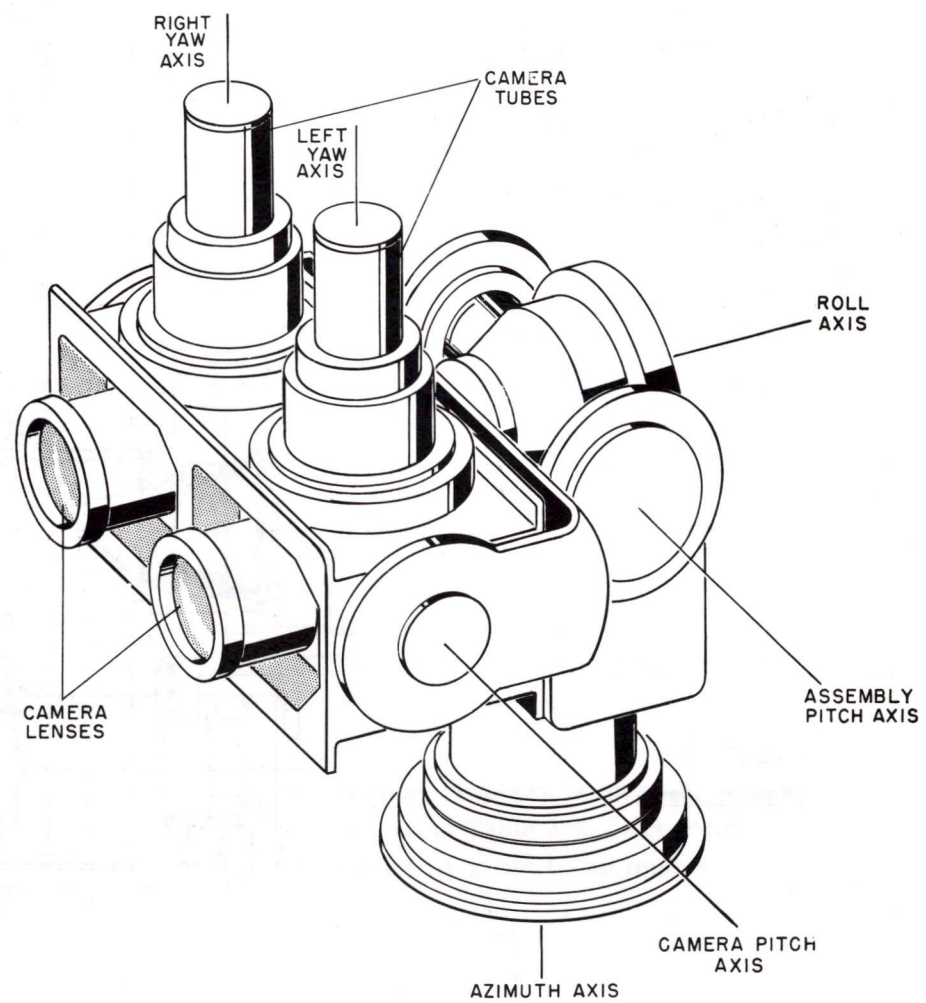


Fig. 3-10. Camera-and-gimbal assembly E1.

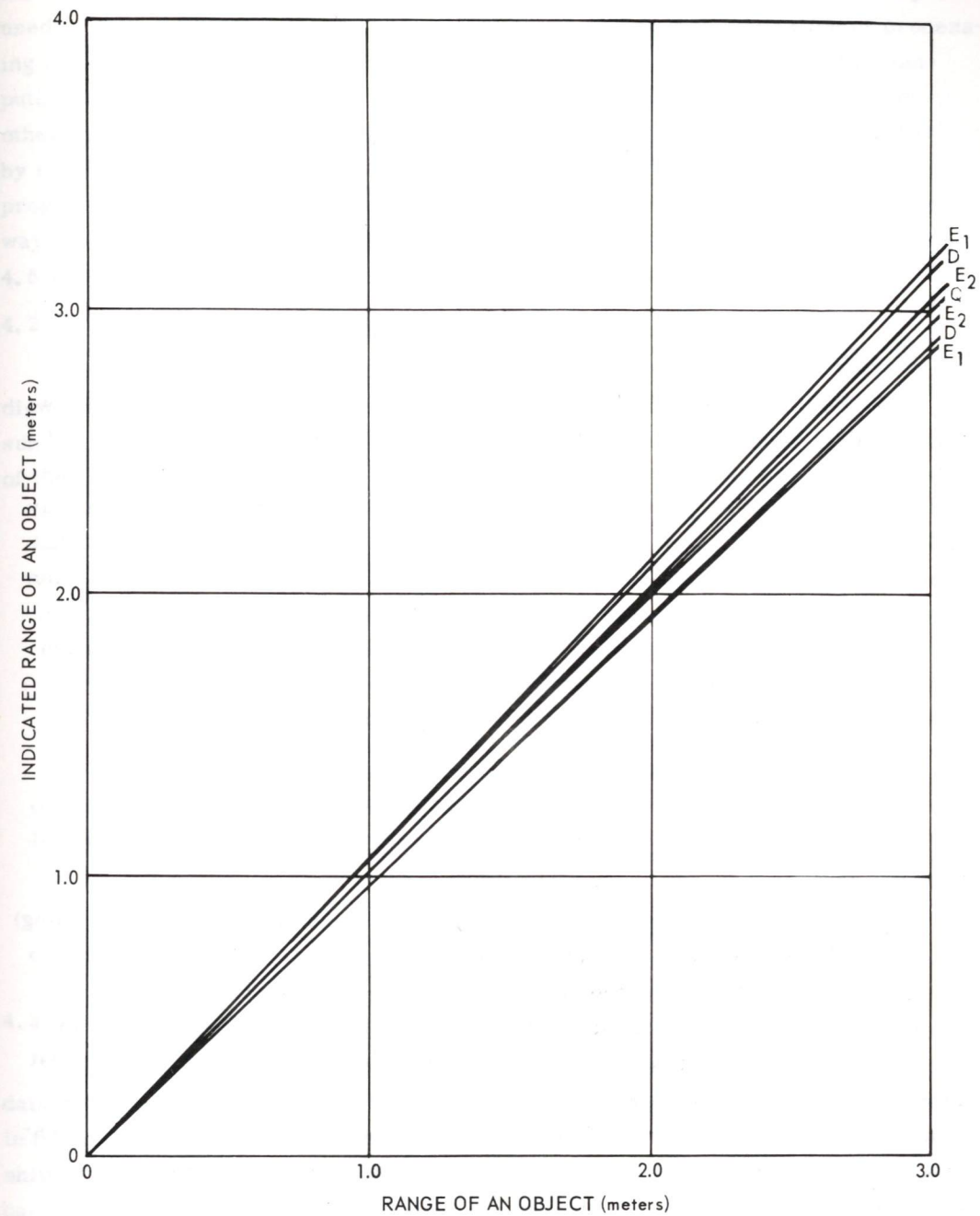


Fig. 3-11. Maximum uncertainties in indicated range for camera-computer chains D1-3 and E1-3. The curves D-D are the bounds of maximum uncertainty about the actual range q_0 . E_1-E_1 and E_2-E_2 are the corresponding bounds for the low and high resolution fields of the E1-3 chain respectively. (See Appendix B.2 for computation of the points plotted here.)

SECTION 4

VISUAL DATA PROCESSING

by

Roberto Moreno-Diaz and James Bever

4.1 General

This section is concerned with methods of processing visual data in such a way that significant features can be enhanced and abstracted. In particular, it considers processing of visual data for the generation of a "point model" of the environment. What is meant by a point model can be expressed in a simple way as follows: imagine television records of the scene in front of the camera, that have been divided into small areas or portions, each containing many resolution elements. The position of each small area is given by a pair of numbers (x, y). By means of a range finding process, a third number could be assigned to each small area indicating how far away the part of scene there is imaged. The point model consists of the set of "points" (x, y, z) describing the scene.

The point model represents the scene by giving a "replica" of it without much abstraction. The realization of an internal model in a more abstract way, so that storage can be further reduced, is the aim of such investigations as that reported in Section 5.

The ranging process is based on finding pairs of homologous (corresponding) small areas – or windows – in a stereoscopic pair of views. To decide that two windows – one in the left and the other in the right – are homologous, the window in one view is kept fixed while being compared to a variety of them in the other view, until matching is obtained. The matching process requires that the edges appearing in both windows are well defined. In general, the initial visual data is not so, and some initial processing to enhance the edges is necessary. For this reason, we began by considering different processings of the data to enhance contrast.

A conventional technique to detect – or enhance – the edge in a picture is that of differentiation in two directions.^(11, 12) This method is, however, very susceptible to noise. By considering how animal retinae process the data (see Section 8) a more generalized method of contrast enhancement was arrived at,

that of lateral inhibition. Actually, similar procedures have been extensively used in optical filtering techniques.^(13, 14, 15) A natural generalization of processing of the type indicated in subsection 8.2 has resulted in the writing of a computer program able to handle not only contrast enhancement but also perform other kinds of "filtering" of the data – the word filtering having been suggested by similar techniques in optical processing and range finding. The general program is described in the following subsections, along with an account of the way it was arrived at. The method of finding homologous windows in subsection 4.5 was the first attempt to solve the range finding problem.

4.2 Characteristics of the Visual Processing Program

Visual processing must include such manipulations of data within a two-dimensional array as those performed by the set of routines associated with a sub-array called "windo." As illustrated by Fig. 4-1, the routines are capable of the following:

1. The "windo" array may be varied in size from 1×1 to $18 \times N$, to (some multiple of 18) $\times N$ for special applications.
2. a. "Windo" can be loaded and unloaded from four directions. Hence, "windo" can be "moved" through a larger array;
b. Also, "windo" can be modified selectively.
3. A variety of (a) irregular and (b) regularly-shaped regions can be defined within "windo" which can be processed directly or mapped onto an array called "windo-within".
4. Arbitrary positions within "windo" can be processed together.
5. "Windo" can function as (a) a one- or (b) two-dimensional shift register. In particular, one "windo" can be superimposed upon and shifted across another.

4.3 Initial Approach

Our initial approach to processing visual data was to reduce a small array of data to some sort of line, representing a summary of grey level contours present in the array. Grey level contours, frequently termed "edges", occur as sudden shifts in grey level, extending typically along the boundary between a shape and a background of contrasting luminance. It should be noted that specks of noise also cause such level shifts.

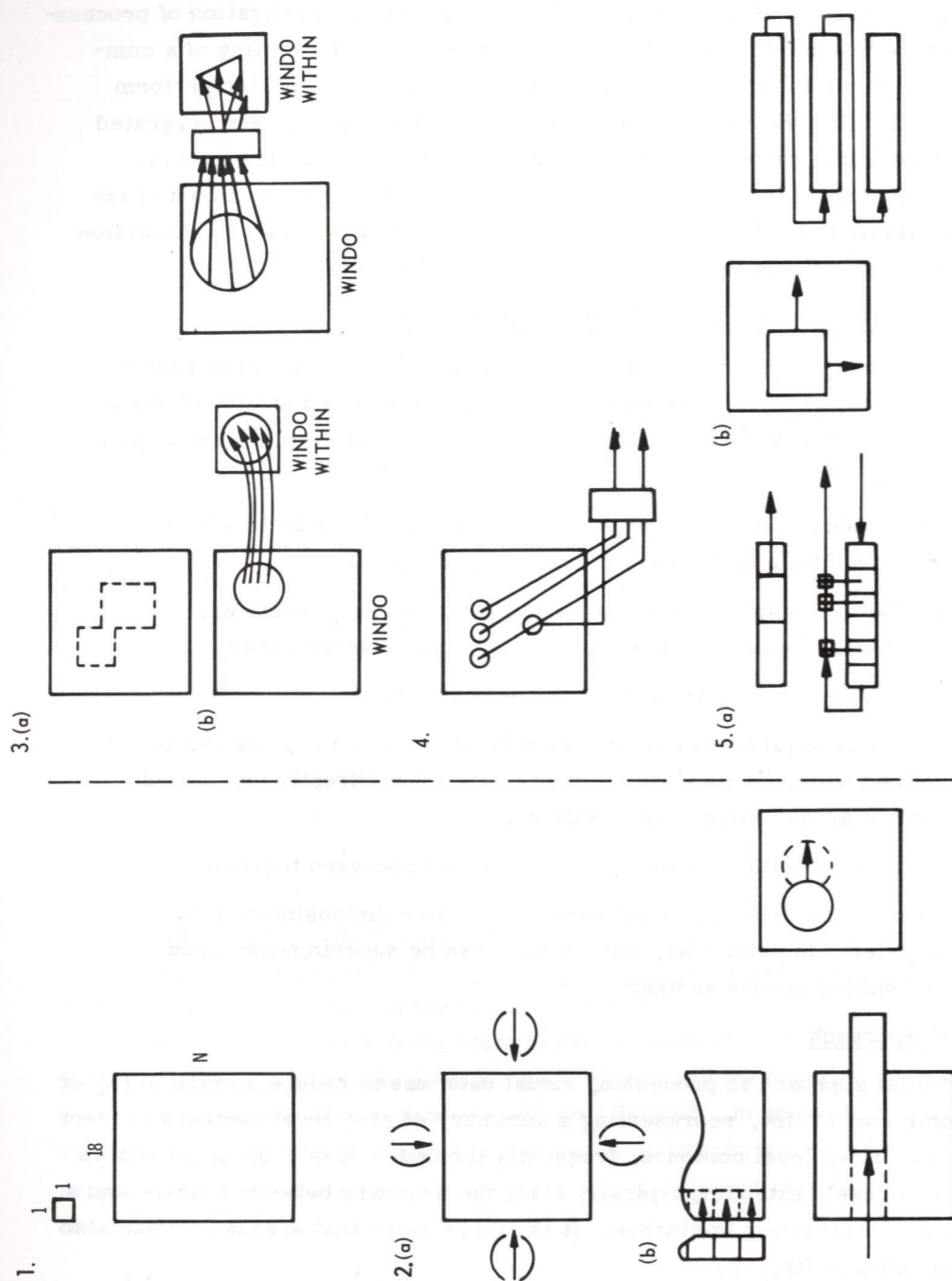


Fig. 4-1. Some properties of the information-processing space.

A computer program which reduces data to the line form first loads "windo" with data, then sweeps across this data searching for abrupt shifts in grey level. The program marks such a shift by placing a cross, consisting of five binary one's, in an array called "windo-within." A chain of binary-coded crosses may form a contour which will extend unbroken between two sides of "windo-within." A subsequent process reduces contours to straight lines, and counts them as components of a "vector."

A further process endeavors, by means of comparing addresses of data from which matching left and right "eye" pointers* have been computed, to establish the range of the data.

The line reduction scheme above cannot be used to interpret all the data configurations possible in "windo." It is especially ineffective where high resolution increases the complexity of the data.

4.4 A Second Approach

The line approach suggested that a form of correlation would be useful. A method of correlation was then developed whose first step is to differentiate the data, such that the presence of an edge is represented by a 1, the absence by 0. Afterwards, a frame is transformed into a string of raster lines taken in succession. Correlation consists of holding say, the string of the left frame fixed while passing the string of the right frame along it, one (bit) position at a time; at each position, corresponding elements are "anded" together; thereafter, a count of non-zero results is made and entered in a "correlation function." The latter is a block of registers whose length is the same as the number of bits in the longer string.

A "correlation function" can be searched as a list to determine locations of peaks and other features. In the case of a simple form of range-finding, the location in the "correlation function" of an absolute maximum determines, in principle, the range of a feature in the field scanned by the pair of cameras.

While such a correlation method has been found useful, difficulties, such as multiple peaking of the correlation function, arise when the strings are formed, since it destroys the two-dimensional character of the input. Also the presence of an absolute maximum value of the correlation function does not guarantee similarity between two forms being compared.

Input to the correlation process has been provided by a program which first differentiates data then recodes as binary the result of differentiation which consists of edges present in the data. Because of the sensitivity of this process, the results are often crowded with trivia, and do not characterize simply the original data.

* See subsection 2.3.

4.5 The Present Approach

Experience obtained in modelling animal visual systems^(16, 17) suggested the use of integral processing. Routines have been written called "Latin", short for "lateral inhibition", which is a particular case of the lateral interaction of afferent fibers along pathways leading from animal sensors.

"Latin" routines substitute, for a piece of data, a weighted average of data over a region surrounding it. Weighting functions which are positive value in the center and negative in the surround smooth noise and yield results consisting mainly of contours and other specifiable features. "Latin" is now being tested in the processing of photographs and will be tried in the processing of television images when camera-computer chains C and D go into operation.

4.6 Range Finding Planned for Camera-Computer Chains C and D

To find a pair of homologous windows in the left and right views, a window is first fixed in, say, the left view, according to some criterion such that it be the area of highest contrast or of highest luminance; then a "likely area" is defined in the other view (see Fig. 4-2). Since this area contains as many potential windows as horizontal resolution elements, the problem is to determine which of these corresponds to the one fixed in the other view. The simplest way is to compare, point by point, the fixed window to all those contained in the area of suspicion, generate some error function and use this as a criterion to decide when the matching is best attained. A point by point comparison is very susceptible to shifts of the position of the points and to changes in the gain of the video amplifiers. To lessen these difficulties, we may substitute the value of the output of the camera at a point by an average over a few positions around the one being compared, or, in other words, smooth the initial data from the cameras. This smoothing is similar to defocusing the original image.

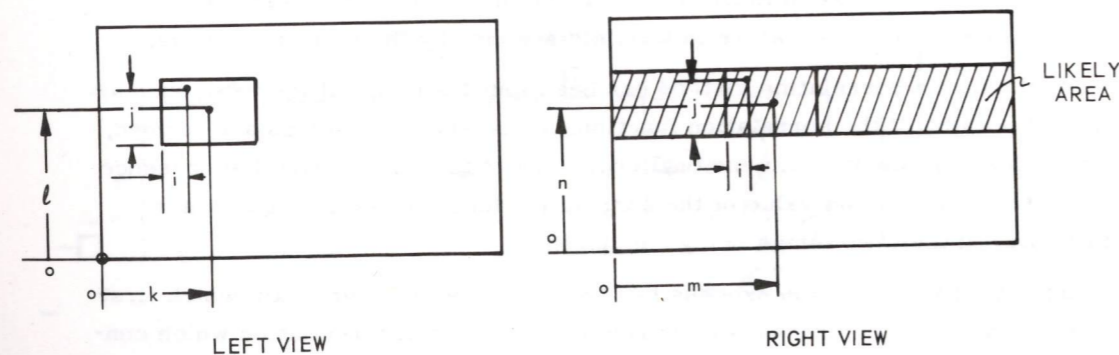


Fig. 4-2. Finding a pair of homologous windows.

Let us represent by $f_L(i, j, k, l)$ the value of the output of the camera at the point (i, j) of the window (k, l) . (See left view of Fig. 4-2.) Let $f_R(i, j, m, n)$ be the value of the camera output at the point (i, j) of the window (m, n) in the right view. By smoothing, we obtain similar distributions $\Phi_L(i, j; k, l)$ and $\Phi_R(i, j; m, n)$. The problem is then to find the coordinates (m, n) of the window in the right view that correspond to the window (k, l) in the left view. If the cameras are horizontal, $l = n$, and the width of the likely area equals the width of a window. For every window (m, l) in the likely area, we compute the absolute value of the difference,

$$|\Phi_L(i, j; k, l) - \Phi_R(i, j; m, l)|$$

and integrate it over the size of the window, that is, we form the "error":

$$\epsilon(k, l; m, l) = \frac{1}{M \times N} \sum_{i=1}^{i=M} \sum_{j=1}^{j=N} |\Phi_L(i, j; k, l) - \Phi_R(i, j; m, l)|$$

where $M \times N$ are the dimensions of the windows (M and N are odd) and (17×17) are lines used for experimentation.*

Since, to start, k and l are fixed, $(k, l; m, l)$ is computed for all values of m , that is, from $\frac{m-1}{2}$ to $F - \frac{m-1}{2}$, where F is the number of resolution elements in the horizontal dimension of the parts of the frames viewing overlapping areas. The errors so computed are stored in a list which is searched to find the absolute minimum value of error. To the absolute minimum error there corresponds a pair of addresses (k, l) , (m, l) of windows in the left and right view respectively. These windows are regarded as homologous.

For example, assume that the frames overlap horizontally in 200 resolution elements, that the window size is 17×17 , and that a left window has been fixed at $k = 30$, $l = 50$. For the likely area, $n = l = 50$, and the list of errors is formed by computing

$$\epsilon(30, 50; m, 50) = \frac{1}{17 \times 17} \sum_{i=1}^{i=17} \sum_{j=1}^{j=17} \Phi_L(i, j; 30, 50) - \Phi_R(i, j; m, 50)$$

for values of m between $m = \frac{17-1}{2} = 8$, and $m = 200 - 8 = 192$, that is, 184 values. Assume that their absolute minimum value is $\epsilon(30, 50; 35, 50)$. Windows in position $(30, 50)$ and $(35, 50)$ are considered homologous. The parallax is $35 - 30 = 5$ elements which, since the geometry of the camera is known, can be converted into distance (see Section 3).

*This method of generating errors is actually equivalent to template pattern recognition, where the template is provided by the fixed window.

Present studies indicate the need of further preprocessing the digitized video data before a point-by-point comparison is done to find range. In the approach described here, the preprocessing is only an averaging. When a complicated scene is viewed - for example, containing many edges - enhancement of the edges and reduction of the number of them prior to the range finding is necessary.

A variety of ways of achieving visual data reduction are being investigated with the computer program described in subsection 4.2. It is hoped that experimentation will lead to a family of weighting functions applicable to different types of scenes. This study will lead to specification of the characteristics of the computer in the camera-computer chain.

SECTION 5

TOWARD A METHOD OF MODELLING AN ENVIRONMENT

by

Louis Sutro

5.1 General

Since identification of life and geological forms is ultimately performed on Earth, the task of a robot in a planetary mission is to acquire pictorial data and reduce it both as a basis for decision there and for transmission to Earth. The reduced data must be reconstructed on Earth into a recognizable display.

To reduce the pictorial data we are here concerned first with the creation of a "point model" of the scene, then the determination of properties that are intrinsic, that is, invariant, say, to changes in illumination. Such a property is the reflectance of each triangular area formed by a group of three neighboring points. The point model would enable a robot to select a path to follow and, transmitted to Earth, provide a three-dimensional map of the terrain. The point-and-reflectance model, by outlining areas of different reflectance, should permit discriminating surface patterns.

The point model is considered in subsections 5.2 through 5.6, the point-and-reflectance model in subsections 5.7 through 5.10. The final subsection 5.11 considers the Earth station needed to display these models.

5.2 Proposed Method of Acquiring a Point Model

A necessary first step is to aim the cameras at an object of approximately known position, such as a flashing cross on the forefinger of one hand. The forefinger will then be moved until it touches an edge or other feature that can be identified. The feature will have to be found in two views by the method of subsection 4.6 since the cameras cannot be aimed at the point of the finger with great accuracy.

A way of finding a second point, described in subsection 2.3, is to move the windows along the edge used to identify the first point. Another way is to place the left window over any area with a sharp discontinuity in luminance and then search for the homologous point with the right window, while keeping the first point where it can be recovered.

When three points have been acquired, they can be stored in the computer memory as points of a triangle. Using two of these points as the base of a second triangle, a new third point can be acquired, and so on. As the number of points grows, the uncertainty of position also grows until it becomes desirable to aim again at a known point.

5.3 An Example

Assume that the robot has landed on Mars and has transmitted to Earth both a panoramic view of its environment and its approximate position as determined by two observations of the sun. The Earth operator has commanded the robot to proceed along azimuth 43° for 50 meters, if feasible, to map and describe the terrain as it proceeds, and to transmit this description together with sample television pictures.

Looking in azimuth 43° , the robot views a scene as that in Fig. 2.3. It needs to form a point model and from that determine a route to follow.

Let us assume that points are then determined on the surface of the three nearest mounds and on the ground among them. Figure 5-1 shows points on the surface of the mound at the left center of Fig. 2-3, drawn by the same artist who made that illustration. He, of course, knew where to place the points. Where the choice of points is made by the computer it is better to obtain more points so that the connections between them, if made, are not as significant. For example, triangle GEF is a shelf whose appearance is destroyed by connecting points E and J. For a point model, connections between points do not need to be made as explained in subsection 2.3. But for a point-and-reflectance model such lines serve as bounds of the planes for which reflectance is determined.

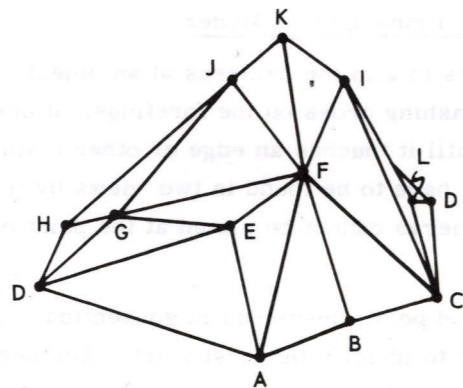


Fig. 5-1. Triangulated view of the mound in Fig. 2-3.

Table 5-1 presents the location of points F, G, and J in a coordinate system whose origin is in the robot, whose Y axis is a horizontal line through this origin in the direction the robot is headed, whose X axis is a horizontal line perpendicular to the Y axis and whose Z axis is perpendicular to the plane of the other two axes. For use in an extension of this example, the centroid of the triangle formed by these points is included.

Table 5-1. Coordinates of points F, G and J and their centroid.

	METERS		
	x	y	z
F	0.14	3.45	-0.09
G	-0.46	2.74	-0.25
J	-0.14	4.64	0.28
Centroid of ΔFGJ	-0.15	3.94	-0.02

The coordinates presented in Table 5-1 are computed from the data of Appendix E.1. They are located with an uncertainty given in the lower right corner of Table B. 1 of Appendix B.

5.4 Interrogation of a Point Model

To find a passage between mounds, the computer can form what a surveyor calls "profiles." A transverse profile will be defined as one in a plane perpendicular to the horizontal line of sight into the scene which will be the robot's Y axis if the robot is level. Figure 5-2 shows such a profile drawn on the imaginative illustration of Fig. 2-3. This profile does not have to be drawn; it can be stored as a set of points.

An algorithm to determine a profile can be developed as follows. A point p in the profile is $p(x,y,z)$ where y is constant. To find a point p on the profile for each value of x, a volume of the model such as that shown in Fig. 5-3 has to be searched, a task which may be facilitated by storage of the model in the form of a list indicating the adjoining point in each direction. Such a list can be searched for the highest point in each elementary volume shown in Fig. 5-3.

When the profile of Fig. 5-2 has been formed, it can be searched to determine whether it contains an expanse wide and level enough to permit the robot to cross. When such an expanse has been found, more transverse profiles can be "drawn" nearer and further away. The apertures found in each case can then be connected by longitudinal profiles, and these can be examined to determine if their pitch permits the robot to move ahead.



Fig. 5-2. Supposed views of the surface of Mars with a profile drawn in.

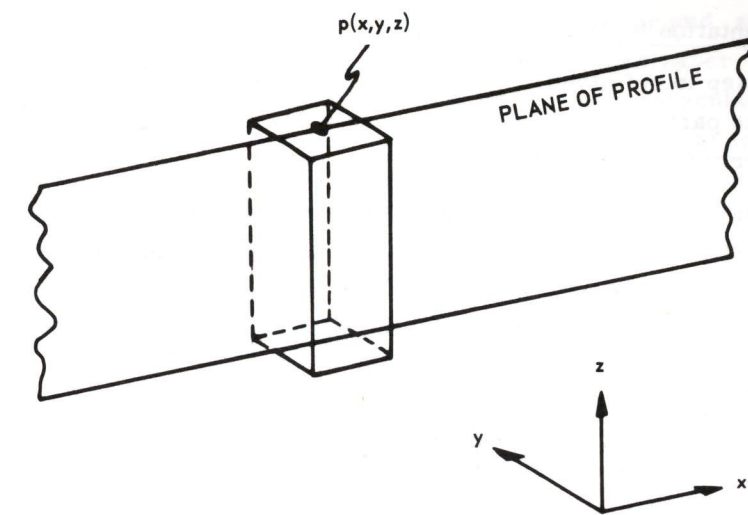


Fig. 5-3. Volume surrounding point p in the plane of a profile.

Another form of interrogation that can be performed on the model is generalizing from it both to locate the peaks, bases and approximate curvature of the mound as well as to conclude that except for the mounds all of the surface shown in Fig. 2-3 is an approximate plane.

5.5 Interrogation a Point Model in the Search for Evidence of Life

A profile may be "drawn" of a free-standing object, obtained in a manner similar to the making of a transverse profile of the scene. The profile can then be checked against a stored criterion such as: Is the object larger above than at its base? This is a characteristic of much of plant and animal life on Earth.

5.6 Mapping

Mapping of the surface of Mars can be carried out by the methods of a surveyor. From a starting station he runs a traverse of straight line segments, measuring the length of each and the angle between the forward and backward segments. The intersection of each pair of segments is called a "station" and is marked on the ground.

A Mars station can be defined as a place where the robot stops to look, and the distance between stations may be determined by an odometer. Enough stations should be marked to permit the robot to retrace its path. A suggested form of marking is a transponder which can be excited by a radio on the robot and homed in on.

5.7 Representation by Points and Reflectances

A next step in refining the model of the environment is to record a property of the surface particularly one that is invariant under changes in illumination. Such a property is reflectance.

The minimum area for which reflectance needs to be determined is a triangle formed by three neighboring points. The amount of computation required to determine the reflectance will, of course, be weighed against the data reduction achieved. The principles to be employed follow.

5.8 Computation of Reflectance^(18, 19, 20)

A television camera detects a mosaic of values of what used to be called brightness but is now called luminance, L . The term brightness is now limited to subjective response while luminance is used where the response can be measured. The unit of luminance in the United States is the footlambert. Except when it is the property of a light source, luminance is the product of illumination onto a surface and the intrinsic property of that surface called reflectance. Let us consider each of these factors separately, then together.

Illumination is a vector whose unit in the United States is the footcandle, whose representation is \bar{E} (for éclairage) and whose component normal to a surface is \bar{E}_N . The latter two measures are related by the Lambert cosine law:

$$E = E_N \cos \theta_i \quad (5-1)$$

where θ_i is the angle of incidence (See Fig. 5-4).

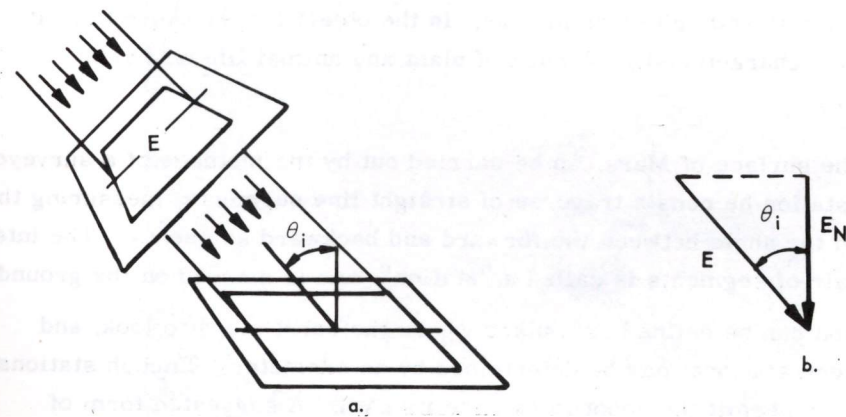


Fig. 5-4. a. The Lambert cosine law showing that light flux striking a surface at angles other than normal is distributed over a greater area. b. Representation of illumination E as a vector and its normal component as a smaller vector of magnitude $E \cos \theta_i$.

Three forms of reflectance are identified: diffuse, specular and, spread. (See Fig. 5-5.) Diffuse reflectance ρ_D is that such as blotting paper; that is, it yields the same luminance from whatever angle it is viewed. Specular reflectance ρ_{SR} is that of a polished surface; that is, it reflects light at an angle from the normal equal to the angle of incidence. Spread reflectance ρ_{SD} is that of a roughly polished surface, reflecting so that the angle of reflectance only roughly equals the angle of incidence. Other forms of reflectance are combinations of these.

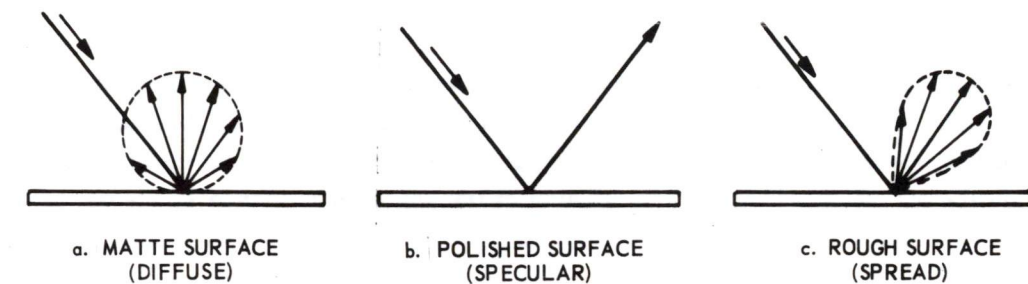


Fig. 5-5. The type of reflection varies with different surfaces: (a) matte surface (diffuse or Lambert); (b) polished surface (specular); (c) rough surface (spread).

The product of normally incident illumination E_N onto a surface, expressed in footcandles, and the diffuse reflectance of the surface ρ_D is luminance expressed in footlamberts,

$$\rho_D E_N = L \quad (5-2)$$

Stated another way, the diffuse reflectance of the surface is the ratio of the luminance from it to the illumination upon it. Figure 5-6⁽³¹⁾ shows the use of this equation in determining the reflectance of a diffuse or non-glossy surface such as a painted wall. In Fig. 5-6(a), a General Electric type 213 light meter has been placed with its photosensitive surface against the wall and then drawn back two or three inches until no shadow falls on the wall. A typical reading might be 40 fL of luminance. In Fig. 5-6(b) the meter is held with its base against the wall. A typical reading in this position might be 65 fc of illumination. Substituting the two measurements into Eq. (5-2) yields

$$\rho_D = \frac{40 \text{ fL}}{65 \text{ fc}} = 60\%$$

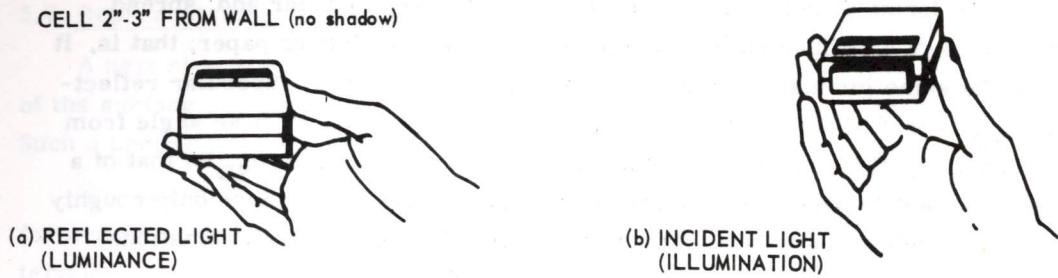


Fig. 5-6. Reflectance measurement with G.E. Type 213 light meter.

This is an approximate determination because the areas measured are large and the accuracy of this instrument is not high.*

A television camera can be used on Mars to determine diffuse reflectance in a similar manner. Aimed at the surface in question the camera measures luminance. Aimed at the sun with a suitable protecting filter it can measure illumination E . Multiplying E by the cosine of the angle of incidence θ_i yields the normally incident illumination E_N .

Figure 5-7 diagrams the measurement of diffuse reflectance by a television camera. The energy represented by the vectors E_N is reflected in an amount determined by the value of ρ_D and in a manner represented by the spherical clusters of vectors. The magnitude of each vector in a cluster bears a cosine, or Lambert, relation to the maximum vector in the cluster. It can be seen that if the camera is aimed at a steep angle it detects few long vectors, at a shallow angle many short vectors. The luminance detected in both cases is the same.

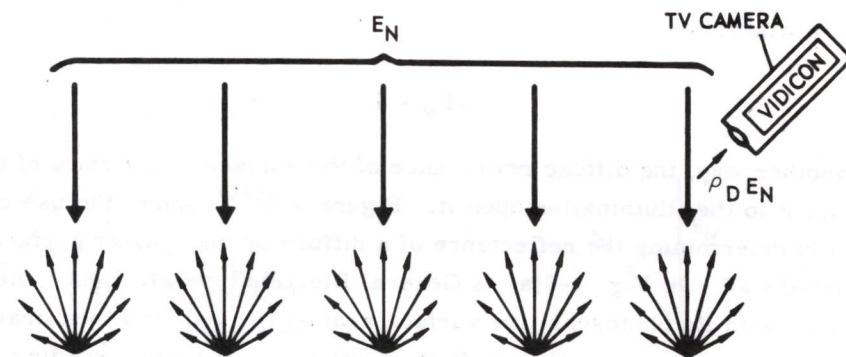


Fig. 5-7. Conversion of illumination to luminance when the surface has reflectance ρ_D is large and is uniformly illuminated.

*An instrument that measures luminance in a narrow angle of view at high sensitivity is the Spectra Brightness Photometer.

To measure specular and spread reflectance requires the ability to move around an object and measure vectors whose magnitude is different in different directions as in Fig. 5-5. When specular and spread reflectance are present luminance is determined by the polynomial:

$$L = \rho_D E_N + \rho_{SR} E + \rho_{SD} E \quad (5-3)$$

where

$$\rho_{SR} = \text{specular reflectance}$$

$$\rho_{SD} = \text{spread reflectance}$$

Note that the illumination E from the light source to the reflecting surface is used in the second and third terms. To measure specular reflectance it is necessary to move the camera to a point such that the light source, the measured surface, and the camera are in the same plane, with the angle equal to the angle of reflection. To measure spread reflectance, the same conditions must be followed and, in addition, the angle of reflection must be varied about the value equal to the angle of incidence to measure the "spread" of the light.

In most illuminating engineering work, luminance is computed from only the first term in the above equation. There may, however, be three such terms, one for each light primary. These are referred to as "tri-chromatic coefficients"⁽¹⁹⁾. Color measurement, which this implies, will be considered later.

The point model can be acquired under any illumination sufficient for edges, differences in shape, or differences in reflectance to be detected as contrast. The point-and-reflectance model, on the other hand, requires measurement of the illumination. To acquire the second kind of model, therefore, the robot must first determine the illumination upon the surface or, better yet, of the illumination upon the scene.

From its observations of the sun at two times in one day the robot can predict at any other time, the azimuth and declination of the sun and the resulting illumination normal to any surface of known position. Because there is almost no atmosphere on Mars, it is expected that there will be almost no sky light, except during dust storms. Then the lenses of the cameras will probably have to be capped to protect them from injury. The only sources of light are expected to be the sun and light reflected from the surface of Mars. In the example that follows only sunlight is considered.

The method presented here is similar to that of a painter who sketches a scene so that he can return to his studio to paint it. He observes the direction of the sunlight and, if he is acquainted with the effect of sunlight, keeps it in the

back of his mind as he concentrates on shapes and their reflecting properties. In doing this he is recording a scene invariant to illumination. Returning to his studio he can re-introduce sunlight as he paints the highlights, shades and shadows.

No painter that we know is aware of the mathematical methods of this task. Such methods are common in illuminating engineering, for example, where to achieve a desired appearance of a room, bulbs, wattages and reflectors as well as the reflectances of walls, ceiling and floors are considered.

5.9 Computation of the Diffuse Reflectance of ΔFGJ

The example of subsection 5.3 will now be extended by determining

1. The direction components of the normal to triangle FGJ , computed in Appendix E.2,
2. The average luminance L of this triangle, measured by the left camera, as 1300 footlamberts,
3. The illumination E from the sun, measured also by the left camera, as 4000 footcandles,
4. The direction components of sunlight, determined from the position of the sun stored in the computer memory,
5. The angle between the direction of sunlight and the normal to ΔFGJ computed in Appendix E.3 to be $\cos^{-1} 0.71$,
6. The illumination normal to the triangle, employing Eq. (5-1),

$$E_N = 4000 \times 0.71 = 2840 \text{ footcandles}$$

7. The average diffuse reflectance of ΔFGJ , employing Eq. (5-2),

$$\rho_D = \frac{1300}{2840} = 0.46$$

Measurement of the luminance of the remaining triangles, together with continued application of the above equations, will yield the average diffuse reflectance of every surface of the model.

5.10 Interrogation of a Point-and-Reflectance Model

A point-and-reflectance model can be interrogated for properties that can not be extracted from the point model. For example, in a point-and-reflectance model of a scene containing dark rock and light sand, the computer could determine their boundary and report it to Earth; or, on command from Earth, the robot could explore one area and ignore the other.

5.11 The Earth Station

Two ways of transmitting pictorial data to Earth are being planned. One is transmission of television pictures, either monocular or stereoscopic, the other transmission of the model of the surface of Mars employed by the robot in making decisions. The model can be presented on the same screens (Fig. 5-8) as the television pictures, provided a computer is employed on Earth to store it and arrange it for display. The arranging process can be such as to give the viewer the impression of moving through the scene.

If the model is composed of points and reflectances then the Earth operator will command under what illumination he views it. Figure 5-9 illustrates a possible display whose luminance to the eye of the operator is computed from the reflectances of the surfaces and from illumination that he has commanded to come from the upper right of the scene.

5.12 Acknowledgement

The concepts presented above and corollaries yet to be described evolved with the help of

Prof. John T. Rule, former Head, Dept. of Graphics, MIT

Dr. Adelbert Ames, former Director, Dartmouth Eye Institute

Dr. S. Howard Bartley, former Research Associate, Dartmouth Eye Institute

Nickerson Rogers, former Research Fellow and stereoscopic photographer, Dartmouth Eye Institute

Paul Sample, former Artist-in-Residence, Dartmouth College

Willard Allphin, Illumination Research, Sylvania Electric Products Co., Salem, Mass.

Dr. Parry Moon, Associate Professor of Electrical Engineering, emeritus, MIT

Dr. Domina Eberle Spencer, Professor of Mathematics, University of Connecticut, Storrs, Conn.

Robert N. Davis, Research Staff, Lincoln Laboratory, MIT

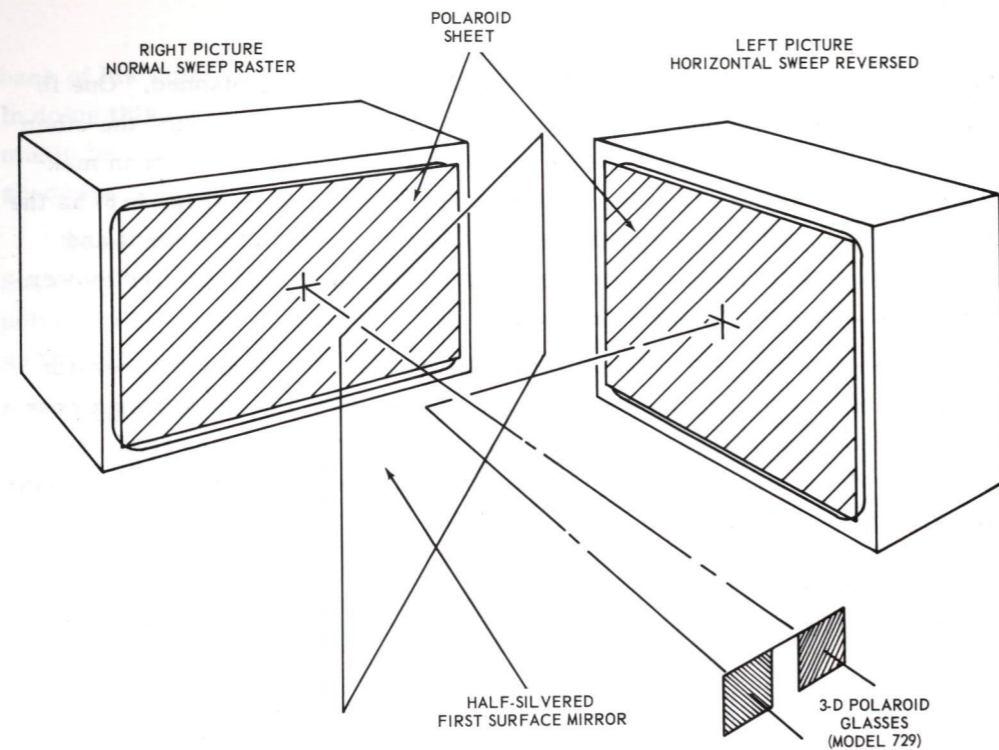


Fig. 5-8. Stereo picture presentation using two 23-inch monitors and Polaroid filters. Orientation of each filter in glasses corresponds to orientation of filter before each monitor.

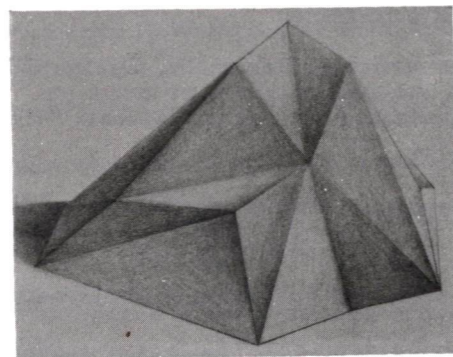


Fig. 5-9. Triangularized view of the mound in Fig. 2-3 shaded as though by an Earth computer.

SECTION 6

ADDITION OF A LEARNING ABILITY TO THE DECISION-MAKING SUBSYSTEM

by

William L. Kilmer, Jay Blum and Warren McCulloch

6.1 Introduction

In a temporally changing environment, a static decision-making computer such as the model described in Refs. 3 and 4 may be at a significant disadvantage. Consequently, that model is being extended from that which has been S-RETIC (where S denotes a static environment) to a model of a decision computer capable of several forms of conditioning and learning. The new model is called STL-RETIC, where T denotes a temporally changing environment, L a conditioning and learning ability of the kind described in Section 9.

As simulated on the Instrumentation Laboratory general purpose computer, the static S-RETIC model consists of twelve modules (M_i), inputs (γ 's) to the modules from the simulated environment (Σ), and outputs (P) to threshold elements. The decision computer commands a mode of activity only when a threshold has been exceeded.

To make learning possible, the model is being modified to include logic and memory elements reinforceable as a function of experience. The logic elements will operate in the A-part of each module, while memory elements will store past inputs from the environment, messages from other A-parts, and past A-part computations. Figure 6-1 shows each γ input to an M_{i_a} paralleled by a memory element and each connection (ω) from another M_{i_a} , similarly paralleled. The number of γ inputs from each M_{i_a} has been increased from five to seven. Note that the design continues to be one that can be built of hardware as well as simulated.

6.2 Requirements

The augmented STL-RETIC structure ought to satisfy the following requirements:

First, if the environment is rich in stimuli and if the responses which follow these stimuli are at least partially unknown, it is not possible to establish an input-output correspondence table defining for the decision computer which acts should follow which stimuli. The computer must be trained.

Second, the decision mechanism should respond differently to different overall environments; it should be able to adjust to a new input-output specification, so that it need not be reprogrammed.

Third, if one or more of the robot's sensory or action subsystems fails, the decision computer should compensate by changing its responses to input stimuli.

6.3 Nature of the Problem

A difficulty in augmenting S-RETIC to be able to learn is that each module appreciates only part of the input-output correspondence. How can all of the modules function harmoniously in the same direction, so that the effect is cooperative and not chaotic? Generally, each module is most sensitive to different input stimuli; for example, module 5 may discriminate among overall (Σ) input sets 67, 68, and 69, whereas module 4 may discriminate among inputs 23, 24, and 25. Moreover, neither may discriminate at all among other input sets. If one module makes necessary distinctions between two cases which call for different responses, it has to be able to notify neighboring modules of these distinctions as they occur. This would enable the other modules to act appropriately.

6.4 Approach to the Problem

The problem mentioned in subsection 6.3 may be limited by requiring that the decision-making computer be effective only over several hundred successive stimuli. Each call for a change in response would be stored in memory elements so that, when several hundred had been accumulated, actual alteration of the response would occur. Naturally, the smaller the number of stimuli required, the better the model will be able to yield up-to-date responses. Thus, the benefit of reduced complexity achieved through storage of desired responses must be weighed against the delays intrinsic to such storage.

6.5 Implementation by Operations on P_{AB} Vectors

Each kind of learning will be implemented by operations on the P_{AB} vector which goes from the A to the B part of each module. These operations will be similar to those described in subsection 5.6 of Ref. 4: flatten vectors, elevate vectors on certain modes, depress them on certain modes, measure the peakedness of vectors in order to decide what to do with them, and permit one vector to change into another systematically, smoothly, and gradually.

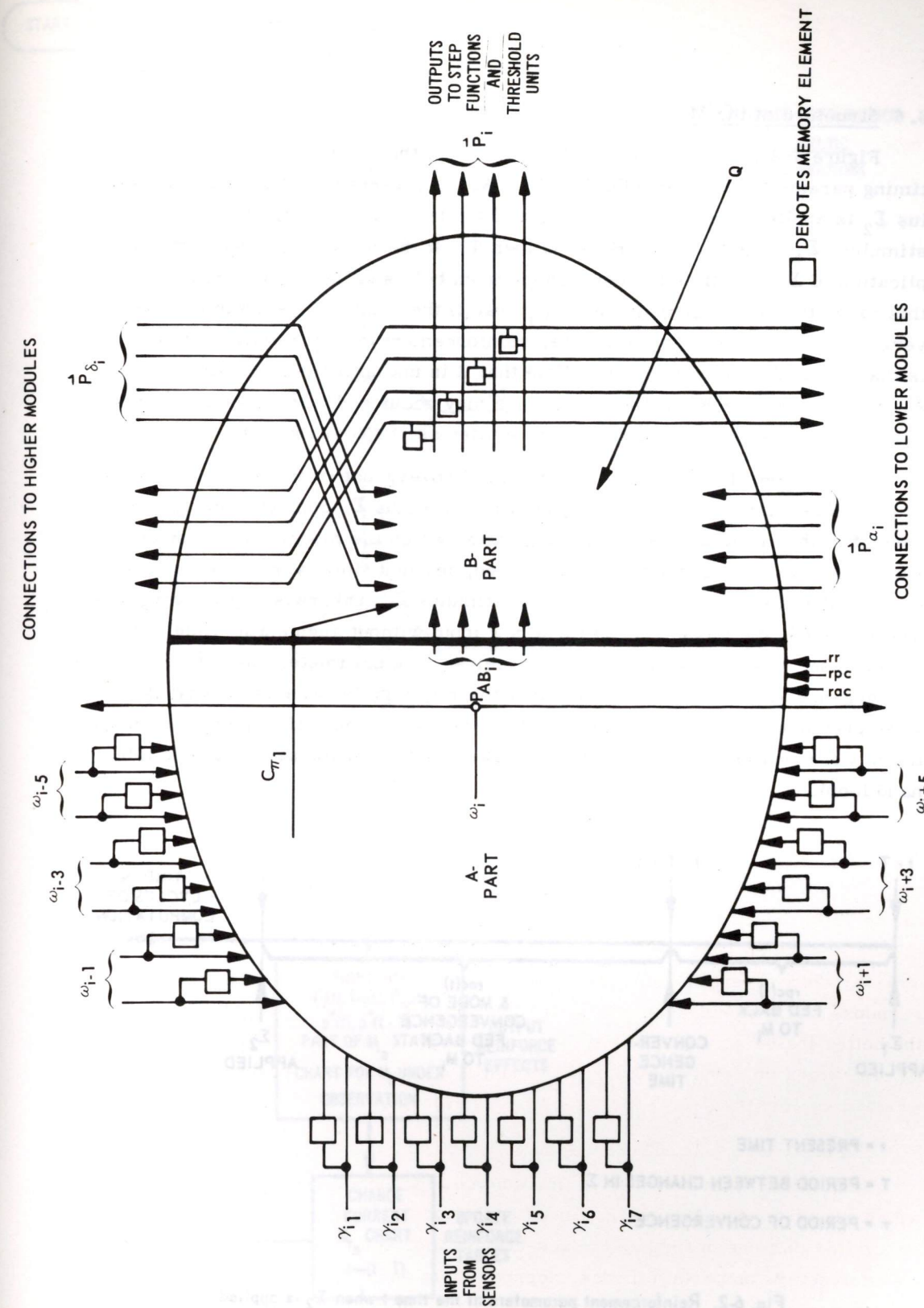


Fig. 6-1. Input and output connections to parts A and B of module M_i .

6.6 Structure of the Model

Figure 6-2 is a schematic representation of the relation between the three timing parameters used in STL-RETIC. At the present time t at which a stimulus Σ_2 is applied, the model looks back to the time of the application of the last stimulus, Σ_1 - the time difference is then T . If the model converged after application of Σ_1 , the time to convergence from $t-T$ is called τ , with $rpc(t)$, in this case, the reinforcement due to Σ_1 , $rac(t)$ the reinforcement due to convergence. If no convergence occurs, reinforcement over the entire period T is that due to Σ_1 , namely $rpc(t)$. Note that it is not until the application of Σ_2 that reinforcement due to convergence may occur and that consequently the reinforcement signals rpc and rac are written as $rpc(t)$ and $rac(t)$.

A condensed flow chart of the stimulus-convergence-reinforcement process is given in Fig. 6-3. After getting the first stimulus Σ , computations based on receipt of the stimulus are carried out, after which the model checks for convergence. If convergence has not occurred, present time is incremented one unit and the model looks for a new input stimulus Σ . Otherwise, if convergence has occurred, the model checks for a new input without augmenting time. If, in either case, a new input has not been received, the parameters used to determine convergence are recomputed. Note that these change in value from what they were previously only if time has been incremented. Thus, if convergence occurs, the model remains converged until receipt of another stimulus. (Stays in left-hand loop).

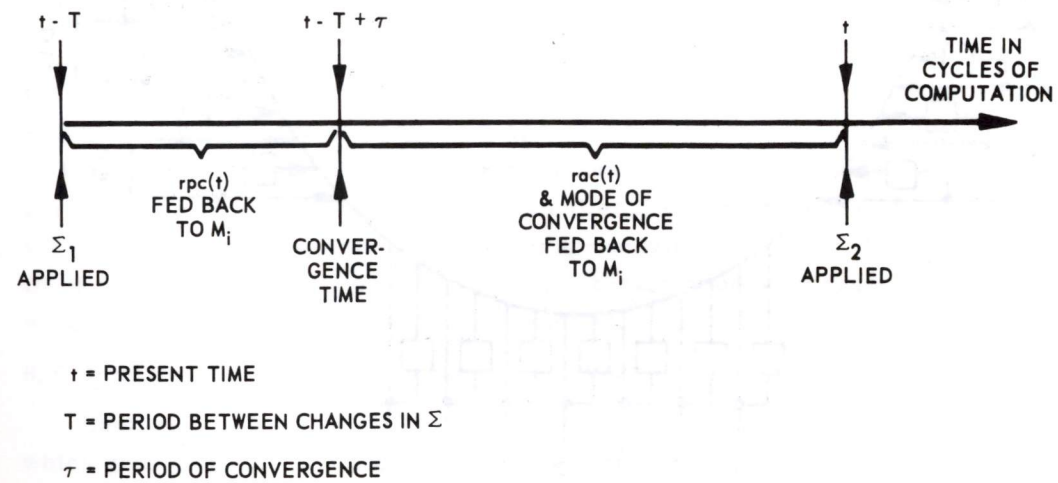


Fig. 6-2. Reinforcement parameters at the time t when Σ_2 is applied.

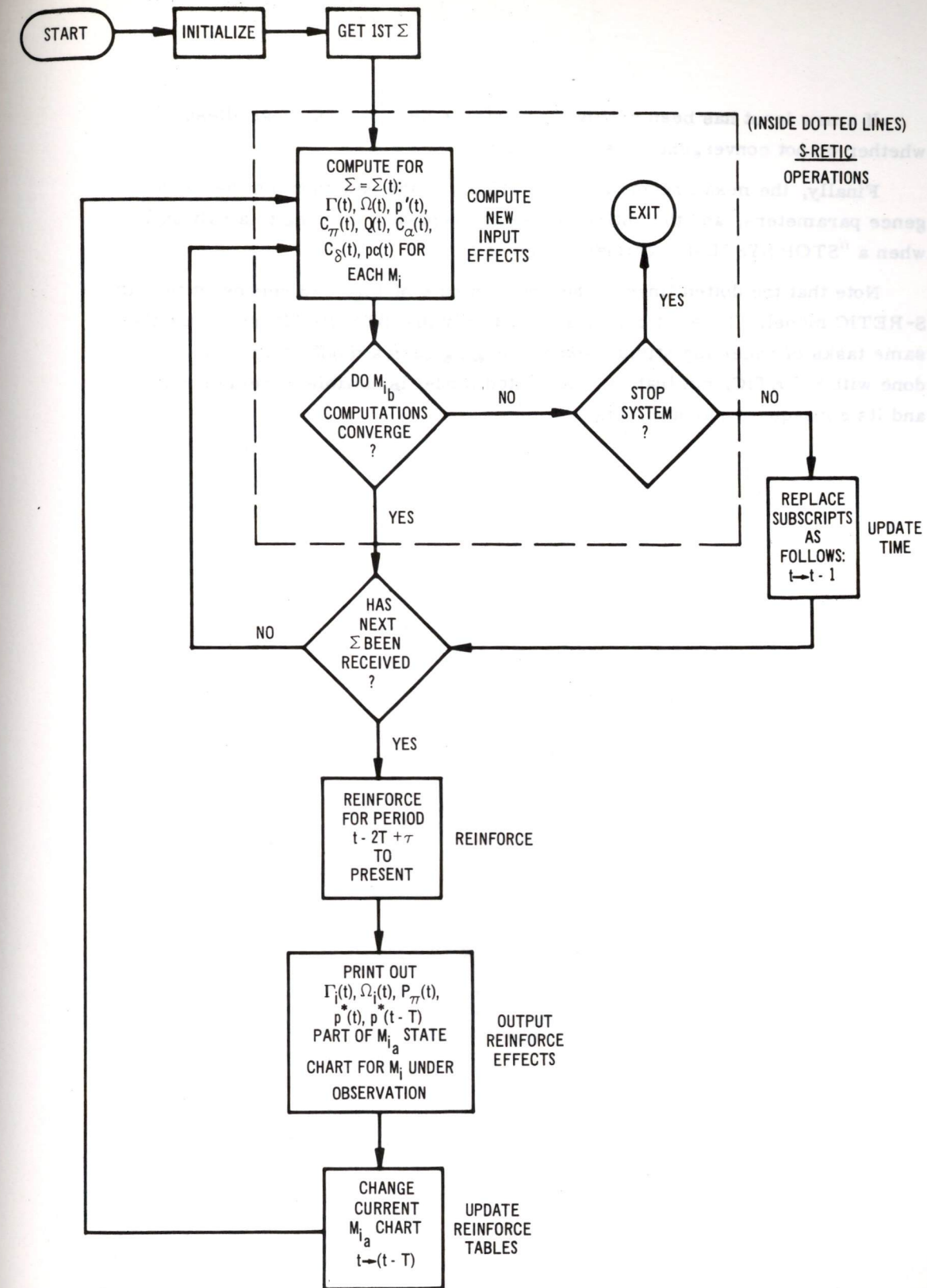
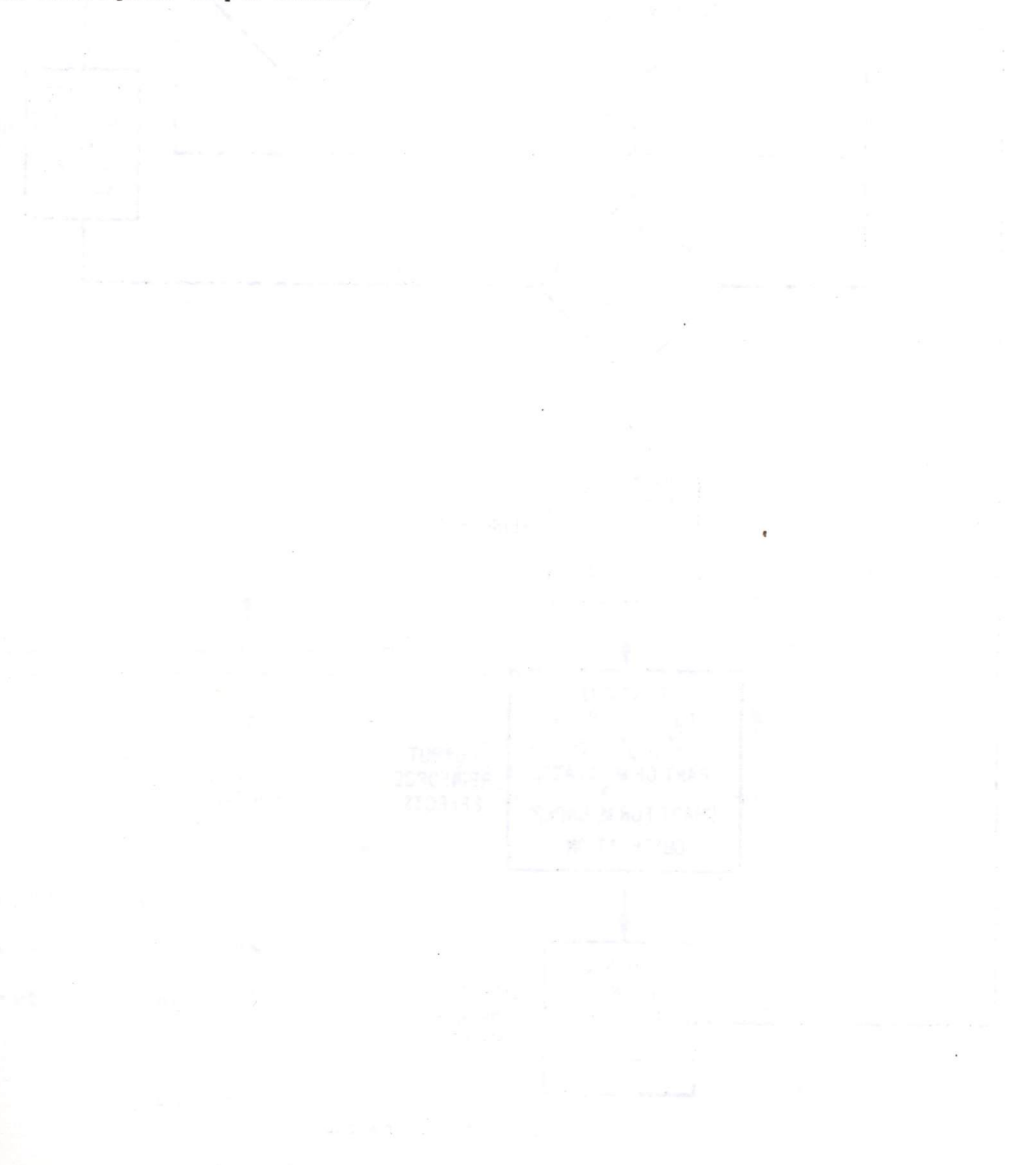


Fig. 6-3. STL-RETIC flow chart.

If a new input has been received, reinforcement occurs, regardless of whether or not convergence has been attained.

Finally, the model recomputes the effects of the new input on the convergence parameters, and the entire process is repeated, coming to a halt only when a "STOP SYSTEM" command is given to it.

Note that the dotted lines in the upper middle of Fig. 6-3 represent the old S-RETIC model. Thus, it can be seen not only that STL-RETIC performs the same tasks of receiving stimuli and converging upon a mode of activity as were done with S-RETIC, but that the augmented model is capable of reinforcement and its consequent output effects.



PART II

RESEARCH INTO THE NATURE OF ANIMAL SENSORY,
DECISION AND CONTROL SYSTEMS

SECTION 7

ANIMAL COMPUTATION

by

Louis Sutro and David Lamport

with the technical assistance of Warren McCulloch

7.1 Background to a Computer Model of a Vertebrate Brain

From his life-long study of the human nervous system, Dr. Warren McCulloch has concluded that the essential properties of human computation must serve as a basis for the electronic acquisition and storage of human knowledge. Although as a neurologist, psychologist, and physiologist he is aware of the dangers involved in embodying mental functions in physical devices, he has nevertheless developed a simplified model of a vertebrate brain. His intention is merely to suggest an organizational hierarchy necessary for efficient robot performance.

Figure 7-1 outlines his model of the vertebrate nervous system, identifying what he feels are five principal computational areas and their functional connections. At the left is the retina, consisting of three layers of cells, two of which are engaged in computation. The eye is shown as representative of the senses because its computational capacity qualifies it as a principal computer; it is the foremost data source to the primate brain, providing two million of its three million inputs.

At the upper left is the cerebrum, which Dr. McCulloch calls the "great computer" and in which computation is carried out in many layers; each of which, in man if unfolded, would be about the size of a large newspaper.

The computer which controls all others is shown at the center right. It is the reticular core of the central nervous system and extends from the base of the brain through the spinal cord. It makes the major decisions, or rather, shifts the focus of attention so as to determine what is to be done from moment to moment. By deciding which command function is to take over, it thus controls all other computers and, through them, the whole organism. A model of this computer was presented in Refs. 3 and 4 and further developed in Section 6 of this report.

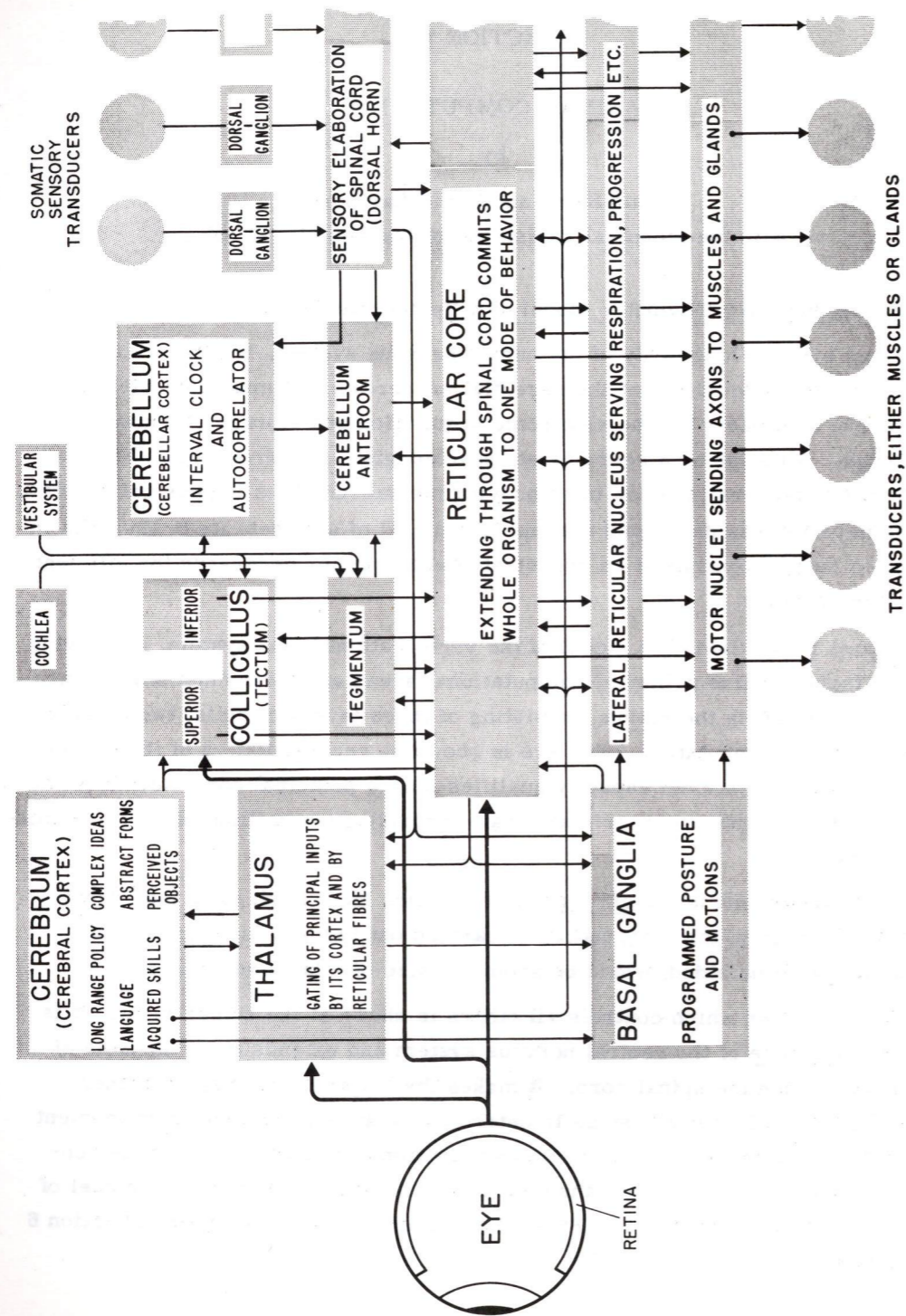


Fig. 7-1. Block diagram of generalized vertebrate nervous system.

Clusters of nerve cells at the base of the cerebrum comprise the basal ganglia, a computer shown at the lower left of the figure. Here are programmed all innate or learned movements, such as feeding, walking or throwing a ball. Additional programs may be acquired through the growth of connections to the motor control nerve cells, shown along the bottom of the illustration.

Completing the list of principal computational areas is the cerebellum, shown at the top of Fig. 7-1. It programs the completion of a movement, such as reaching to touch an object, and requires inputs both from the vestibular system, to detect tilt and acceleration of the head, and from skin and muscle sense cells, to detect posture and information as to what is being touched.

Interconnected with the principal computers are switching structures, such as the thalamus, colliculus, and cerebellum anteroom. In lower vertebrates (fish, amphibians, birds), the colliculus perceives form and movement; in higher animals, it is subdivided into a superior colliculus, which determines the direction of gaze under the general control of the cerebral cortex, and an inferior colliculus, which is probably concerned with auditory and vestibular inputs as well as the somatic body image. The base around the colliculus is the tegmentum.

Around the reticular core are specialized structures that could also be called computers, such as the nucleus of nerve cells that control respiration and other routine bodily functions, and the dorsal horn of the spinal cord, through which pass inputs from sense cells. Note that the reticular core acts on all other computers and that they report to it. It reaches decisions with the aid of raw data from sensory systems as well as processed data from other computers, the latter of which comprising its primary input source.

The computers of Fig. 7-1 are shown as arranged in animals with horizontal spines. Monkeys and man have the same computers in approximately the same relation, but the arrangement is tilted, with the cerebrum, now very much larger, at the top.

All these computers have a common ancestry. All evolved from the central computer, the reticular core, and in so doing have established only those interconnections necessary for efficient communication with "retic." In this manner the reticular core has been constantly able to maintain the complexity necessary to meet the aggregate demands of the entire system.

7.2 A Possible Engineering Equivalent

Figure 7-2 is a diagram equivalent to Fig. 7-1 but labelled with engineering terms to suggest how the system can be simulated. In place of the retinae are the cameras and the visual first-stage computer.

First-stage computers receive inputs from all of the senses – auditory, vestibular and somatic sensory. Each is called a computer rather than a pre-computer or preprocessor to indicate that it receives feedback from the central computers.

Other substitutions are as follows:

decision computer	for	reticular core
associative computer	for	cerebral cortex
timing, coordinating and autocorrelating computer	for	cerebellum
computer of effector sequences	for	basal ganglia (nucleii)
computer of specialized controls	for	lateral reticular nuclei

7.3 Memory in Biological Computers

If models of the computers described above are to be built, what will be the form of the memory?

Computer designers recognize three principal forms, but there is a fourth more important for our present work. The first form is storage on tape where data is found by searching linearly. The second form is random access where data is found by its address. In the third form, data is found by searching for part or all of its content. This has unfortunately come to be called "associative" memory.

The fourth form is the true associative memory employed in animal brains. Here data is found by a relation. For example, if you are looking for a friend in a crowd, you may find him by recognizing the proportions of his face or the time of day that he passes by. These are relations. You acquire this form of memory by constantly updating a model of the world in your head. A change in the model is considered to be the formation of a trace embodying the relation. (The connection for the trace may have existed but have had a different threshold than that required.) Conditioning, which is the most elementary method of forming trace, is described in Section 9. This form of memory will be simulated, at first, by the other three. Hardware (or chemistry) more appropriate to a true associative memory may evolve.

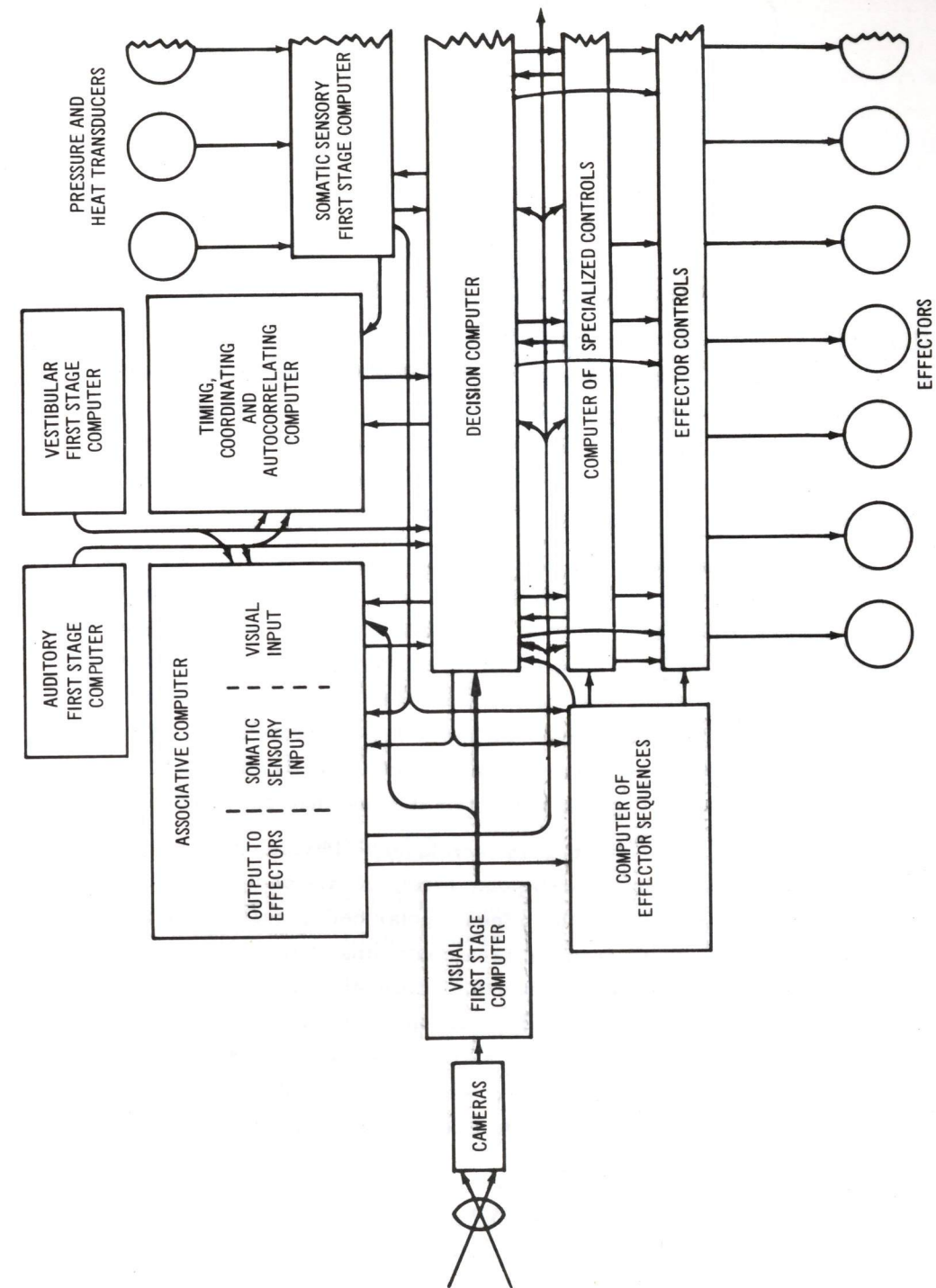


Fig. 7-2. Block diagram employing functional engineering nomenclature.

SECTION 8

VISUAL PROCESSING IN ANIMALS

by

Roberto Moreno-Diaz

8.1 Development of a Computational Model

Existing models of visual processing in the frog's retina and tectum (Refs. 17, 21, 22) suggest a concept which may be applied to retinas of higher animals as well. It is hoped that a model of this concept will soon be realized which, founded in known anatomy and physiology, may be further extended to apply to tasks such as those described in Section 4 of this report.

The concept of visual processing being developed may be regarded as a layered computation; that is, elements of a similar nature comprise a given layer, and computation occurs first through interconnections between elements in the same layer, then between those of different layers. In a gross sense, a succession of such layers will come to represent the higher animal retina. Although there are only three significant cell types in vertebrate retinæ (photoreceptor, bipolar, ganglion), the number of computational layers is larger, since interaction may occur between dendrites at more than one level.

What is the nature of computation in each layer? Physiology suggests the possibility, at the photoreceptor-bipolar level, of lateral inhibition. Assume a mosaic of photoreceptors which feed bipolar cells, such that a particular bipolar receives inputs from an area containing many photoreceptors by means of photoreceptor ends, or "fibers." Lateral inhibition may then be formulated as follows. Consider a typical bipolar cell, called "B" in Fig. 8-1. It receives both an excitatory signal from the photoreceptor - or photoreceptors - situated directly above it as well as inhibitory signals from the surroundings. If the effect of the inhibition is subtractive, the total signal arriving at the bipolar B would be

$$B = \sum_{A_1} F_i - \sum_{A_2} F_j \quad (8-1)$$

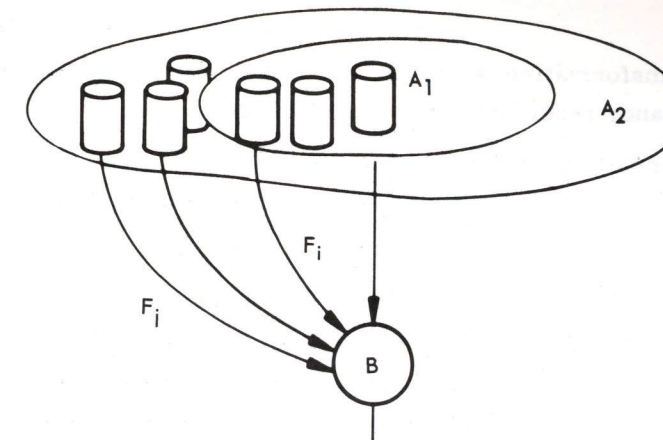


Fig. 8-1. Dendritic level, defined by areas A₁ and A₂, with dendrites leading to cell body B.

where F_i is the output of a photoreceptor transmitting an excitatory signal, F_j the output of an "inhibitory" photoreceptor, and A₁ and A₂ the areas of excitatory and inhibitory photoreceptors respectively.

Equation (8-1) refers to interaction only between the photoreceptor-bipolar levels. Regardless of the location of such interaction in the retina, it is possible to postulate a corresponding layer which produces the same effect. Let F(α, β) be the output of an ideal photoreceptor situated at the point (α, β) (in a suitable coordinate frame superimposed over the retina). This corresponds to the input of an idealized computational layer, called a layer of lateral inhibition. Let B(x, y) be the output of that layer corresponding to the point (x, y). The transformation rule that defines the computation at that layer is then

$$B(x, y) = \frac{1}{N_1} \sum_{(\alpha, \beta) \in A_1} F(\alpha, \beta) - \frac{1}{N_2} \sum_{(\alpha, \beta) \in A_2} F(\alpha, \beta) \quad (8-2)$$

where summation is over all points (α, β) in areas A₁ and A₂, and N₁ and N₂ correspond, respectively, to the number of elements in A₁ and the number in A₂.

The values of the coordinate variables x, y, α, and β, according to the discrete nature of the retinal mosaic, are first considered discontinuous, taking, for example, values 1, 2, 3, If the area A₁ is reduced to a single element, that at position (x, y), Eq. (8-2) becomes

$$B(x, y) = F(x, y) - \frac{1}{N_2} \sum_{A_2} F(\alpha, \beta) \quad (8-3)$$

If the elements of the retina are sufficiently small and numerous, the summation may be replaced by an integral, to give

$$B(x, y) = F(x, y) - \frac{1}{A_2} \iint_{A_2} F(\alpha, \beta) d\alpha d\beta \quad (8-4)$$

The effect of this transformation is known to produce contrast enhancement (Ref. 23) and redundancy reduction of the initial image.

In the above formulation, variables B and F are considered time independent. This restriction will be eliminated in future modeling to take time into account.

8.2 Generalization of Interaction Coefficients

Equation (8-2) defines computation over specific areas, giving equal weight to each photoreceptor in a field. That equation may be generalized by introducing a set of coefficients which reflect the relative strength of a particular input fiber connected to the output fiber under consideration. Let $C(x, y, \alpha, \beta)$ be the strength of the input in position (α, β) over the output in position (x, y) . Equation (8-2) then becomes:

$$B(x, y) = \sum_R C(x, y, \alpha, \beta) F(\alpha, \beta) \quad (8-5)$$

where the summation is over the entire retina.

That Eqs. (8-2 and 8-3) are particular cases of Eq. (8-5) is easily verified. To obtain Eq. (8-2), let $C(x, y, \alpha, \beta)$ be $+1/N_1$ over the area A_1 , $-1/N_2$ over A_2 , and zero elsewhere; Eq. (8-2) is then the consequence. Let $C(x, y, \alpha, \beta)$ be $+1$ for $C(x, y, x, y)$, $-1/N_2$ over A_2 , and zero elsewhere; the result is Eq. (8-3).

Equation (8-5) is called the general expression for lateral interaction, of which lateral inhibition is a specific case. In the continuous retina, Eq. (8-5) becomes

$$B(x, y) = \iint_R C(x, y, \alpha, \beta) F(\alpha, \beta) d\alpha d\beta \quad (8-6)$$

Theoretical work continues at present to investigate various forms of the weighting coefficients to obtain different processes. It has been found that Eq. (8-6) explains several properties of the visual system - such as reliability, in the sense that the destruction of a fiber $B(x, y)$ or a group of them, produces not blind spots but merely a general loss of resolution. Such points will be examined in detail in future reports.

Computations as expressed in the above equations are not easy to carry out analytically in any but the most simple forms of $F(\alpha, \beta)$ - which are not the most interesting ones, but clues about the behavior of the transformation have been obtained analytically through treating the simple one-dimensional case. For a real image in the retina, however, the form of $F(\alpha, \beta)$ is extraordinarily complicated and requires a simulation of the kind presented in Section 4.

8.3 Examples of Lateral Interaction

For the one-dimensional case, Eq. (8-4) becomes

$$B(x) = F(x) - \frac{1}{2s} \int_{x-s}^{x+s} F(x) dx$$

where $2s$ is the size of the inhibitory area, which in this case is a line. Statisticians call the second term a "moving average". $B(x)$ may be regarded as the result of a transformation T_s upon $F(x)$. The transformations of some simple functions are:

a) When $F(x) = Ag(x)$, with $A = \text{constant}$, then the transformation is:

$$T_s(Ag(x)) = AT_s(g(x)).$$

Proof:

$$\begin{aligned} T_s(Ag(x)) &= Ag(x) - \frac{1}{2s} \int_{x-s}^{x+s} Ag(x) dx \\ &= A \left[g(x) - \frac{1}{2s} \int_{x-s}^{x+s} g(x) dx \right] \\ &= AT_s(g(x)) \end{aligned}$$

b) When $F(x) = Ag(x) + Bh(x)$, with A, B constant, then $T_s(F(x))$ is:

$$T_s(Ag(x) + Bh(x)) = AT_s(g(x)) + BT_s(h(x)) \text{ as suggested by the above proof}$$

c) For $F(x) = A$, with $A = \text{constant}$,

$$T_s(A) = 0$$

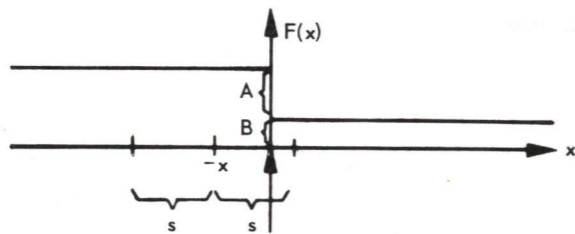
d) For $F(x) = Ax$, with $A = \text{constant}$,

$$T_s(Ax) = 0$$

Proof of d:

$$\begin{aligned} T_s(Ax) &= Ax - \frac{1}{2s} \int_{s-x}^{s+x} Ax dx \\ &= Ax - \frac{1}{2s} \left[\frac{A(s+x)^2}{2} - \frac{A(s-x)^2}{2} \right] \\ &= Ax - \frac{A}{4s} [s^2 + 2sx + x^2 - s^2 + 2sx - x^2] \\ &= Ax - \frac{A}{4s} [4sx] \\ &= 0 \end{aligned}$$

e) If $F(x)$ has a step discontinuity:



For x negative, the transformation of the above function is:

$$T_s(F(x)) = A + B - \frac{1}{2s} \int_{x-s}^0 (A+B) dx - \frac{1}{2s} \int_0^{x+s} B dx$$

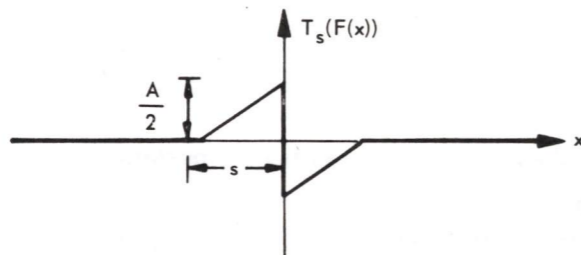
$$= \frac{A}{2} \left(1 + \frac{x}{s}\right) \quad -s \leq x < 0$$

For x positive:

$$T_s(F(x)) = B - \frac{1}{2s} \int_{x-s}^0 (A+B) dx - \frac{1}{2s} \int_0^{x+s} B dx$$

$$= \frac{A}{2} \frac{x}{s} - 1 \quad 0 < x \leq s$$

Plotting T_s over the entire domain of x ,

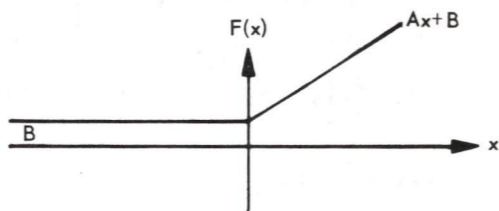


For $x < -s$ or $x > s$,

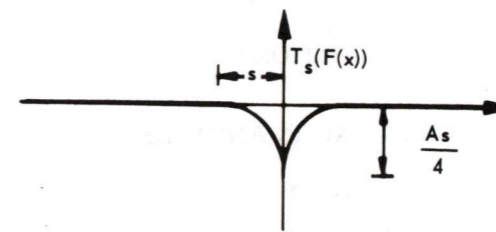
$$T_s(F(x)) = 0$$

because the function is a constant.

f) If $F(x)$ is a ramp



then T_s is



SECTION 9

LEARNING IN ANIMALS

by

Charles Sigwart

9.1 General

At the end of Section 7 memory traces in the nervous system were mentioned as the embodiment of relations comprising a model of the environment. The possibility of changing such a model with new environmental situations implies memory and learning. Attempts to define learning and memory have occupied the efforts of many neurophysiologists and psychologists for a long time. Thus there exists a body of literature concerning mathematical learning theory based on behavior and a range of hypotheses about the relation of learning and memory to structural, chemical and functional properties of the nervous system.

Learning is a term which encompasses a range of phenomena extending from simple changes in responsiveness to a stimulus to the very complex capacity for dealing with abstractions possible in man. If we can accept a broad generality, "We can define learning as that process which manifests itself by adaptive changes in individual behavior as a result of experience..."⁽²⁴⁾

Learning cannot be defined more precisely since mechanisms underlying many complex learning phenomena are not understood. The more elementary forms of learning such as habituation and conditioning have been subjected to neurophysiological study. In the following paragraphs of this section various concepts relevant and desirable in the behavior of a robot will be described. A more detailed discussion of recent research on learning may be found in Ref. 25 and Ref. 26.

9.2 Classes of Response

The decision computer's responses to stimuli may be classed as either momentary or predictive. Those which are momentary are a consequence only of the stimuli which they directly follow; those which are predictive occur when the decision computer receives a stimulus recognized to be a precursor of a later event and acts in accordance with that event. STL-Retic will implement

momentary response properties such as habituation and classical conditioning as well as predictive responses by incorporating operant conditioning and generalization. Implicit in providing the capacity to change response patterns is the ability to modify generalizations when they become inappropriate.

9.3 Classical Conditioning

Classical conditioning (also called Pavlovian and Type I conditioning) involves the association of a reflex response with a new stimulus. For this to occur the new stimulus which is to become a conditioned stimulus (CS) must regularly precede the unconditioned stimulus (US), which normally produces the unconditioned response (UR). It is assumed that the 'stimulus' may include many factors in a vertebrate animal.

"Sensory signalling that takes place during conditioning involves at least two differentiable sensory channels: specific and nonspecific. The nonspecific channel apparently establishes background conditions for widespread reception and retention of particular sensory and motor signals which may enter by way of specific and nonspecific channels... The net result of all this is an alteration in distribution and influence of incoming and outgoing signals which increases the likelihood that certain responses will occur in relation to certain previously experienced stimuli."⁽²⁷⁾

A working definition for classical conditioning is as follows:

"A stimulus class (CS) that before the conditioning procedure does not produce a response, or produces it with low incidence will, after the conditioning procedure, produce the response with a frequency above some specified criterion. The conditioning procedure consists in pairing of the CS with another stimulus (US) that ordinarily produces an unconditioned response (UR). The paired presentations will be in a certain order (CS - US) with a CS-US interval with specified limits both with respect to mean duration and variability. The CR must meet criteria of resemblance to the UR."⁽²⁸⁾

The production of new relations between a stimulus situation and a response may be considered as a memory allowing the building up of a class of stimulus situations to which a response is elicited.

9.4 Habituation

Habituation is "...the relatively persistent waning of a response as a result of repeated stimulation which is not followed by any kind of reinforcement. It

is specific to the stimulus and relatively enduring... distinct from fatigue and sensory adaption. ⁽²⁹⁾ It is assumed to be a central phenomenon, exclusive of receptor or effector fatigue.

The main function of habituation is to eliminate insignificant responses by reducing sensitivity to a specific stimulus which is repetitive. The man who hears a jack hammer going in the street stops paying attention to it after a while. Similarly the decision computer should ignore stimuli which it finds unimportant. When the robot does something new in its environment, or when that environment changes, such stimuli will occur and the decision computer should eradicate their traces as they enter over sensory subsystems.

9.5 Discriminative Learning

The capacity to generalize, given a set of stimuli appropriate to a particular response, so as to include in that set related stimuli, is known as discriminative, or perceptual, learning. It is directly involved with conditioning and habituation, in that effective response to an appropriate stimulus requires correlated sensory data.

Since changes resulting from generalization are reversible, those that are inappropriate should be unlearned as soon as a contradiction arises between results of behavior and expectation. Thus, the capacity to generalize depends on perceptual discrimination as well as information from the senses.

9.6 Instrumental Conditioning

Instrumental conditioning (also known as operant or Type II conditioning or trial and error learning) involves the modification of a response pattern dependent on a reinforcement. This is in contrast to classical conditioning, where reinforcement shapes the selection of the appropriate stimulus. Instrumental conditioning and its variants involve modification of a response that was initiated by the animal. In essence the animal is changing the probability of a particular response to a given situation consequent upon the reinforcing result, whether painful or pleasurable. Thus this type of conditioning includes avoidance conditioning, where a potential hazard is recognized and elicits a response to avoid it.

It is clearly important for the robot to eliminate undesirable responses to hazards. A task of the decision computer must be to seek other responses to lessen hazards in situations it encounters.

SECTION 10

SUMMARY AND CONCLUSION

The search for evidence of life and geological changes on Mars we interpret as a problem in reducing the data collected initially to a smaller amount of significant data to be sent to Earth, and of changing the mode of operation of the data-acquiring device (robot). Reducing the amount of visual data is called visual processing and modelling; changing the mode of operation, decision. We aim by data reduction to employ the limited transmitted information more efficiently than by conventional TV. By permitting the robot to change its mode of operation we aim to enable it to decide what to do next.

Components have been described which, when tied together, will comprise the visual subsystem of a robot. The computer-controlled television camera is sufficiently open in its design to permit analysis and resynthesis as demands upon it change. A succession of stereo-optics designs, from C1 through D1 to E1, has led back to a type with fixed convergence, called C2, as the one most likely to be developed next. Means have been devised of processing stereo pairs of pictures to extract range. Two structures have been planned to store a model of the environment, one employing points, the other points and reflectances. Data reduction is achieved first in the formation of a model and second in the selection of parts of that model (interrogation) for transmission to earth. One purpose of this interrogation is to search for evidence of life and geological changes. Another is to find paths that the robot can follow.

The contact subsystem is a hand-and-arm and a means of modelling the environment contacted. The hand is being planned at present both to feel the shape of objects and to provide a signal, such as a flashing cross, on which the cameras can converge at the start of the exploration of a scene.

The decision subsystem, previously devised to respond to a sequence of static environments, is now being augmented by logical and storage elements so that it can learn. The structure of this subsystem is such that while it is now in the form of a computer program it can later be implemented in special-purpose hardware.

The mobility subsystem is at present, a vehicle, wheel drives and controls, originally designed for a lunar landing by the General Motors Defense Research Laboratory and since adapted by them for a Martian landing. The synthesis of visual, contact, decision and mobility subsystems we call a robot. Three configurations of this robot have been pictured, the preferred one having the camera assembly directly above the front axle of the vehicle. All are capable of mapping by measuring distances with an odometer and angles with a radio direction finder aimed at transponders previously dropped by the robot. Accuracy can be improved by sighting landmarks.

While this program is expected to be primarily one of computer development, only commercially-available computers are used at present. The purpose of this is to determine by simulation on existing machines the specifications of the desired machine.

The first part of this report concerns the developments just summarized. The second part describes research into the organization of animal nervous systems, into the nature of visual processing and the nature of learning in animals. The organization of animal nervous systems is that found by Warren McCulloch and presented here in computer language. Visual processing is that found by physiologists and presented here analytically by Roberto Moreno-Diaz. "Learning in animals" is a summary of this subject as presently understood in psychology and physiology. These investigations will be the basis of further development of the kind described in the first part of the report.

APPENDIX A
MEASUREMENT OF VIDICON RESOLUTION

by

David Tweed

The method employed in the measurement of vidicon resolution is that described in Ref. 25. Figures A-2 and A-3 in that reference are replaced respectively by a photograph of an oscilloscope presentation of the output of the C1-1 camera chain and a graph derived from this presentation. Quoting from Ref. 25:

"A new type of test pattern was devised (Fig. A-1) which consists of nine line-groups; each of these groups contains a black and white bar to represent 100% response factor, followed by ten sets of four black and three white lines which represent 100 to 1000 TV lines (across the width of the chart), in 100-line increments.

"This new chart makes it possible to obtain data for a complete square-wave aperture response curve with one oscilloscope presentation. Figure A-2 is typical of the type of presentation that is obtained from an oscilloscope that is fed the video signal and set to select one horizontal scan line when an image from the line selector pattern is being transduced.

"The procedure for obtaining the aperture response curve is as follows:

1. The tube is set up under the desired test conditions, in this case optimum overall performance.
2. The test pattern is the "line selector" pattern of Fig. A-1.
3. The video signal from a selected horizontal line is presented on an oscilloscope and a photograph is taken. See Fig. A-2.
4. The photograph is measured for the relative response of the line-sets to that of the black and white bars, and a curve is plotted. See Fig. A-3."

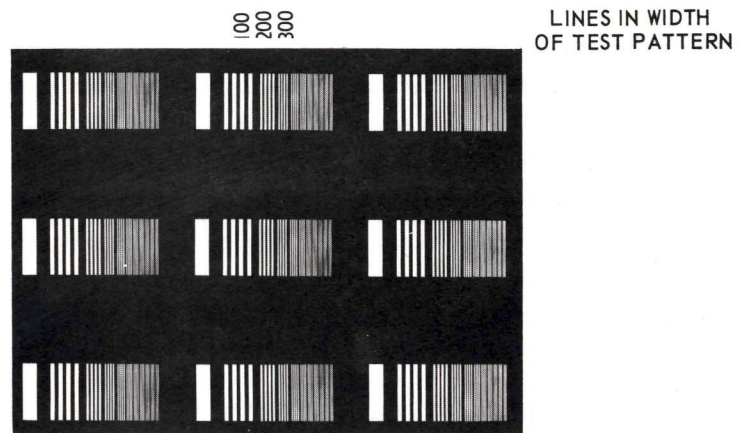


Fig. A-1. "Line Selector" test pattern (Westinghouse resolution chart ET-1332 purchased from Tele-Measurements Inc., 145 Main Ave., Clifton, N.J.), reproduced 1/3 full size.

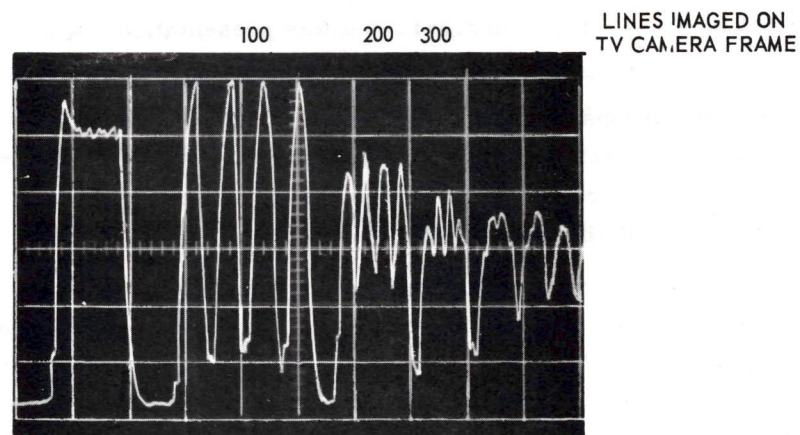


Fig. A-2. Tracing of oscilloscope presentation of video signal from part of single scan line of Fig. A-1, employing the amplifier described in subsection 3.5.

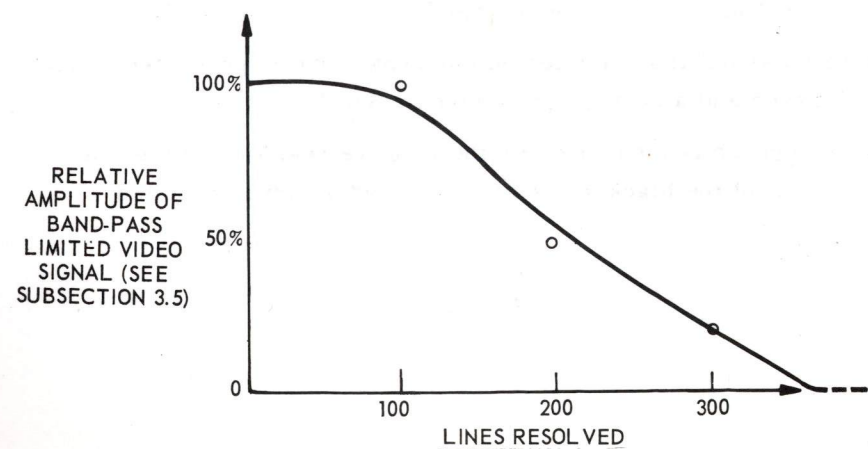


Fig. A-3. Square-wave aperture response curve derived from Fig. A-2.

Note to Section 3

APPENDIX B

RANGING AND RANGE ERROR

by

Charles Sigwart

B.1 General

To obtain the range of objects in the environment it is possible to use a stereo optical system and in essence triangulate. Thus the range z to point P is determined by measuring α_1 and α_2 , knowing that the base separation of the two optical systems A_1 and A_2 is $2b$. The measurement of range is uncertain to the extent that the measurement of convergence angles α_1 and α_2 is uncertain. In general there will necessarily be some small angular uncertainty $\pm\delta$ in measuring the orientation of the optic axis. Thus the localization of P by the method is in a volume bounded by the intersection of the two cones whose axes intersect at P . The cones have a half angle δ as shown in the plane of the axes in Fig. B-1. The intersection of the plane of the optic axes with this volume produces the quadrilateral of uncertainty surrounding P .

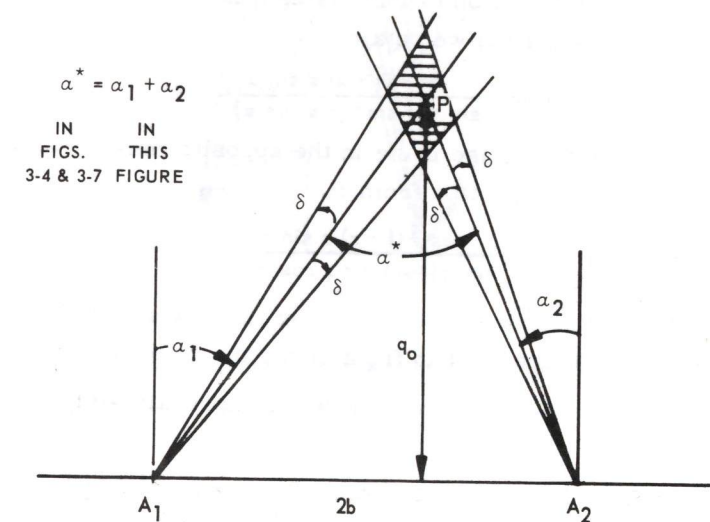


Fig. B-1. Quadrilateral of uncertainty surrounding point P when viewed by cameras at A_1 and A_2 .

This zone of uncertainty varies with α_1 and α_2 as indicated in Fig. B-2. Thus it is apparent that ranging on objects along the line q_0 away from the z axis results in an increasing uncertainty of measurement and that the range error is minimized by restricting the convergence angles such that $\alpha_1 = \alpha_2$. The effect shown in Fig. B-2 is of course amplified by the large angular uncertainty δ shown, but the general result is that range uncertainty increases significantly with deviation from the condition of equal convergence angles. Thus the following discussions will deal with range uncertainties for points on the center line.

B.2 Range and Uncertainty for Points on the Q Axis

To keep the range uncertainty minimal as indicated above, we will consider only the range of points on the Q axis, employing equations derived by Roberto Moreno-Diaz and David Tweed. The range of an object at P shown as OP in Fig. B-3 is determined from

$$q = \frac{b}{\tan \alpha} \quad (B-1)$$

The range is measured from the midpoint of a line of length $2b$ connecting the yaw axes of the imaging system. Each yaw axis intersects an optic axis at point A distance x from the image plane. The convergence angle α is measured from the forward direction (parallel to Q). N is the nodal point of the lens.

Considering the similar triangles in Fig. B-3, $\triangle MCN \sim \triangle SEN$, we see that $MC/CN + ES/EN$ or

$$\frac{q_0 + \Delta q - (f-x) \cos \alpha}{b - (f-x) \sin \alpha} = \frac{f \cos \alpha + s \sin \alpha}{f \sin \alpha - s \cos \alpha} \quad (B-2)$$

The range uncertainty in the direction of increasing q is denoted $+\Delta q$, shown as PQ. Thus from Eqs. (B-1 and B-2) we obtain

$$+\Delta q = \frac{bs - (f-x)s \sin \alpha}{\sin \alpha (f \sin \alpha - s \cos \alpha)} \quad (B-3)$$

Similarly a displacement in the image plane in the opposite sense, $-s$, gives the uncertainty in range in the $-Q$ direction from P.

$$-\Delta q = \frac{-bs + (f-x)s \sin \alpha}{\sin \alpha (f \sin \alpha + s \cos \alpha)} \quad (B-4)$$

The second term of the numerator is minimal when the yaw axis is through the point on the optic axis a distance $x = f$ in front of the image plane, then $(f-x) = 0$.

It can be shown that range uncertainty increases approximately with the square of the range.

Substitutions into Eqs. B-3 and B-4, noting that $\alpha = \tan^{-1} b/q_0$, using values of q_0 (from 0 to 3 meters at intervals of 0.5 meter), b and f from the text and Δs from Table 3-2, leads to the range uncertainty of camera assemblies D1 and E1 tabulated in Table B.1 and plotted in Figs. 3-5 and 3-11.

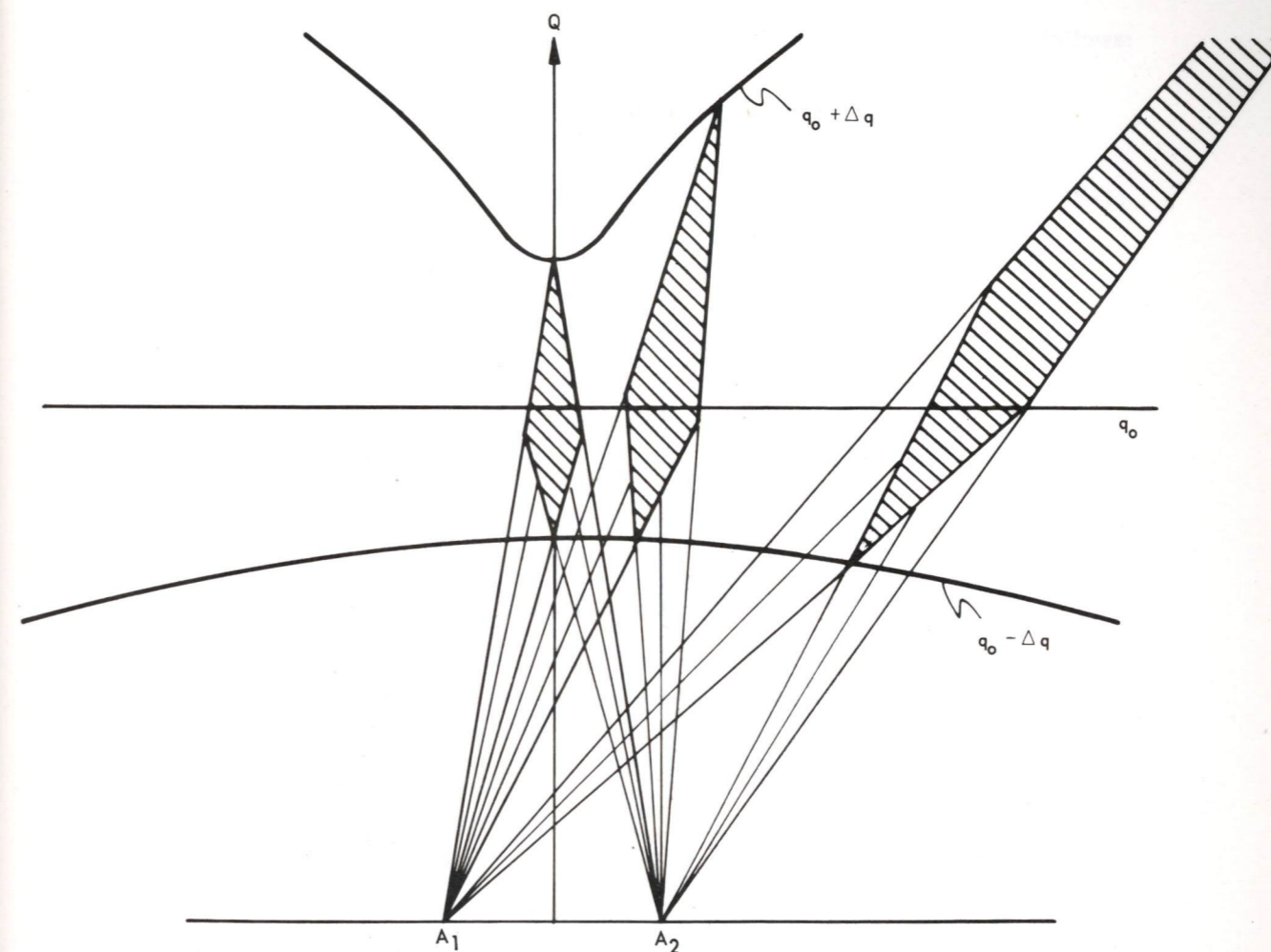
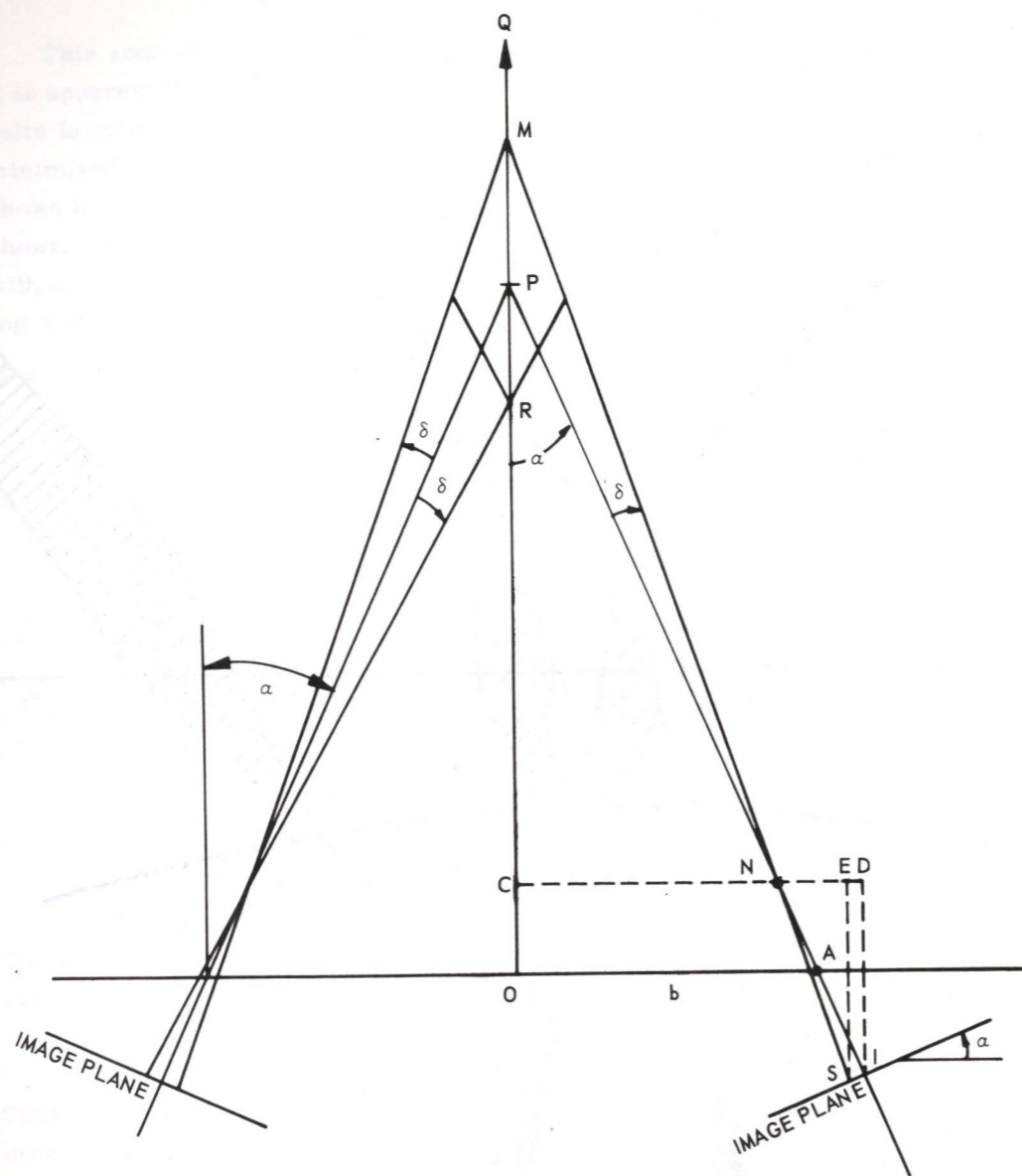


Fig. B-2. Quadrilaterals of uncertainty for points on the line q_0 .



SYMBOL DEFINITIONS

- | | |
|---|--|
| A = POSITION OF YAW AXIS OF CAMERA | IS = s, A DISPLACEMENT ON THE IMAGE PLANE EQUIVALENT TO ANGLE δ |
| AI = x, SEPARATION OF IMAGE PLANE AND YAW AXIS ALONG THE OPTICAL AXIS | PO = q |
| AO = b, ONE HALF OF SEPARATION OF YAW AXES | PM = $+\Delta q$ |
| I = POINT ON IMAGE PLANE PIERCED BY OPTIC AXIS | PR = $-\Delta q$ |
| IN = f, FOCAL LENGTH OF LENS SYSTEM | α = CONVERGENCE ANGLE |
| IP = OPTIC AXIS | δ = UNCERTAINTY OF ANGLE α |

Fig. B-3. Geometry of variable convergence optics for range finding.

B.3 Systems with Parallel Optic Axes

When the optic axes are held parallel to the Q axis, as in camera assembly C1, $\alpha = 0$, and an object at a point P on the Q axis will be represented by corresponding points S in the image planes. Uncertainties in the position angle $\pm \delta$ will produce equivalent displacements $\pm \Delta s$ in the image plane.

The range q of a point P, as seen in Fig. B-4, may be determined as follows:

Since $\Delta MOS \sim \Delta NAS$, $\frac{f}{s} = \frac{q}{b+s}$

$$q = \frac{f(b+s)}{s} \tag{B-5}$$

The range $q + \Delta q$ of a point M is

$$q + \Delta q = \frac{f[b+(s-\Delta s)]}{s-\Delta s} \tag{B-6}$$

Subtracting Eq. (B-6) from (B-5) we obtain the range uncertainty Δq due to uncertainty in a position S on the image plane (or equivalently an angular uncertainty δ in determining β):

$$\Delta q = \frac{fb \Delta s}{s(s-\Delta s)}$$

and from Eq. (B-5)

$$s = \frac{fb}{q-f}$$

so that

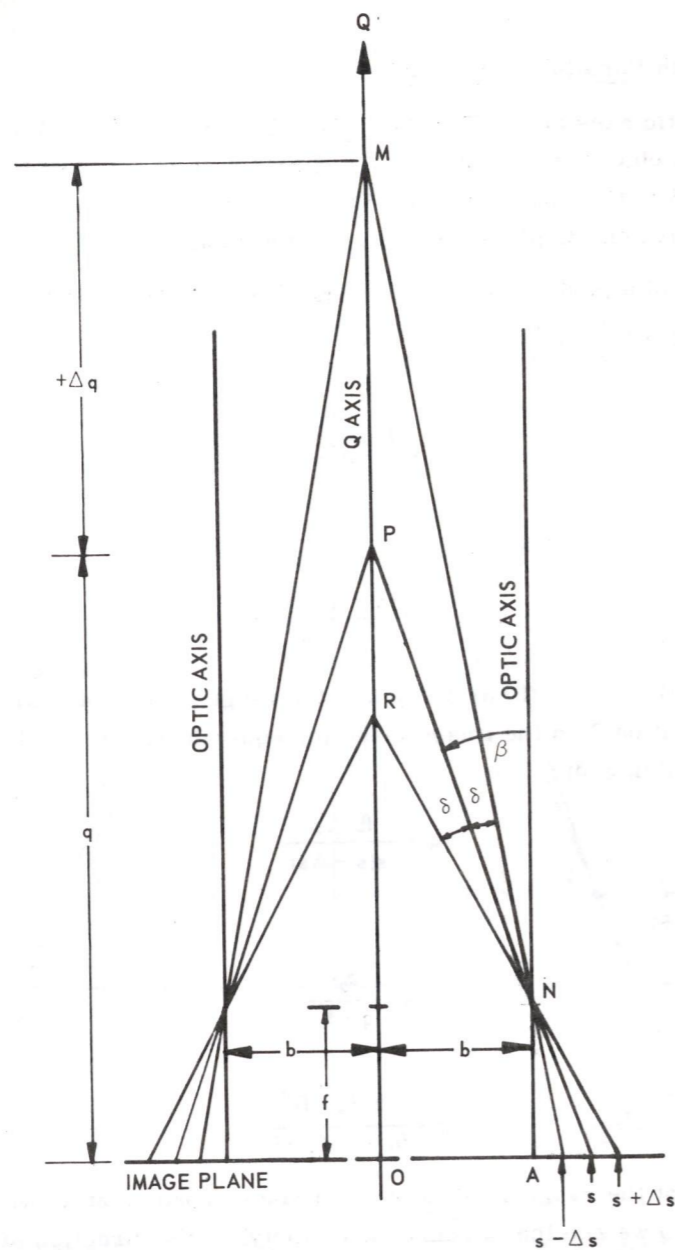
$$\Delta q = \frac{\Delta s(q-f)^2}{fb - \Delta s(q-f)} \tag{B-7}$$

Thus it is seen that the range uncertainty increases approximately with the square of the range. Also we see that a range uncertainty, in the direction of increasing range, corresponds to an uncertainty in s towards the optic axis.

Substitution into Eq. (B-7) of values of q (from 0 to 3 meters at intervals of 0.5 meter), b and f from the text and Δs from Table 3-2, leads to the range uncertainty of camera assembly C1 tabulated in Table B. 1 and plotted in Fig. 3-5.

B.4 Stereo Optics with Fixed Convergent Axes

In a system with fixed convergence the field of view of the lenses will in general be limited. As shown in Fig. B-5, there will be some near point NP on



SYMBOL DEFINITIONS

- q = RANGE OF A POINT P FROM IMAGE PLANE
- A = POINT WHERE OPTIC AXIS PIERCES IMAGE PLANE
- s = DISTANCE ON THE IMAGE PLANE FROM POINT A TO IMAGE OF POINT P
- Δs = UNCERTAINTY IN THE POSITION OF POINT s CORRESPONDING TO THE UNCERTAINTY ANGLE δ
- $+\Delta q = PM$
- $-\Delta q = PR$

Fig. B-4. Geometry of fixed optics for range finding.

Table B-1. Range and range uncertainty for three camera-computer chains (meters).

q_0	C1-1		D1-3		E1 ₁ -3		E1 ₂ -3	
	$+\Delta q$	$-\Delta q$	$+\Delta q$	$-\Delta q$	$+\Delta q$	$-\Delta q$	$+\Delta q$	$-\Delta q$
0.5	.0417	.0366	.0032	.0031	.0048	.0048	.0011	.0011
1.0	.118	.0948	.0137	.0133	.0243	.0231	.0046	.0045
1.5	.384	.2041	.0320	.0306	.0553	.0515	.0104	.0102
2.0	.540	.350	.0578	.0548	.0997	.0908	.0209	.0205
2.5	.898	.494	.0922	.0845	.1321	.1195	.0288	.0280
3.0	1.431	.734	.1346	.1235	.1813	.1617	.0377	.0367
5.0	9.16	1.96	.3743	.3440	.6741	.5901	.1321	.1275

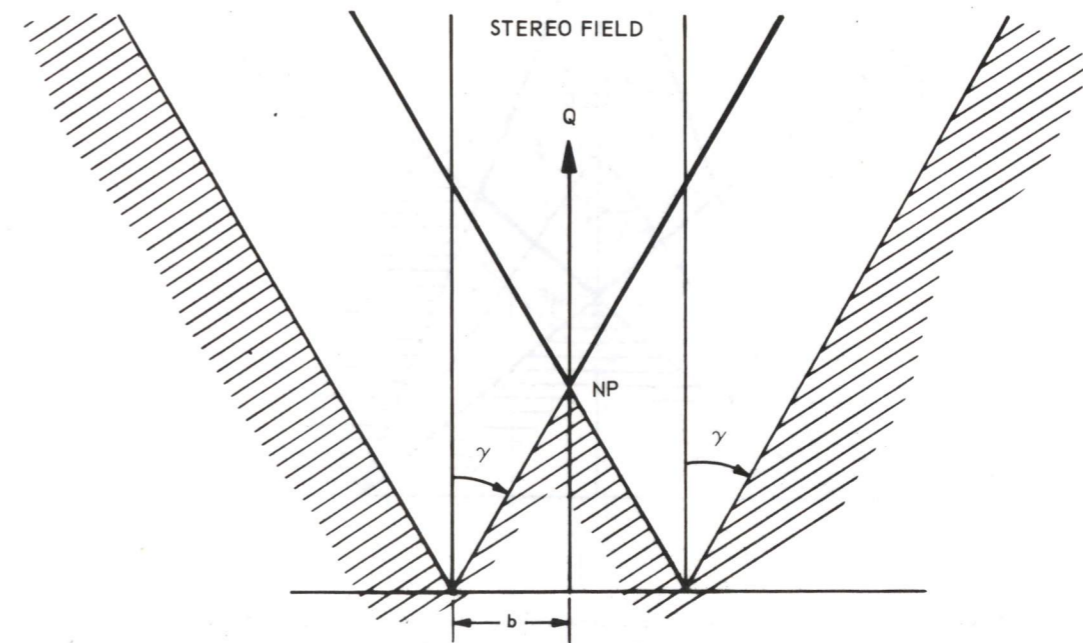


Fig. B-5. Field of view of parallel optical systems. Shaded area is outside view of either system, heavy line bounds area viewed by both.

the Q axis which is the minimum range at which a stereo image can occur. Thus there is a minimum distance at which the range of an object can be determined. This is at the point

$$q_{NP} = \frac{b}{\tan \gamma}$$

where γ is the half angle of the visible field of the lens system. When the optic axes are not parallel ($\alpha \neq 0$) but have fixed convergence a there may also be a maximum point at which stereo images can be obtained, FP in Fig. B-6.

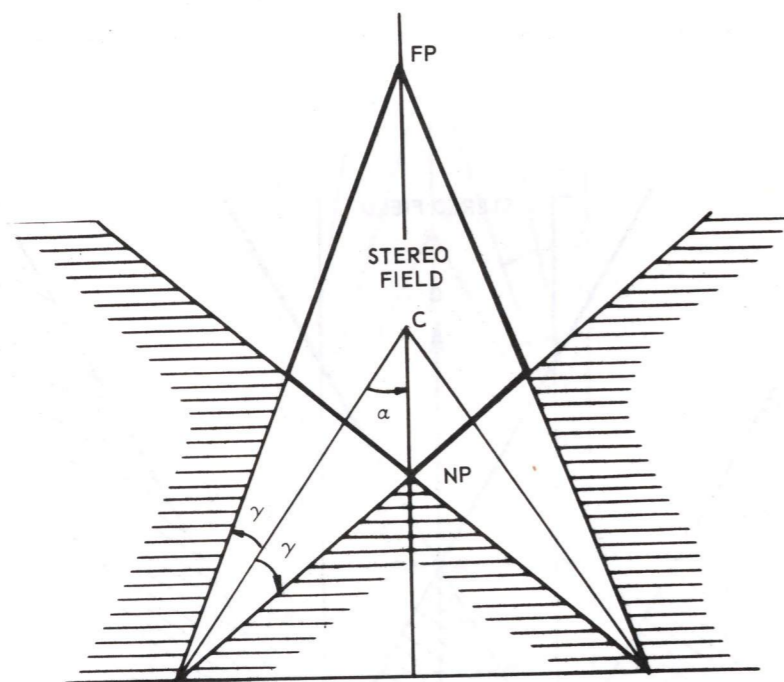


Fig. B-6. Fields of view of optical systems whose axes converge at C. Since $(\alpha + \gamma) < 90^\circ$ there is a point FP which is the maximum point for a stereo image. Area outside view of either system is shaded. Stereo field is bounded by heavy line.

Note to Section 3

APPENDIX C

OTHER STEREO CAMERA CONFIGURATIONS

by
Joseph Convers

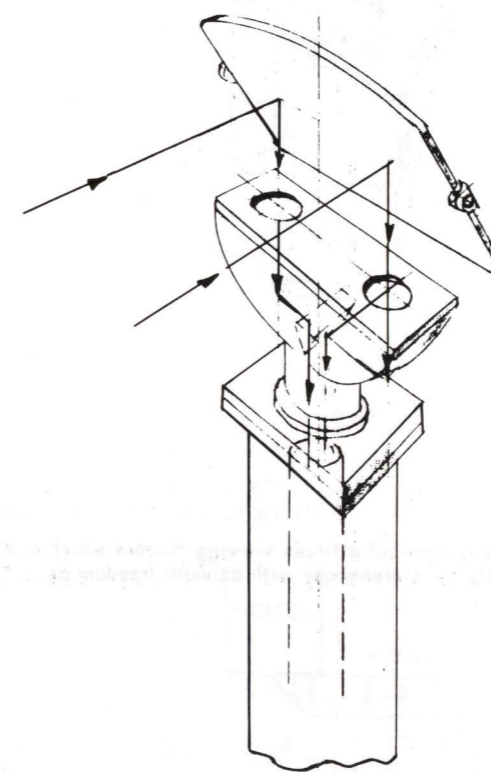


Fig. C-1. Single vidicon camera in vertical attitude receiving information from a single nodding mirror and giving 1/2 of surface to each stereo axis.

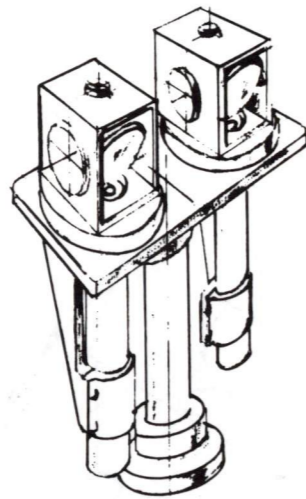


Fig. C-2. Two vidicon cameras in vertical attitude viewing mirrors which nod together and rotate individually for stereoscopy with azimuth freedom as in Figs. 3-6 and 3-10.

APPENDIX D

PROPOSED OPTICAL TRAIN FOR CAMERA E1

by

Robert Magee

The proposed optical train (Fig. D-1) is a dual system combining a wide-field camera with a high resolution but smaller-field camera. Both systems share the same objective lens by means of a beam splitting element. The physical size of the composite focal plane fits within a vidicon tube format size of approximately 0.5 inches by 0.5 inches.

The wide angle objective lens consists of a scaled version of the Minolta MC-Rokkor PF lens. This lens is designed for a speed of $f/1.4$ and is a form of the Zeiss Biotar. The focal length has been reduced from 58 mm in the Minolta design to 40 mm for Camera E1. One particular advantage of the lens is its relatively long back focal length. This lens-to-focal-plane distance is necessary in the single lens reflex camera with which the Rokkor PF is used. Since the optical arrangement of Camera E1 poses design problems similar to those encountered with a single lens reflex, a logical choice for investigation was the form of lens used in the reflex camera.

WIDE ANGLE FIELD IS 20° BY 10°
HIGH RESOLUTION FIELD IS 2° BY 1°

DRAWING: TO SCALE

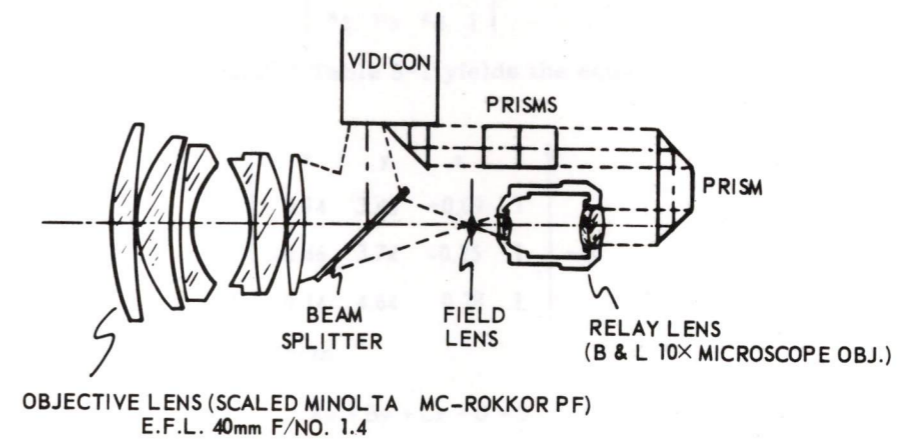


Fig. D-1. Proposed optical train for camera E1.

Light rays forming the wide angle image at the vidicon reflect from a beam splitter located to the right of the objective lens. Most of the light, however, is transmitted through the beam splitter and imaged at a field lens. A small central portion of this image is then relayed to the vidicon by means of a ten power microscope lens. This produces a ten to one scale relation between the two field images at the vidicon. The numerical aperture of the microscope lens is not adequate to handle all of the rays emanating from the $f/1.4$ objective lens and it appears that the latter should be stopped down to $f/2$. This lens would probably be used at $f/2$ for the high resolution field in any case because of the slight improvement in image quality.

A number of right angle prisms and mirrors are then used to conduct the light from relay lens to vidicon. The composite field of view at the vidicon then appears similar to that shown in Fig. 3-8. It is important to keep vidicon signal levels from both fields of view roughly equivalent. To accomplish this, the beam splitter is left uncoated. A very thin, or pellicle, form of beam splitter appears advisable to minimize reflection interference from the second beam-splitter surface.

APPENDIX E

CALCULATION OF THE DIRECTION COMPONENTS
OF THE NORMAL TO TRIANGLE FGJ AND THE
ANGLE BETWEEN THIS NORMAL AND SUNLIGHT.

by
Louis Sutro

E.1 Location of Points and Triangles

Figure E-1 is a plan view of the robot aimed at azimuth 43° and viewing the mound pictured in Figs. 2-3 and 5-1. Two reference lines have been drawn in the north and east directions from which measurements were made to determine the relative latitude and longitude of points F, G and J in Table 5-1. Figure E-2 is a side view employed by the artist to generate the perspective projections in Fig. 5-1. It is presented in lieu of a stereoscopic view which could communicate the spatial relations of all of these figures.

E.2 Calculation of the Direction Components of the Normal to Triangle FGJ

The equation of the plane through any three points (x_1, y_1, z_1) , (x_2, y_2, z_2) and (x_3, y_3, z_3) is, in determinant form:

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0 \quad (\text{E-1})$$

Substitution of the data of Table 5-1 yields the equation of the plane through points F, J, and K:

$$\begin{vmatrix} x & y & z & 1 \\ 0.14 & 3.45 & -0.09 & 1 \\ -0.46 & 3.74 & -0.25 & 1 \\ -0.14 & 4.64 & 0.28 & 1 \end{vmatrix} = 0 \quad (\text{E-2})$$

This equation is of the form

$$Ax + By + Cz + D = 0 \quad (\text{E-3})$$

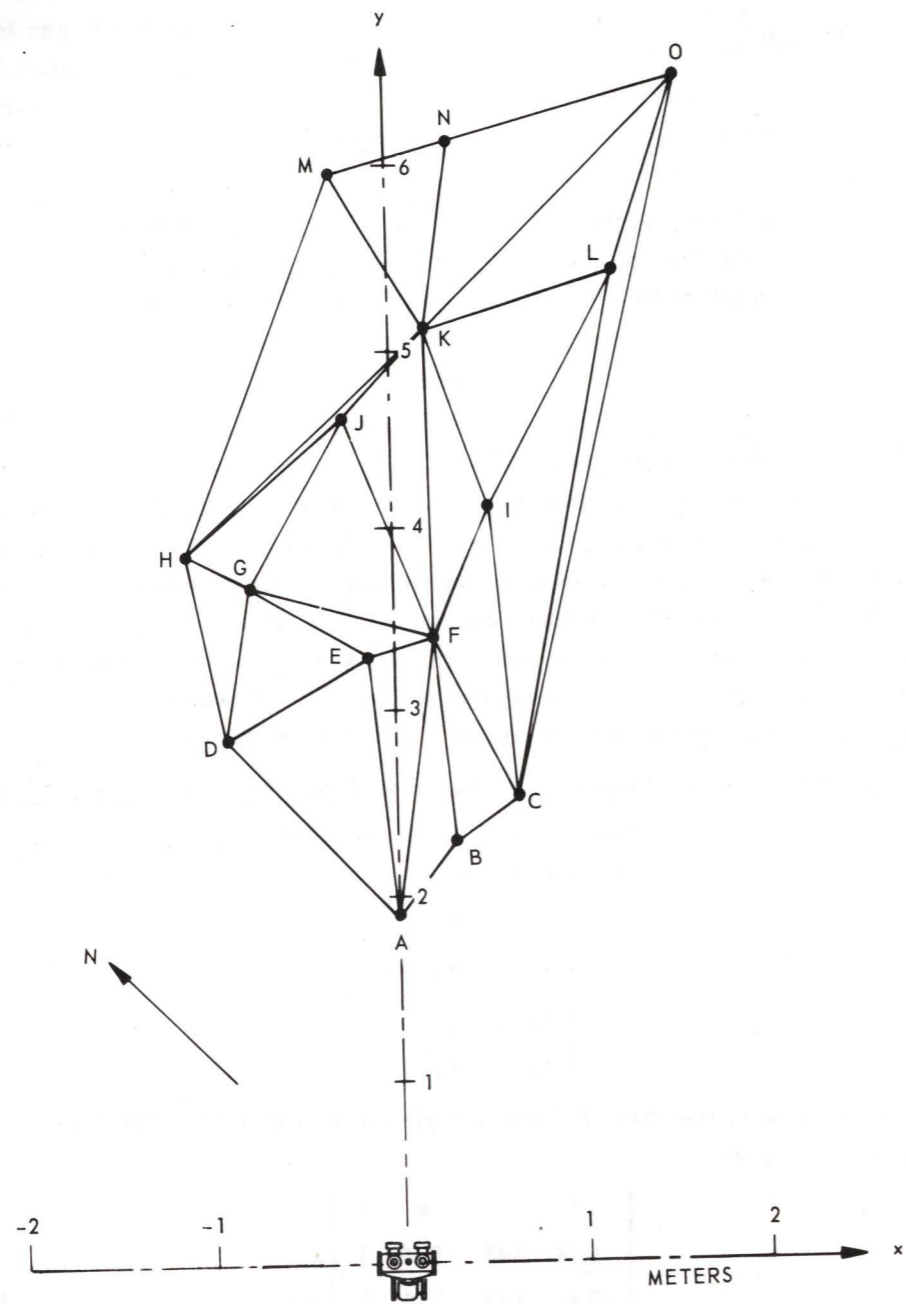


Fig. E-1. Plan view of robot and the mound shown near the center of Fig. 2-3.

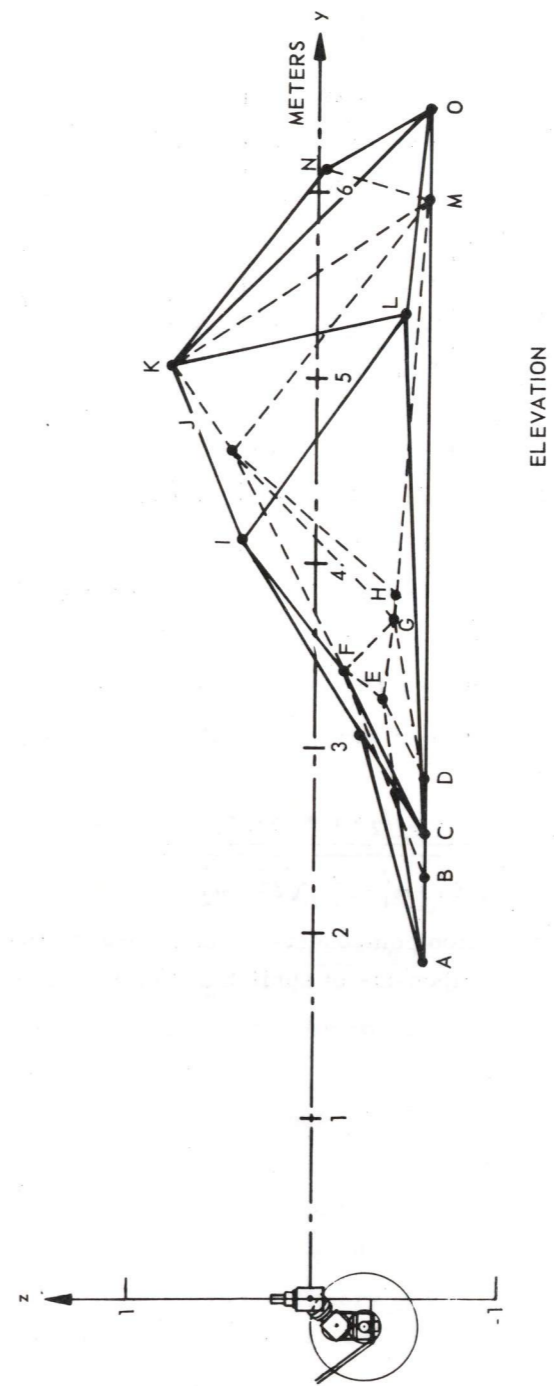


Fig. E-2. Side view corresponding to the plan view of Fig. E-1.

where

$$A = \begin{vmatrix} 3.45 & -0.09 & 1 \\ 3.74 & -0.25 & 1 \\ 4.64 & 0.28 & 1 \end{vmatrix} = -0.2977 \quad (E-4)$$

$$B = - \begin{vmatrix} 0.14 & -0.09 & 1 \\ -0.46 & -0.25 & 1 \\ -0.14 & 0.28 & 1 \end{vmatrix} = 0.2668 \quad (E-5)$$

$$C = \begin{vmatrix} 0.14 & 3.45 & 1 \\ -0.46 & 3.74 & 1 \\ -0.14 & 4.64 & 1 \end{vmatrix} = -0.6328 \quad (E-6)$$

The components of a vector perpendicular to the above plane are proportional to A, B, and C.

E. 3 Calculation of the Angle Between This Normal and Sunlight

The angle between the direction of sunlight and the normal to the plane is computed from the equation

$$\cos \theta = \pm \frac{l_1 l_2 + m_1 m_2 + n_1 n_2}{\sqrt{l_1^2 + m_1^2 + n_1^2} \sqrt{l_2^2 + m_2^2 + n_2^2}} \quad (E-7)$$

where l_1, m_1, n_1 are the direction components of the normal to the surface and l_2, m_2, n_2 are the direction components of sunlight. $\cos \theta = 0.71$.

REFERENCES

1. Sutro, L. L., Information Processing and Data Compression for Exobiology Missions, R-545, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge 39, Massachusetts, June, 1966.
2. Sutro, L. L., et al., Sensory, Decision and Control Systems, April to July 1966, R-548, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge 39, Massachusetts, November, 1966.
3. Sutro, L. L. (June 1966), op. cit. pp. 21-28.
4. Kilmer, W., Blum, J. and Peterson, D., The Decision Subsystem, Section 5 of Ref. 2, pp. 19-29.
5. Gregory, R. L., Eye and Brain, World University Library, McGraw Hill Book Co., New York, 1966, pp. 50-59.
6. Tompkins, D. N., Lunar Surface Panoramic Photography, Philco-Ford Corp., Newport Beach, California, Presented to Convention of American Society of Photogrammetry, Dayton, Ohio, September 23, 1965.
7. Tompkins, D. N., The Mars Facsimile Camera, Philco-Ford Corp., Newport Beach, California, Presented to American Institute of Aeronautics and Astronautics and American Astronautical Society, Baltimore, Maryland, March 28, 1966.
8. Sutro, L. L., et al. (Nov., 1966), op. cit., pp. 14-18.
9. Catchpole, R. J., Detection of Motion and Luminance Variation Using Vidicon Camera, T-470, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge 39, Massachusetts, September, 1966.
10. Beers, Y., Introduction to the Theory of Error, Addison-Wesley Publishing Co., Reading, Massachusetts, 1957.
11. Graham, D. N., Image Transmission by Two-Dimensional Contour Coding, Proc. IEEE, March, 1967, p. 336.

12. Huang, T. S. and Tretiak, O. J., Research in Picture Processing, in Optical and Electro-Optical Processing, The M.I.T. Press, Cambridge, Massachusetts, 1965.
13. Vander Lugt, A., Potz, F. B. and Klooster, A., Jr., Character Reading by Optical Spatial Filtering, The M.I.T. Press, Cambridge, Massachusetts, 1965.
14. Holmes, W. S., Babcock, T. R., Richmond, G. E., Pownall, L. A. and Vorie, G. C., Optical-Electronic Spatial Filtering for Pattern Recognition. The M.I.T. Press, Cambridge, Massachusetts, 1965.
15. Kulikowski, J. J., Adaptive Visual Signal Preprocessor with a Finite Number of States, IEEE Trans. on Systems Science and Cybernetics, December, 1966, p. 96.
16. Sutro, L. L., editor, Advanced Sensor and Control Systems Studies, R-519, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, January 1966, pp. 16-36.
17. Moreno-Diaz, R., An Analytical Model of the Group 2 Ganglion Cell in the Frog's Retina, E-1858, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1965.
18. IES Lighting Handbook, Fourth Edition, published by the Illuminating Engineering Society, 345 East 47th Street, New York, 1966.
19. Moon, P., The Scientific Basis of Illumination Engineering. McGraw Hill Book Co., New York, 1936, and Dover Publications, New York, 1961.
20. Sutro, L. L., A Model of Visual Space, Biological Prototypes and Synthetic Systems, Volume 1, Plenum Press, New York, 1962.
21. Moreno-Diaz, R., Conceptual Model of the Frog's Tectum, Section 6 of Ref. 2, pp. 30-40.
22. Moreno-Diaz, R., Contrast Detectors, Quarterly Progress Report No. 85, Research Laboratory for Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, April 15, 1967, pp. 321-324.
23. Barlow, H. B., Three Points about Lateral Inhibition, Sensory Communications, E. Rosenblith, editor, M.I.T. Press, Cambridge, Massachusetts, 1961, pp. 782-786.
24. Thorpe, W. H., Learning and Instinct in Animals, Methuen and Co., Ltd., London, 1963, p. 55.

25. Neurosciences Research Program Bulletin, vol. 4, pp. 105-233, Simple Systems for the Study of Learning Mechanisms, T. H. Bullock, chairman, 280 Newton St., Brookline, Massachusetts, November 30, 1966.
26. Neurosciences Research Program Bulletin, vol. 4, pp. 235-347, Brain Mechanisms in Conditioning and Learning, R. B. Livingston, chairman, 280 Newton St., Brookline, Massachusetts, Dec. 30, 1966.
27. Neurosciences Research Program Bulletin, vol. 4, pp. 235-347, op. cit., p. 245.
28. Neurosciences Research Program Bulletin, vol. 4, pp. 105-233, op. cit., pp. 177-178.
29. Thorpe, W. H., op. cit., p. 61.
30. Doyle, R. J., The Electrostatic Vidicon and Methods of Evaluation, 1962 IRE International Convention Record, American Institute of Electrical and Electronic Engineers, New York.
31. Light Measurement and Control, General Electric Large Lamp Department, Nela Park, Cleveland, Ohio, p. 20.

R-545

INFORMATION PROCESSING AND
DATA COMPRESSION FOR
EXOBIOLGY MISSIONS

by

Louis L. Sutro

Revised June 1966

INSTRUMENTATION
LABORATORY ●

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Cambridge 39, Mass.

R-545

**INFORMATION PROCESSING AND
DATA COMPRESSION FOR
EXOBIOLGY MISSIONS**

by

Louis L. Sutro

Revised June 1966

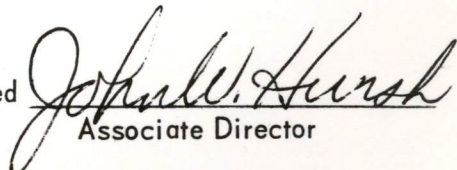
A report of work performed from
October 1965 through March 1966
under NASA Contract Number
NSR-22-009-138.

Presented at the 12th Annual Meeting
of the American Astronautical Society,
Disneyland Hotel Convention Center,
Anaheim, California, May 23-25, 1966

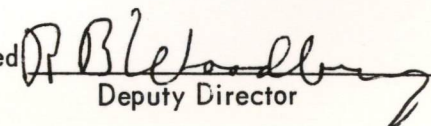
**INSTRUMENTATION LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS**

*Prepared for publication by Jackson & Moreland
Division of United Engineers and Constructors, Inc.*

Approved


Associate Director

Approved


Deputy Director

ACKNOWLEDGEMENT

This report was prepared under the auspices of DSR Project 55-257, sponsored by the Biosciences Division of the National Aeronautics and Space Administration through Contract Number NSR-22-009-138.

Publication of this report does not constitute approval by NASA of the findings or conclusions contained herein. It is published for the exchange and stimulation of ideas.

Biological help was provided by Jerome Y. Lettvin and Warren S. McCulloch, engineering help by William Kilmer, Roberto Moreno-Diaz, Jerome Krasner and Richard Catchpole.

ABSTRACT

Techniques being developed for missions to Mars are part of a larger effort to develop a robot capable of carrying out programs and of self-programming within certain limits. Three development efforts at the Instrumentation Laboratory lead in this direction: Design of a layered processor like that in a frog's retina shows how a model of animal vision may reduce data for transmission to Earth and permit response to the environment. Design of a stereoscopic system suggests how a pair of layered processors, when fed data from two television cameras, may find the positions of objects. Design of a decision and control system like that of animals suggests how a system of this kind may decide in what direction a Mars lander should move, when it should perform each experiment and what it should report.

TABLE OF CONTENTS

<u>Section</u>		<u>Page</u>
1	INTRODUCTION	1
2	THE LAYERED PROCESSING OF A VISUAL SYSTEM	2
	2.1 The Structure of a Visual System	2
	2.2 Properties of a Visual System	2
	2.3 Design of the Model (Ref. 5)	5
	2.4 Computation in Levels 1 and 2(a) *	9
	2.5 Computation in Level 2(a)	10
	2.6 Computation in Level 3	12
	2.7 Instrumentation with Available Components	13
	2.8 Instrumentation with Integrated Circuits	15
	2.9 Application to Missions to Mars	15
3	STEREOSCOPIC SYSTEMS	16
	3.1 Why Stereoscopic?	16
	3.2 Camera-Counter Chain C	17
	3.3 Camera-Computer Chain D	18
	3.4 Application to Missions to Mars	19
4	DECISION AND CONTROL PROCESSORS	21
	4.1 General	21
	4.2 Reticular Formation	21
	4.3 Model of RF*	25
	4.4 Visual-Center	28
	4.5 Application of RF* to Missions to Mars	29
5	THE NEED TO CARRY KNOWLEDGE ALONG	30
	LIST OF REFERENCES	31

INFORMATION PROCESSING AND DATA COMPRESSION
FOR EXOBIولوجY MISSIONS

1. INTRODUCTION

In designing interplanetary probes such as that we plan to put on Mars we are faced with the problem of exploring after the vehicle has landed. We are attempting to devise a self-moving probe capable of exercising judgments in a certain sense and handling a variety of contingencies. It would broadcast the results of its experiments rather than be a simple extension of our senses and movements into the alien world.

It is clear that technology supplies us increasingly reliable parts ever smaller and able to work on little power so that computers become more and more compact. At the same time the art of using these devices expands greatly so that they have ceased to be the novelties for study and have become, instead, common tools. A number of workers, concerned with this art (including us), are attempting to make a true robot, capable of self-programming within certain limits, one that is capable of handling a large sequence of choices, correcting itself when it makes an error in judgment. That is not to say that we are so rash as to claim that we can build an animal or an animal intelligence, but it would be intelligence of a sort, and one that would permit communication with us.

Each experiment proposed for Mars requires both a sensor system, designed for the problem, and a data processing system designed to bridge the gap between the sensor system and the way our minds work. We propose to subject the data of the experiments to a decision system similar to ours. What comes out of this robot must then comply with our criteria of interest and intelligibility.

Can this be done in time for a Mars landing in the 1970's? The only way to find out is to try.

This paper describes three development programs under way at the Instrumentation Laboratory. One is the design of the layered processing of a visual system. The second is the design of a stereoscopic system. The third is a decision and control system.

How would a robot reduce data? We propose that the robot carry with it measures of the usefulness to man of the data it acquires, so that it will send back only data which satisfies those measures. For each of the development programs to be described such measures will be suggested. But a more useful set of measures will have to come from you.

2. THE LAYERED PROCESSING OF A VISUAL SYSTEM

2.1 The Structure of a Visual System

Dr. Warren McCulloch refers to an eye as a "layered computer". By this he means that all the bipolar cells process the data they receive from the photoreceptor layer, passing it along to the ganglion cell layer. The latter processes the output of the bipolar-cell layer passing along the results of its computation to the layered structures of the brain.

We want to model the human visual system, but it comprises far more cerebral layers than we can model now. For that reason we sought an animal with fewer layers in its visual system but equal or greater complexity per layer. We found this in the visual system of the frog.

2.2 Properties of a Frog's Visual System

The species studied, the common leopard frog or *Rana pipiens*, responds to one color and to four pattern aspects of objects: edges, dark convex moving edge (bug), any time varying visual event, and dimming.

The retina of the frog has three layers of nerve cells: photoreceptors, bipolar cells, and ganglion cells, as shown schematically in Fig. 1.

Each cell type is uniformly distributed over the retina.

The photoreceptor layer consists of several types of rods and cones.

An individual photoreceptor apparently may have two output parameters.

The first appears to be a logarithmic function of illumination, related to

the bleaching rate of a pigment. A second parameter appears to be a function of the history of illumination, related to the amount of bleached pigment.

Bipolar cells are so called because their two ends are somewhat alike. Horizontal cells make interconnections among bipolar cells in the layer of photoreceptor outputs. There are interconnections formed by amacrine cells in the layers of connections between bipolars and ganglion cells.

Ganglion cells appear to have five distinguishable dendritic arbors, that is, tree-like branchings. Three of these arbors are shown as O, P, and Q in Fig. 2. On the basis of the connections of bipolar cells to these arbors, Lettvin et al. were able to assign one of five functions to each arbor form;^(1, 2, 3) four concerning the visual pattern and one concerning color. The four pattern outputs go to the tectum (Fig. 1 right) while the color output goes to the geniculate body (not shown). The first four cell groups produce outputs representing:

1. Edge detection
2. Bug detection
3. Event detection (any stimulus change)
4. Dimming detection.

Of the approximately 5×10^5 ganglion cells (Fig. 1) approximately half are bug detectors. Since this has proved to be the most difficult ganglion cell to model,^(4,5) we have concentrated on it.

Figure 3 portrays vertical sections of two bipolar cells and one multi-level E-shaped ganglion cell, as drawn by Ramon y Cajal for the frog.⁽⁶⁾ In the latter there are three dendritic levels. Lettvin et al identified this structure as a bug detector. Note that arborization exists between the two outer dendritic levels.

Two types of bipolar cells will be represented in the model. Their designation "on" and "off", is that of Schwyperheyn.⁽⁷⁾ The distinction between "on" and "off" cells was previously applied by Hartline to ganglion cells.⁽⁸⁾

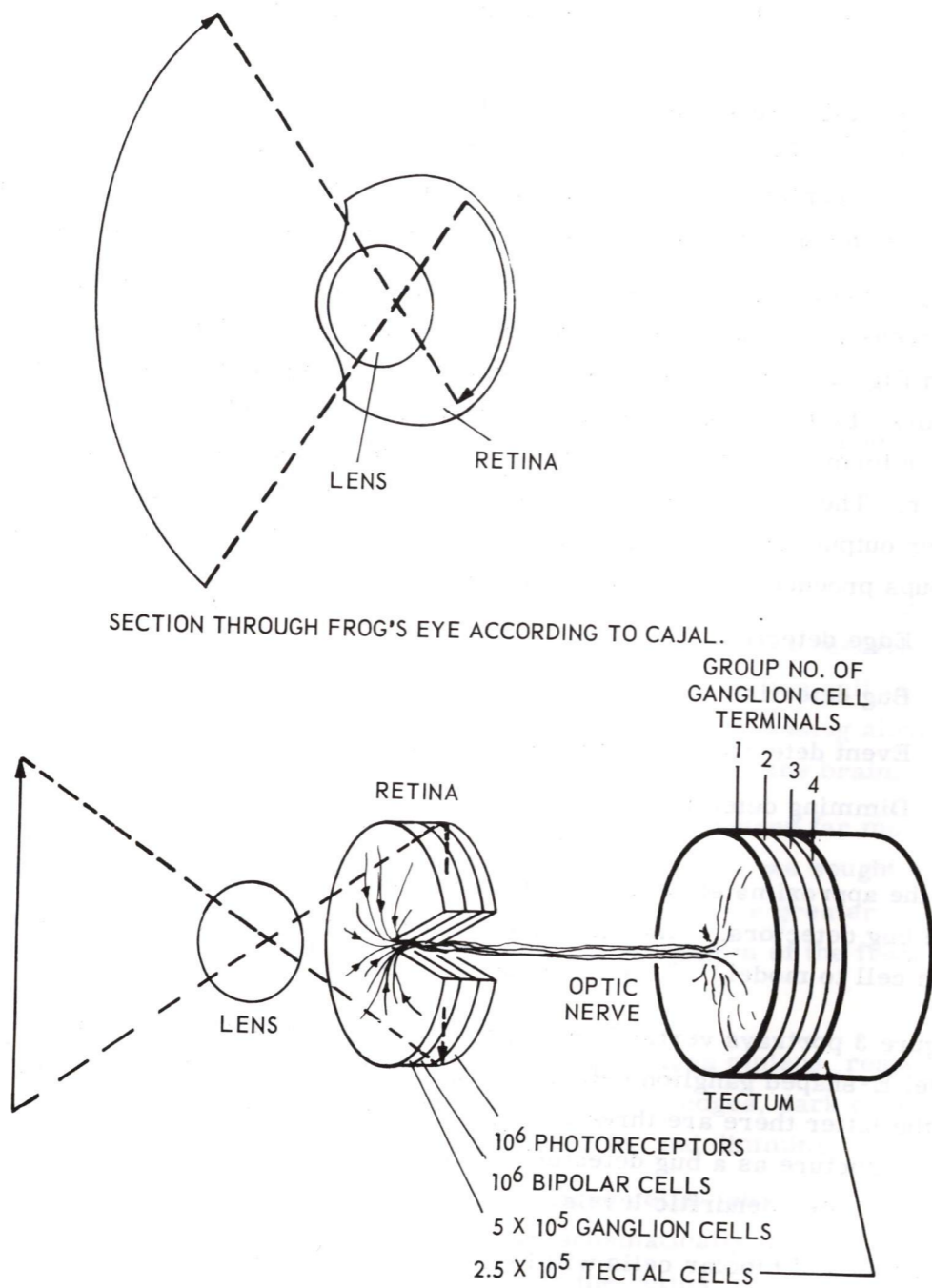


Fig. 1. Schematic of frog visual system.

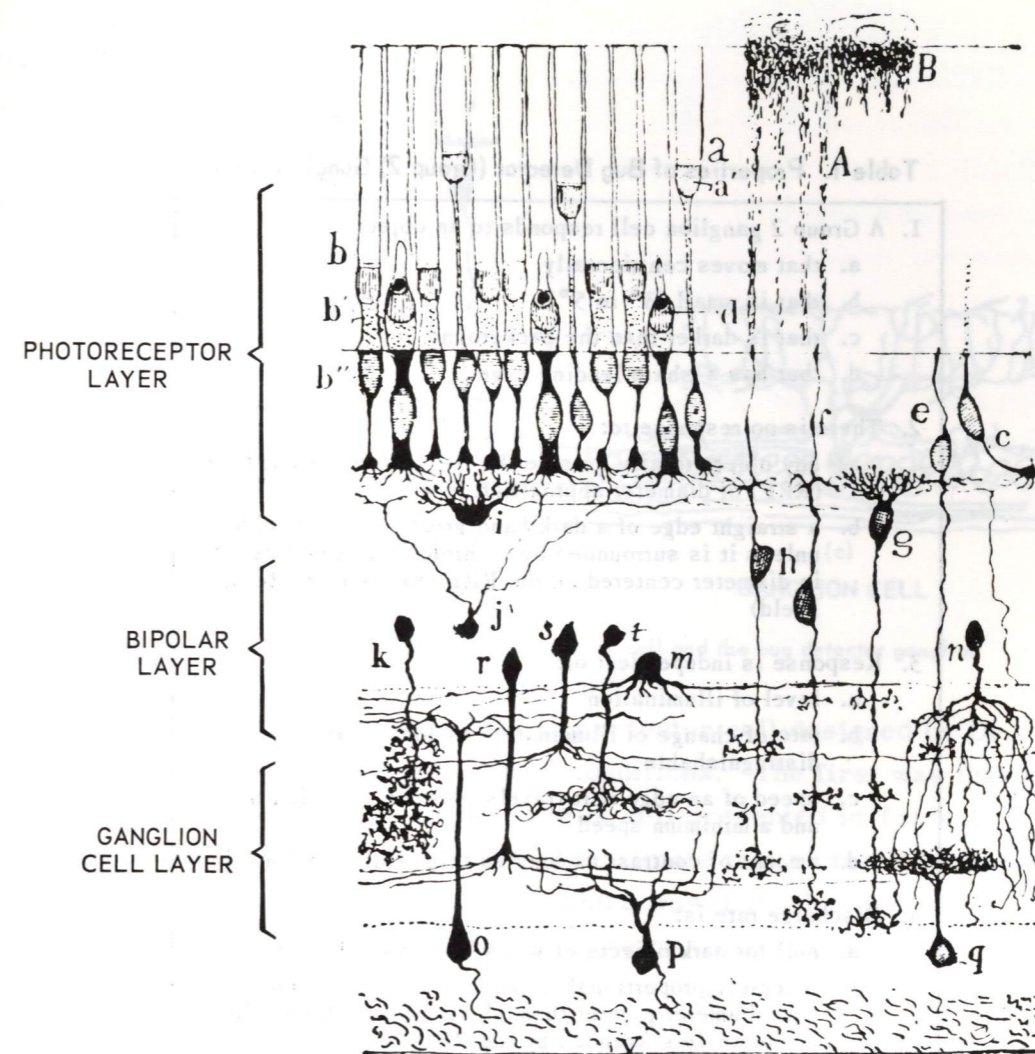


Fig. 2. The retina of frog in Colgi stain - highly schematic but showing spatial arrangement: g and h are bipolar cells; i is a horizontal cell; k, r, s, t, and m are amacrine cells; o, p, and q are ganglion cells. This is the off-hand diagram used for illustration in Ramon y Cajal's *Histologic du Systeme Nerveux* (Paris: Maloine, 1909-1911).

2.3 Design of the Model (Ref. 5)

Essentially, the bug detector cell responds to a small dark object which moves centripetally into its responsive field. In the list of the cell's properties given in Table 1, the smallness of the object is item 1b, darkness item 1c, centripetal movement item 1a. Note that the discharge rate of the cell (items 4a to 4g) depends upon the size and location of the object, both measured in degrees or minutes of arc at the lens of the frog's eye. Five minutes of arc are imaged on one photoreceptor.

Table 1. Properties of Bug Detector (Group 2) Ganglion Cell.

1. A Group 2 ganglion cell responds to an object :
 - a. that moves centripetally
 - b. that is small (3° to 5°)
 - c. that is darker than the background
 - d. that has a sharp leading edge
2. There is no response to:
 - a. any object totally outside the responsive retinal field (RRF) of diameter approximately 4°
 - b. a straight edge of a dark band greater than 2° wide unless it is surrounded by a shield approximately 4° in diameter centered on the RRF (Responsive Retinal Field)
3. Response is independent of:
 - a. level of illumination
 - b. rate of change of illumination as long as this is distinguishable
 - c. speed of an edge if it travels between a maximum and a minimum speed
 - d. amount of contrast as long as this is distinguishable
4. Discharge rate is:
 - a. null for dark objects of width less than about $3'$
 - b. inversely proportional to the convexity of an image with diameter greater than $3'$ but less than one half of the angle subtended by the RRF
 - c. maximal (approx. 40 pulses per second) for an image subtending half the angle subtended by the RRF
 - d. proportional to the convexity of images larger than one half of the RRF angle
 - e. feeble if the object is larger than RRF
 - f. greater to movement broken into several steps than to continuous movement
 - g. feeble in response to light spots
5. The response is maintained for several seconds if the object stops in the RRF

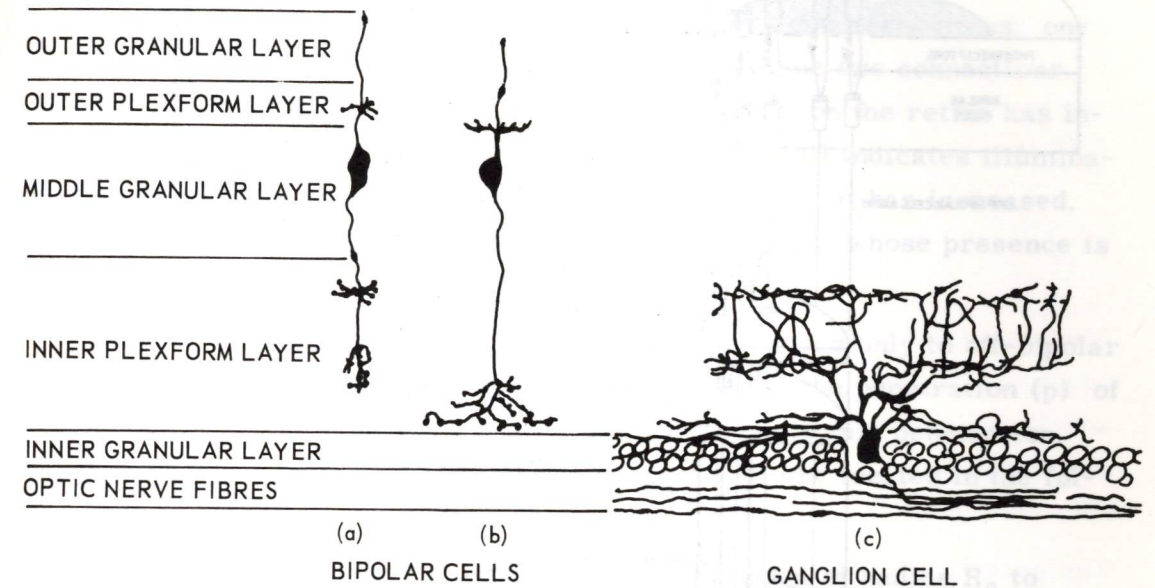


Fig. 3. Retinal sections of two types of bipolar cell and the bug detector ganglion cell in a frog retina (after Ramon y Cajal).

Figure 4 diagrams a model of the bug detector cell designed by Dr. Roberto Moreno-Diaz to meet three conditions. The first was that it fit the experimental data on this type of cell presented in Table 1. The second was that its structure approximate that pictured by Cajal and identified by Lettvin et al. The third was that all of the 1/4 million cells of this type be capable of being built with existing hardware.

Dr. Warren McCulloch feels that in a model of perception, feedback plays an important role. Since this model employs no feedback it cannot be considered an attempt to solve the general problem of perception. It is rather a scheme to match experimental results of physiology with known anatomy, within the restrictions mentioned above.

At the top of Fig. 4, the image of a bug (shaded) is moving toward the center of concentric areas, of radii R_1 and R_2 . The area of radius R_1 (called responsive retinal field or RRF in Table 1) includes approximately 2000 photoreceptor cells, three of which are represented by small cylinders. The larger area represents approximately 20,000 cells; five of which are represented by the three same cylinders and two additional ones.

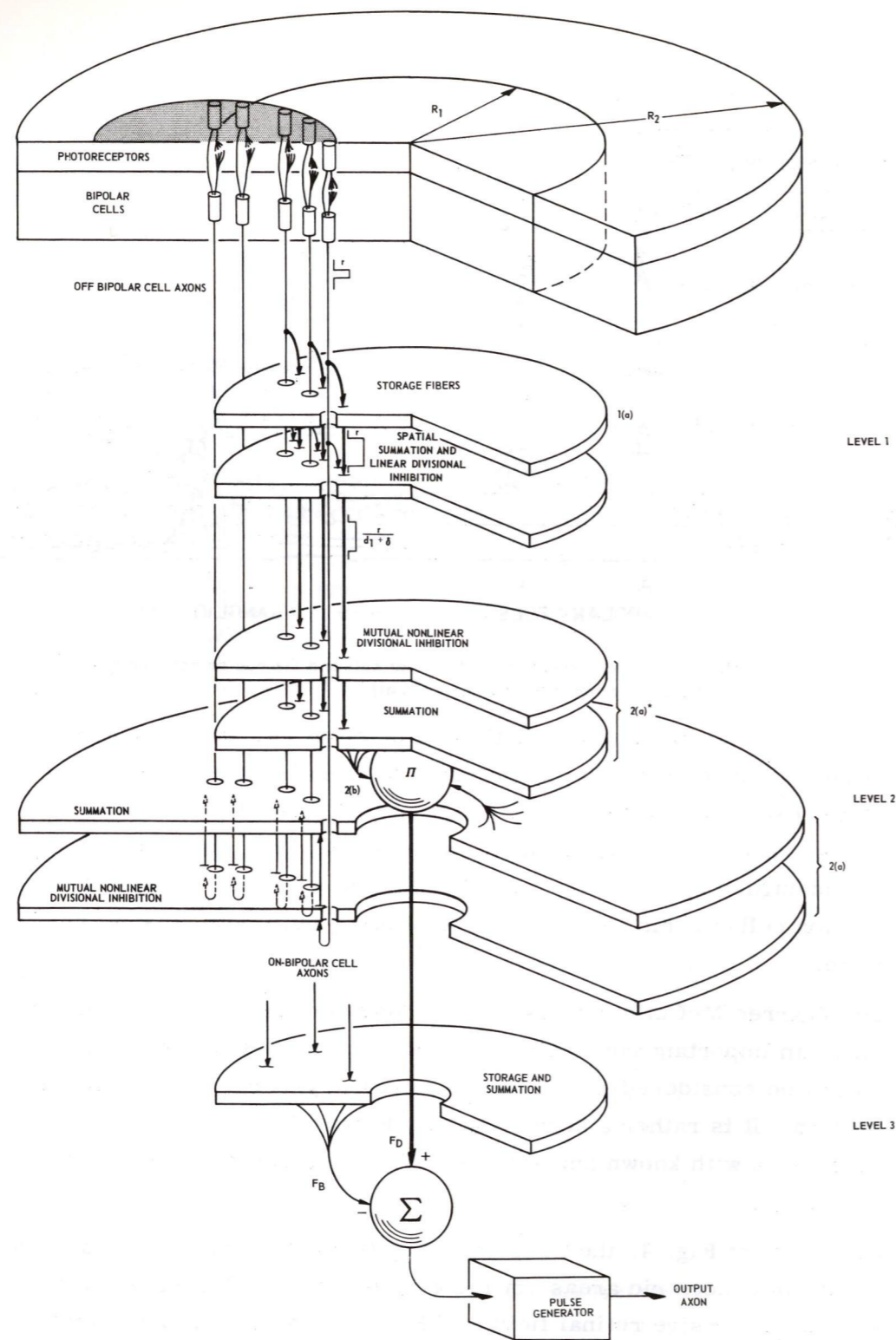


Fig. 4. Diagram of models of photoreceptors, bipolar cells and bug detector ganglion cell.

Between each photoreceptor and bipolar cell are two connections, one fast and one slow. By comparing inputs from these two connections, each bipolar cell determines whether illumination on the retina has increased or decreased. Firing of an off-bipolar cell indicates illumination has decreased, firing of an on-bipolar cell that it has increased. Only off-bipolar cells are shown. On-bipolar cells, whose presence is assumed, connect to level 3.

Levels 1 and 2(a)* of the ganglion cell model respond only to off-bipolar cells in the area of radius R_1 (RRF) to determine the penetration (p) of the bug into this area and the nonlinear response to this penetration. Computations performed in this layer and a half are treated in the following subsection.

Level 2(a) responds to off-bipolar cells in the area of radius R_2 to determine the convexity of the leading edge of the bug. This computation is treated in the second subsection following.

Level 3 responds only to on-bipolar cells in the area of radius R_1 (RRF) to determine, with previous computations, that the motion is centripetal. This computation is described in the third subsection following.

Activity of a bipolar cell is represented by a pulse of amplitude r , as shown in Fig. 4. This pulse is applied through a broad curving branch to level 1(a), through a fine curving branch to level 1(b) and through the lowest slab of level 2(a), to which it returns.

2.4 Computation in Levels 1 and 2(a) *

The pulse of amplitude r that passes through the broad curving branch of the off-bipolar cell axon is stretched in level 1(a), as shown in Fig. 4. While such neural storage usually has an exponential decay, it is here given a square waveform for simplicity. In level 1(b) the amplitude r is divided by the number of photoreceptors in the leading edge of the image, d_1 , plus a constant δ . The second branch — the fine one — makes possible the spatial summation that forms this sum d_1 .

In the first slab of level 2, the advancing pulses interact to inhibit each other in a mutual nonlinear divisional fashion suggested by Schuyperheyn.⁽⁷⁾ The stretched pulses, representing the length of time since

the image moved into the circle of radius R_1 , together indicate the area darkened, D_1 . By the formula for this kind of inhibition,⁽⁵⁾ the amplitude of the signal on the j^{th} wide line in the upper slab of layer 2(a)* is

$$A_j^* = c_1^* \frac{r}{d_1 + \delta} e^{-k_1^* \frac{rD_1}{d_1 + \delta}}$$

The asterisks or stars are applied to the symbols of this layer to distinguish them from the symbols of layer 2(a). The quotient $D_1/(d_1 + \delta)$, which can be approximated D_1/d_1 , is an area divided by the length of the leading edge. It equals approximately the penetration p of this area into the circle of radius R_1 (RRF).

The second slab of level 2(a)* forms a sum A_c^* of all the active A_j^* signals. There are D_1 of these.

$$A_c^* = c_1^* \frac{rD_1}{d_1 + \delta} e^{-k_1^* \frac{rD_1}{d_1 + \delta}}$$

$$\approx c_1^* r p e^{-k_1^* r p}$$

Employing the experimental evidence of Table 1, $k_1^* r$ can be evaluated, leading to

$$A_c^* = c_1^* r p e^{-\frac{p}{0.26R_1}}$$

which is plotted in Fig. 5. The constant c_1^* can be evaluated by a method indicated in Ref. 5.

2.5 Computation in Level 2(a)

The sum of the outputs of the off-bipolar cells, in the area of radius R_2 , in the last time period, is the "length of dimming" d_2 . The signal passed from the lower to the upper slab of layer 2(a) on the j^{th} line is

$$A_j = c_2 r e^{-k_2 d_2 r}$$

where c_2 and k_2 are constants. The interaction represented by this formula is again the "mutual nonlinear divisional inhibition" of J. J. Schypperheyn.

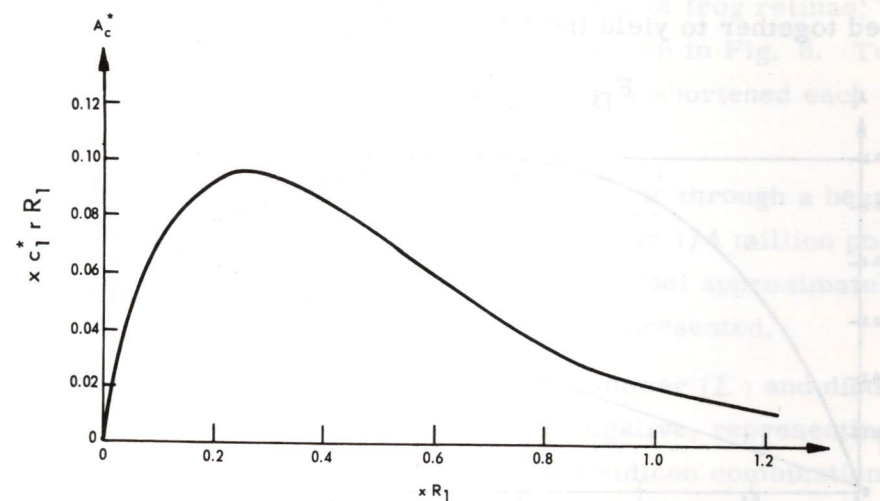


Fig. 5. A_c^* as a function of penetration p .

The upper of the two slabs of layer 2(a) sums the signals on the lines rising from the lower slab and compares the sum to a threshold θ . The sum is represented by A_{net} and can be computed by summing the A_j ,

$$A_c = \sum A_j,$$

and subtracting the threshold

$$A_{\text{net}} = A_c - \theta$$

Figure 6 presents a graph of the sum of the signals, A_c , rising from the lower slab. Figure 7 presents the difference A_{net} . The response is thus made dependent on the size of the object. Experiments on the frog indicate that the smallest object detected is about 3' (Table 1, 4b) of arc wide, that there is an optimum response to objects whose image area is about one half the area of radius R_1 and response falls to zero for objects whose image size is approximately this area, i. e., $d_1 \approx \pi R_1$.

At level 2(b), represented by a sphere, A_{net} , A^* and a constant are multiplied together to yield the function of dimming, F_D :

$$F_D = K_D A_{net} \cdot A^*$$

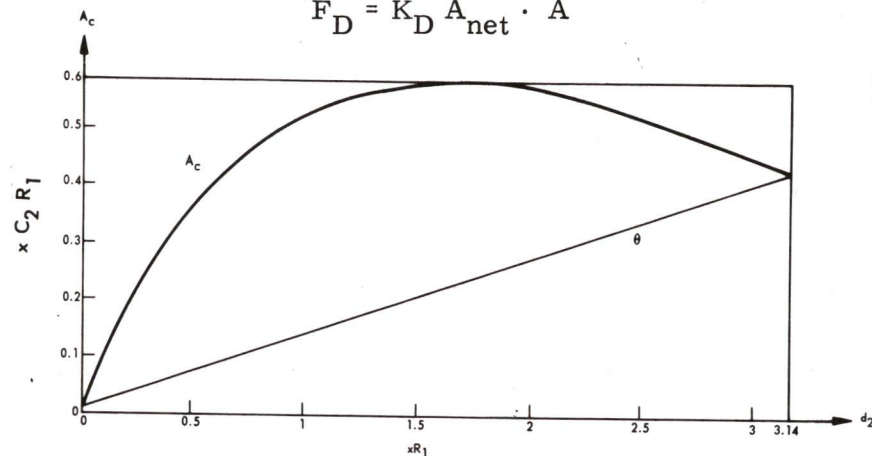


Fig. 6. A_c and θ as functions of the length of dimming d_2 .

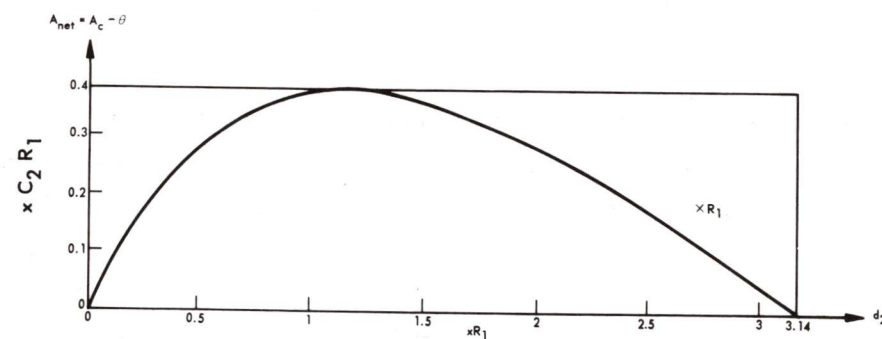


Fig. 7. A_{net} as a function of the length of dimming d_2 .

2.6 Computation in Level 3

The outputs of on-bipolar cells report brightening as the image of the bug advances. Level 3 stretches the pulses from on-bipolar cells, and sums them to give the function of brightening F_B . The summer at the bottom of the diagram takes F_D in a positive sense and F_B in a negative sense to get their difference. The pulse generator emits a frequency proportional to this difference. Thus, only as the dark image of a bug moves toward the center of the field of radius R_1 is there an output. By choice of constants, the pulse rate can be made maximum for an image that subtends half the angle subtended by the area of radius R_1 . This is the condition of maximum response of the living cell.

2.7 Instrumentation with Available Components

Using elements of previous designs of models of frog retinas,⁽⁴⁾ Dr. Moreno-Diaz devised the hardware model shown in Fig. 8. To permit placing the entire design on one page, we have shortened each vidicon into a wide flat bottle.

The two vidicons at the top view the same scene through a beam splitter. The raster of each tube comprises 500×500 or 1/4 million positions. If this resolution is maintained through the model approximately 1/4 million bug detector ganglion cells will be represented.

The bipolar cells are represented by the summer (Σ) and diodes that follow it. If the output of the summer is negative, representing dimming, the output goes to the cathode ray tube and vidicon combination immediately below it. If the output is positive, representing brightening, it goes to the CRT-vidicon in the lower part of the illustration.

The model of the bug detector ganglion cell is divided into three levels as in Fig. 4. In level 1 of Fig. 8, dimming is mapped on the face of the cathode ray tube. Fibre optics communicate the slowly fading image to a vidicon. The combination of tubes, commercially available as a scan conversion tube, performs temporal summation.

The output of the vidicon goes to a pulse generator (PG) which delivers its output to a bank of shift registers which map 50 lines of the vidicon raster. Attached to the shift registers and representative of the area of radius R_1 in Fig. 4 is a summer (Σ) which determines the part of the area of radius R_1 that has been dimmed, D_1 . The bank of shift registers work this way: As the first position of the first line of the raster is scanned, a 1 or 0 is fed to the first shift register. As successive positions are scanned, successive 1's or 0's are fed to the first shift register. At the start of the second line of the vidicon raster, the 1's and 0's in the first shift register are fed to the left end of the second shift register. After 50 lines have been scanned 1's and 0's representing the dimming in these lines will be loaded in the bank of 50 shift registers. From then on the registers overflow but always contain 50 lines.

The length of dimming d_1 is computed in the second bank of shift registers and combined with D_1 in the box labelled \div . The quotient D_1/d_1 is the

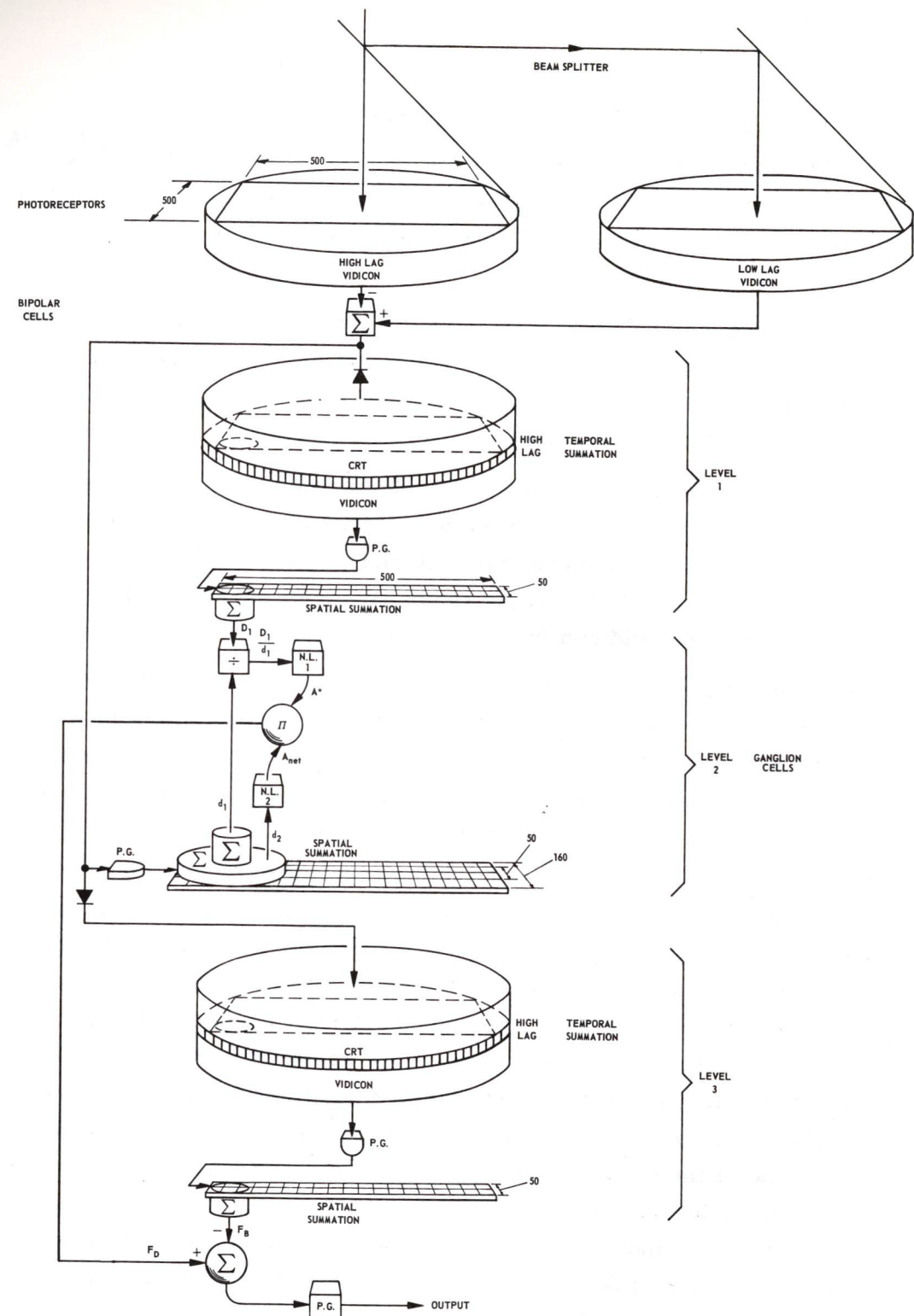


Fig. 8. Possible instrumentation of 1/4 million Moreno-Diaz models of the bug detector ganglion cell.

penetration p of the image of the bug into the area of radius R_1 . The box labelled "N. L. 1" performs the nonlinear transformation of Fig. 5.

The second shift register and its large-area summer (Σ) computes the length of dimming d_2 . This is fed to Nonlinear Circuit 2 which performs according to Fig. 7. The analog multiplier labelled π combines A_{net} , A^* and a constant to form F_D .

In level 3 a second combination of cathode ray tube and vidicon determine the areas brightened in the whole retina. A third bank of shift registers and summer determine the brightened area B_1 for each cell. F_B is proportional to B_1 . The final summer subtracts F_B from F_D . The final pulse generator yields a pulse frequency proportional to this difference.

2.8 Instrumentation with Integrated Circuits

We are now redesigning the system of Fig. 8 to employ integrated circuit elements instead of vidicons and cathode ray tubes. Figure 9 shows an array of 50×50 photo transistors developed by Westinghouse.⁽¹⁰⁾ That company has since built four times as many units in the same area.

We are on the way, we think, of devising a way of using this kind of array. From our point of view, it has two difficulties. One is the difference in response characteristics between phototransistors. The other is defects in phototransistors, i. e., some may not function at all. In a system we are designing the output of each phototransistor will be compared to its own previous output. Thus discrepancies between phototransistors should not enter the computation.

2.9 Application to a Mission to Mars

Design of a layered processor to recognize a bug of a particular size and travelling in a particular direction indicates that other recognition systems can be designed both to recognize predetermined objects and objects newly experienced. Such recognition is a reduction in data, because it permits reporting simply the presence of the objects. The system about to be described should also permit reporting the location of an object.

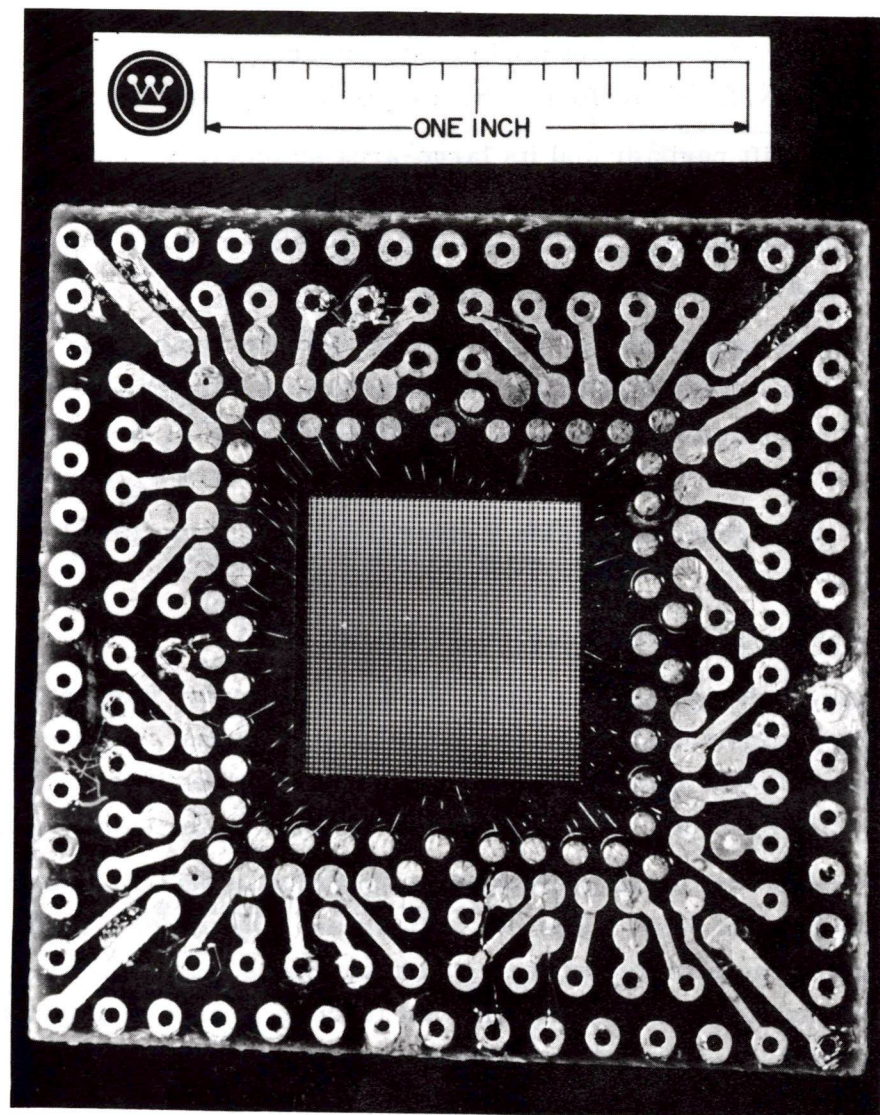


Fig. 9. Array of 50×50 phototransistors developed by Westinghouse.

3. STEREOSCOPIC SYSTEMS

3.1 Why Stereoscopic?

One of the tasks a robot may have to perform is to discover separate objects so as to avoid them, if moving, or describe them in detail if they have properties of interest. Several strategies are possible in the use of a pair of television cameras coupled to a computer. One that we are considering is diagrammed in Fig. 10.

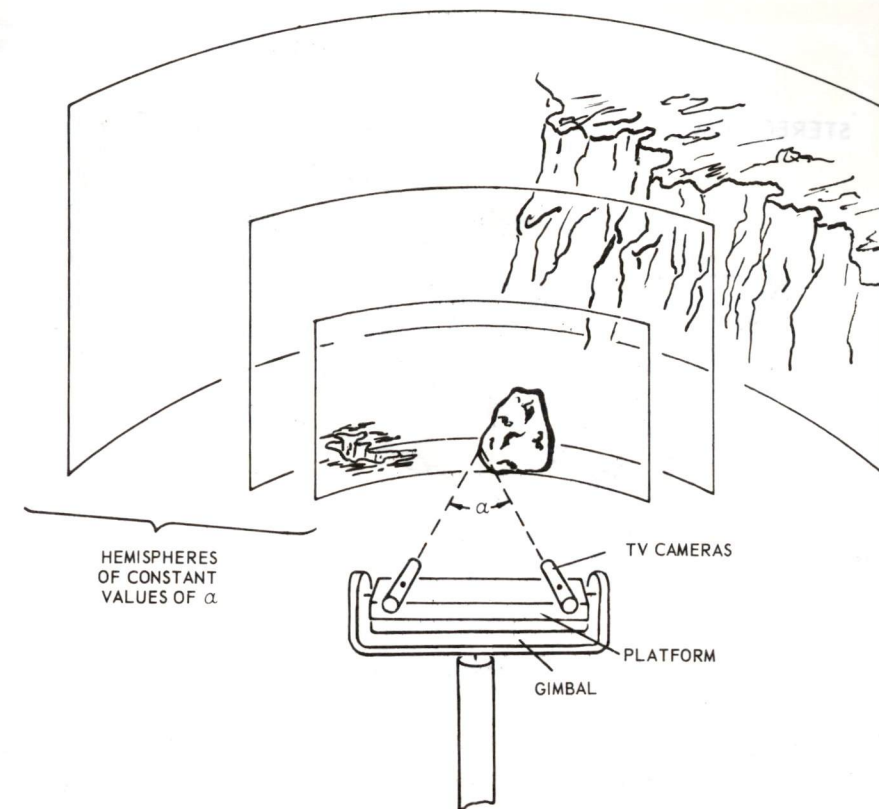


Fig. 10. A strategy to discriminate objects by a stereoscopic pair of TV cameras and a computer.

Here the cameras have mechanically scanned the innermost of three hemispheres until they have encountered the edge of a rock. The hemisphere scanned is that for which the angle of convergence of the two cameras is the value of the angle α shown. For a smaller value of α , the hole could be encountered. For a still smaller value, the edge of the cliff could be encountered. The problem is how to bring about the encounter.

3.2 Camera-Counter Chain C

To carry out initial experiments, we have attached a beam splitter to the front of a television camera, as shown in the illustration of our Model C camera computer chain (Fig. 11). The geometry of the two views is shown in Fig. 12. The prism of the previous illustration is here omitted for simplicity and the single camera shown as two. The left camera looks out through the lower cone, the right through the

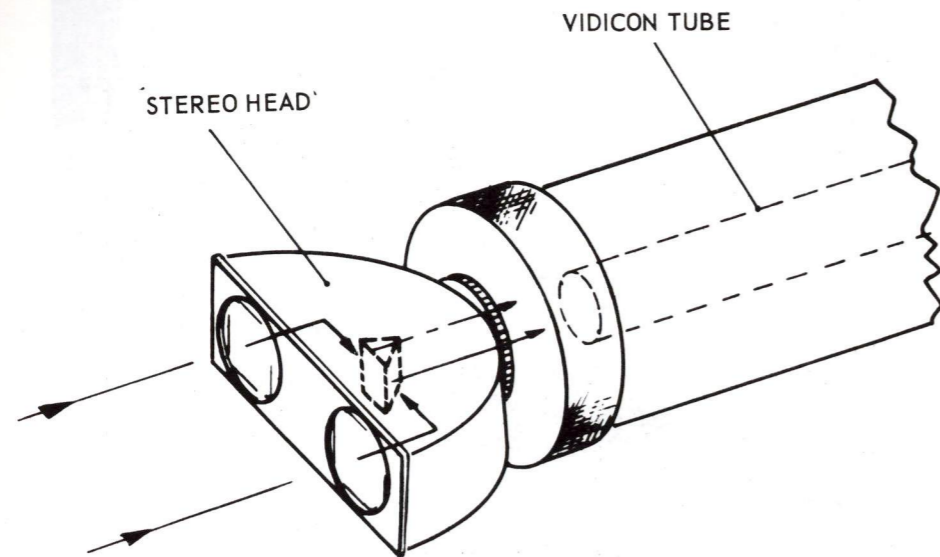


Fig. 11. Front end of Model C camera computer chain.

upper cone, the two cones overlapping some distance in front of the cameras. The point P on the edge of the right cone is near the center of the lower cone. It is imaged at P'_R on the right half of the camera, at P'_L on the left half of the camera.

Left and right TV frames have been drawn in front of the cameras to illustrate the method of computation. To search for an object such as that with the angle of convergence α shown in Fig. 11, windows will be advanced from left to right across the frames. Computation then can be limited, at any instant, to interpretation of the data in the two windows.

3.3 Camera-Computer Chain D

A constraint on camera-computer chain C is that it employ a lens of average focal length. Thus each camera sees neither a full field of view nor the narrow field of view that provides high resolution. To obviate this difficulty, camera computer chain D is being designed. (See Fig. 13.) It has interchangeable lenses and each camera is mounted to swivel. In addition, the cameras are mounted on a platform that nods.

Figure 14 shows a possible configuration of cameras and computers to visually explore a scene and report, according to the strategy described

above, first the presence of free-standing objects then details about one or a few of these objects.

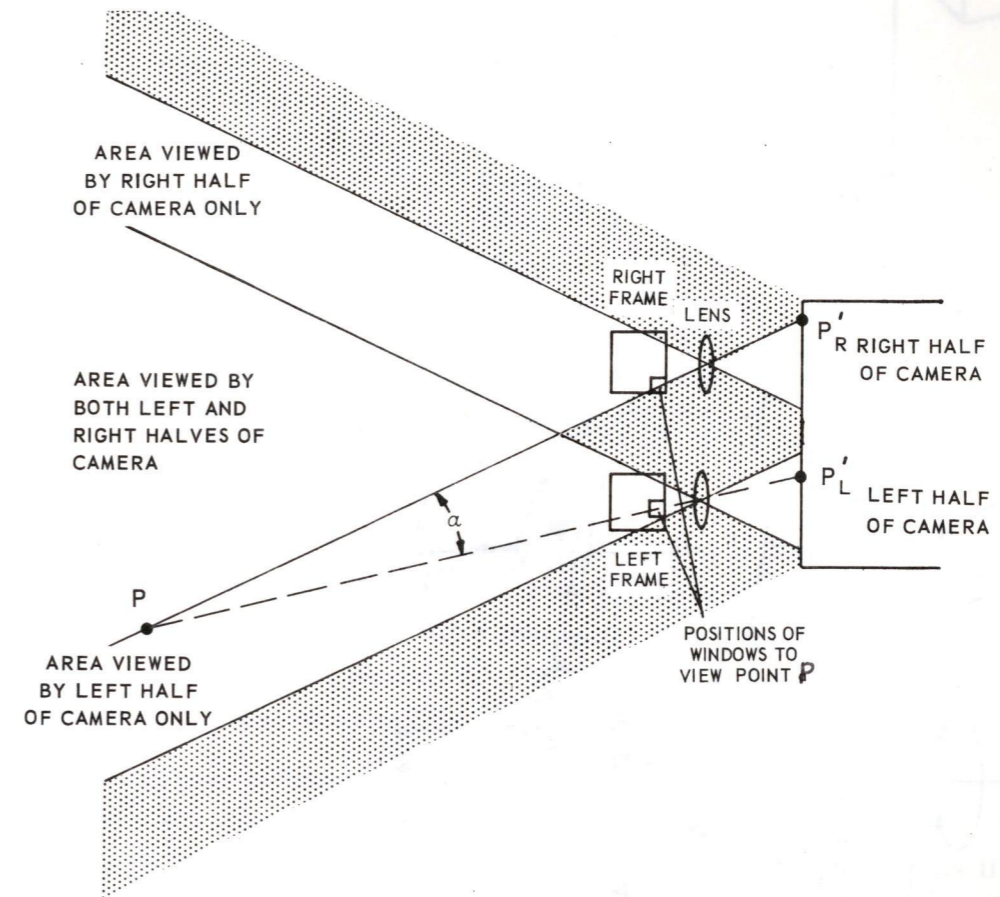


Fig. 12. Schematic of front end of Model C camera computer chain.

3.4 Application to Missions to Mars

If several free-standing objects are detected, how is one to be selected for more detailed examination? An animal could use another sense, such as the sense of touch. We are planning senses other than that of vision. In the following section we refer to these as sense modalities. The problem then arises: How should a robot combine the inputs from different sense modalities to decide what to do?

PLEASE INSERT THIS CORRECTED PAGE IN YOUR COPY OF
M.I.T. INSTRUMENTATION LAB. REPORT E-1858, "AN ANALYTICAL
MODEL OF THE GROUP 2 GANGLION CELL IN THE FROG'S RETINA",
BY ROBERTO MORENO-DIAZ, OCTOBER, 1965

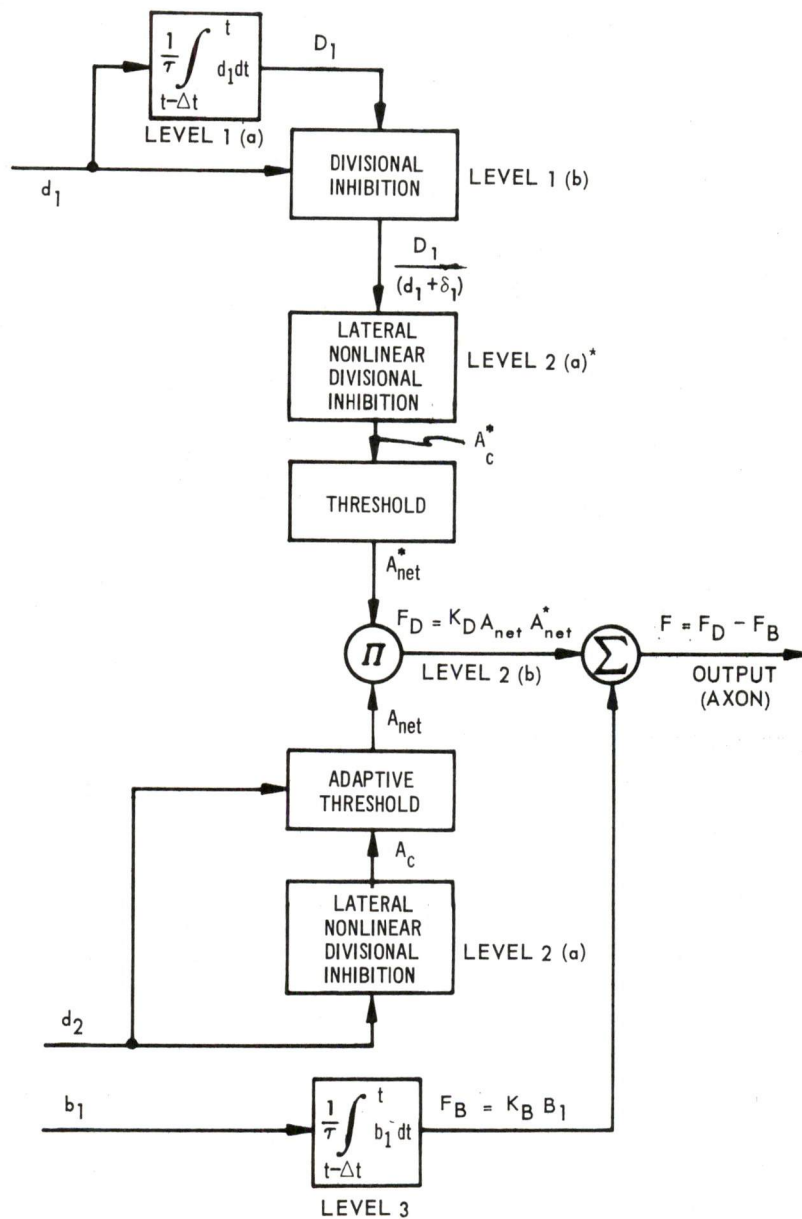


Fig. 16 Equivalent block diagram of Group 2 ganglion cell.

The pulse repetition frequency of the output along the ganglion cell axon is proportional to F (Hypothesis 3i).

4.4 Evaluation of the Constants and Thresholds

4.4.1 Summary

The documented results of physiological experiments on the frog retina are used here to determine the values of the various constants and thresholds of the model. First are described the spatio-temporal properties of an object and its motion that correspond to the numbers of active fibres. Then successively are shown how k_2 depends on the size of the dark disc that maximizes the response; how an adaptation is necessary for the threshold of A_{net} ; how R_2 may be evaluated from the width of the widest straight band that can excite the cell; how k_1^* is restricted by the result that the cell responds only to centripetally moving images and how the relative influence of Level 3 predicts the preferential response to dark rather than bright images.

4.4.2 Geometric Significance of n_{d1} , N_{D1} , n_{b1} , and N_{B1}

The bipolar cells are randomly distributed (Hypothesis 2g) in such a manner that, given a line of constant length in the retina, the number of cells intersected is independent of the direction and position of the line. Consider the sharp edge of an object in contrast with its background, that is moving through the retinal field. The number n_{d1} of off-bipolar cells activated at any instant is proportional to the length of image dimming within the circle of radius $R_1^{(*)}$.

$$n_{d1} \propto d_1 \quad (4.23)$$

where d_1 is the length of the edge causing dimming – or "length of dimming" – within the RRF.

Similarly, n_{d2} is proportional to the "length of dimming" within the circle of radius R_2 . Thus, we write

$$k_2 n_{d2} r = K_2 d_2 \quad [\text{See note ** below}] \quad (4.24)$$

(*) n_{d1} is also a function of image contrast and velocity, but is independent of them within limits determined by photoreceptor and bipolar cell parameters. Image contrast and velocity are assumed constant in this report.

(**) In subsection 4.3 variables are the number of bipolar cells that fire under different conditions; lower case letters express the constants: c_1^* , k_1^* , c_2 and k_2 . In subsection 4.4 variables are the length of the edge and the area scanned; upper case letters express the constants: C_1^* , K_1^* , C_2 , and K_2 .

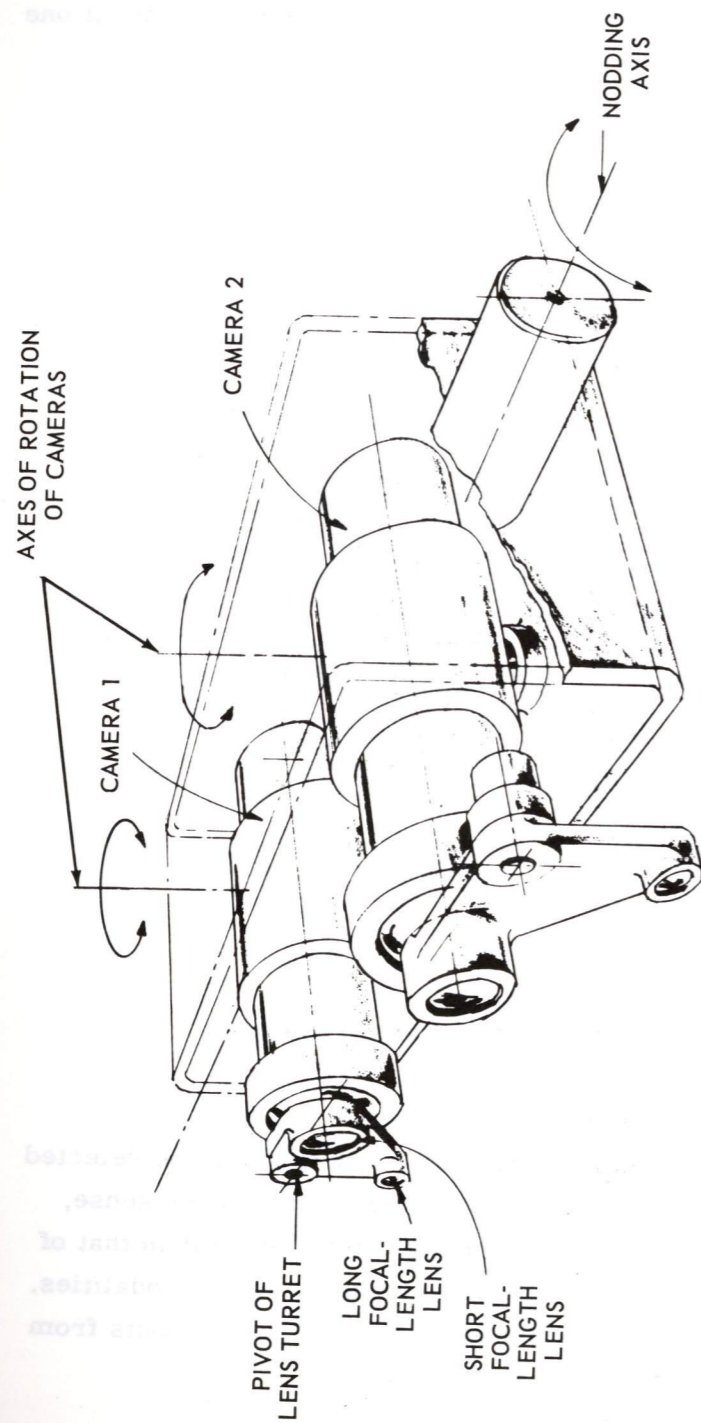


Fig. 13. Front end of Model D camera computer chain.

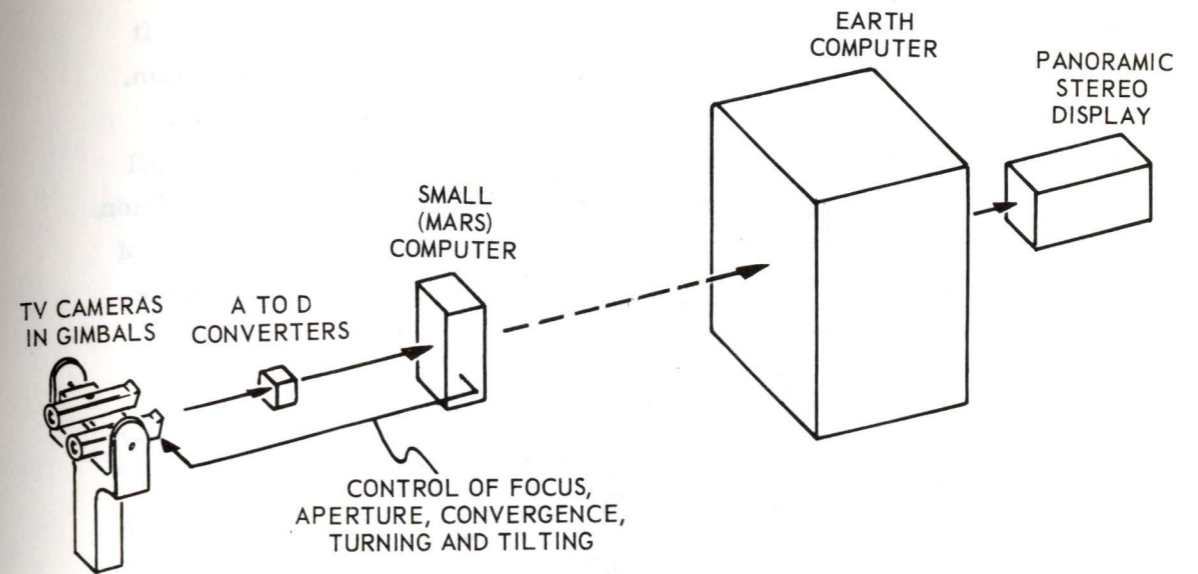


Fig. 14. Block diagram of Mars-Earth camera computer chain.

4. DECISION AND CONTROL PROCESSORS

4.1 General

As we enter into the field of layered processing of information we are in strange territory. There is no mathematical theory of nonlinear integral transforms. There is no theory for handling purely informational (as opposed to power) feedback. Group theory is concerned with better defined cases than what we are handling.

But, if layered computation seems the method used by all brains, it may be one of the necessary conditions for doing what an animal does. There is no way by which sequential operations map parallel operations if the elements in a layer are also mutually connected, as seems the case in most of brain. Thus we have had no recourse but to start simulating and trying before we adduce general design principles.

Two ventures in computer design beside the design of eyes will now be presented. First is the model of reticular formation designed by Drs. McCulloch and Kilmer.⁽¹¹⁾ Second is a processor for the output of our "eye" (described under "Visual Center" below).

4.2 Reticular Formation

The reticular formation (RF) integrates the sensory-motor and autonomic-nervous relations so as to permit an organism to function as a unit instead

of a mere collection of organs. The core of all nervous systems, it extends throughout the length of the spinal cord and into the cranium. (See Fig. 15.) It receives inputs from all of the separate sensory systems, auditory, visual, olfactory, vestibular, etc., and from all of the housekeeping systems involving visceral regulation, respiration, cardiovascular control, etc. Figure 16 shows the wide flat spread of the reticular cells to which the inputs connect and between which the reticular cells interconnect.

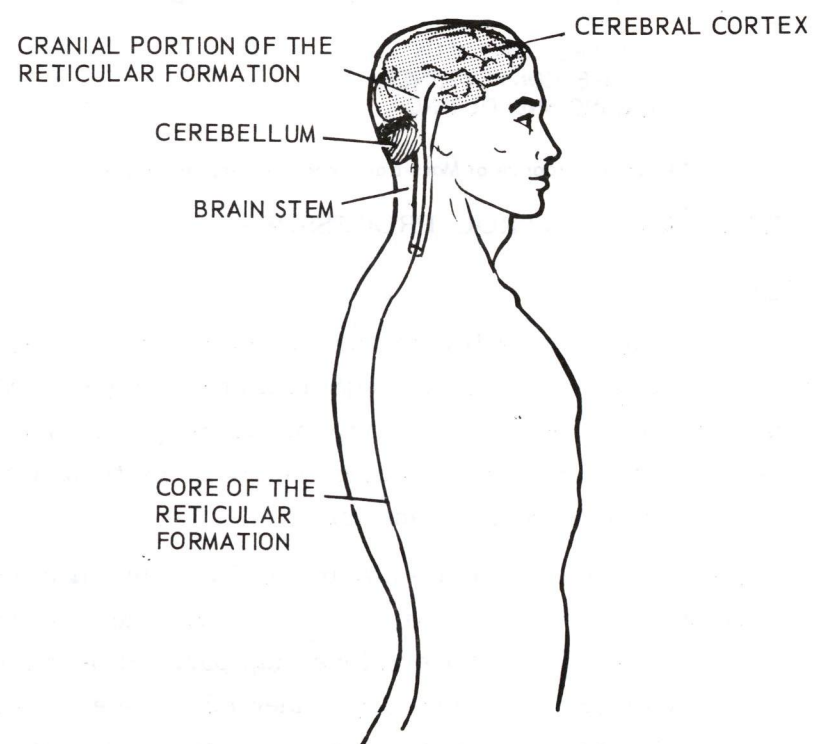


Fig. 15. Core of the reticular formation in the cranium, brain stem and spinal cord of man.

As Kilmer et al⁽¹²⁾ say, "The Scheibels⁽¹³⁾ have so far done the most definitive neuroanatomy that we have on the reticular formation. In their milestone report of 1962 they caricatured its anatomical structure in the brain stem by comparing it to a stack of poker chips. In each chip region the dendritic processes of RF neurons ramify in the plane of the chip face, often covering nearly half of the face area. This causes a very large degree of overlap and intermingling among dendrites of

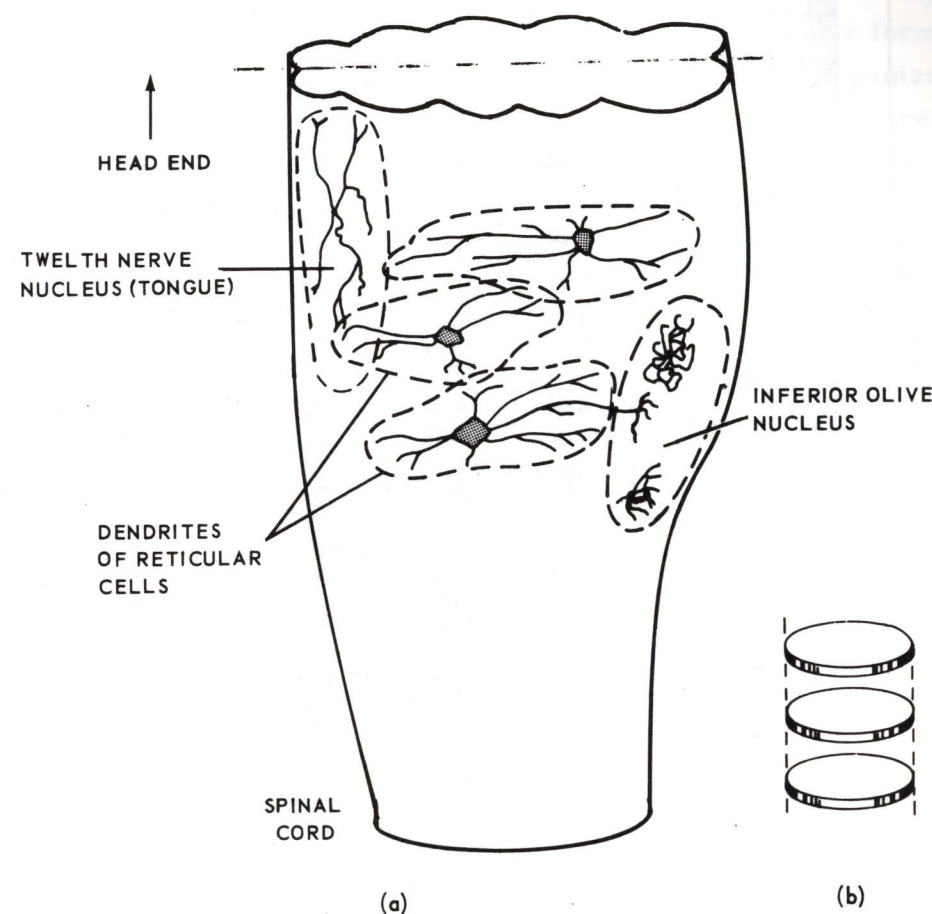
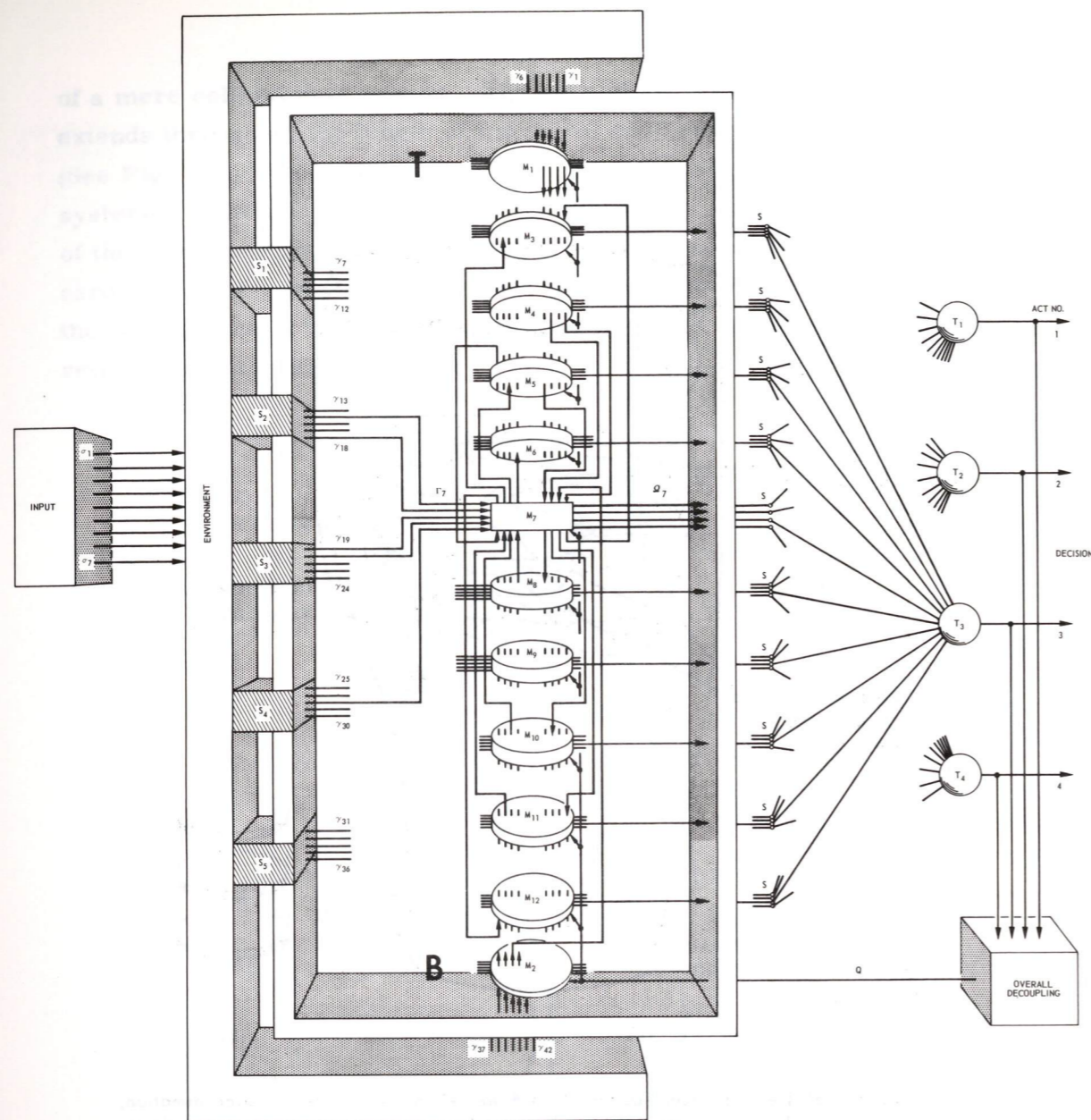


Fig. 16. (a) Dendritic ramifications from three cell bodies in the reticular formation, from two in the inferior olive and from one in the twelfth nerve nucleus (kitten brain, Scheibel, 1962)

(b) Poker chip caricature of reticular cell dendritic ramifications in (a)

nearby neurons, as shown schematically for just the brain stem region in Fig. 16. (This is very similar to Scheibel's Figure 1 in reference 13.) . . . Since as many as half dozen or more input systems may synapse on a single RF neuron, and each RF input nucleus and fibre tract in general feeds very many RF cross-sectional levels, the Scheibels suggest that the RF might tolerate considerable puddling of information at each of its cross-sectional levels, but demand somewhat



Legend (from left to right)

σ_i	factor of the environment	M_i	module i
S_i	sense	Ω_i	four components of the probability vector from module i
γ_i	input	s	step function
Γ_i	inputs to module i	T_i	threshold element

Fig. 17. Schematic of the computer-simulated RF*.

greater informational rigor between levels. Figure 16(b) is a caricature of the dendritic fields, comparing them to a stack of poker chips.

"In its non-specific control of sensory inputs, the reticular formation is analogous to an admiral-of-the-fleet, committing the organization under its command to one act, trusting that fine perceptions are made at their centers of specialty and are accurate. As a computer, the reticular formation is seen to be broad in its domain of command but exceedingly shallow in any specific area under its command. It commits the organism to an act which is a function of the information that has been played upon it in the last fraction of a second or so. After commanding the organism to act, e.g., fight, run, swallow, vomit, sleep, copulate, etc., it sends out control directives to all of the specialized centers of the central nervous system, tuning their activities to this task.

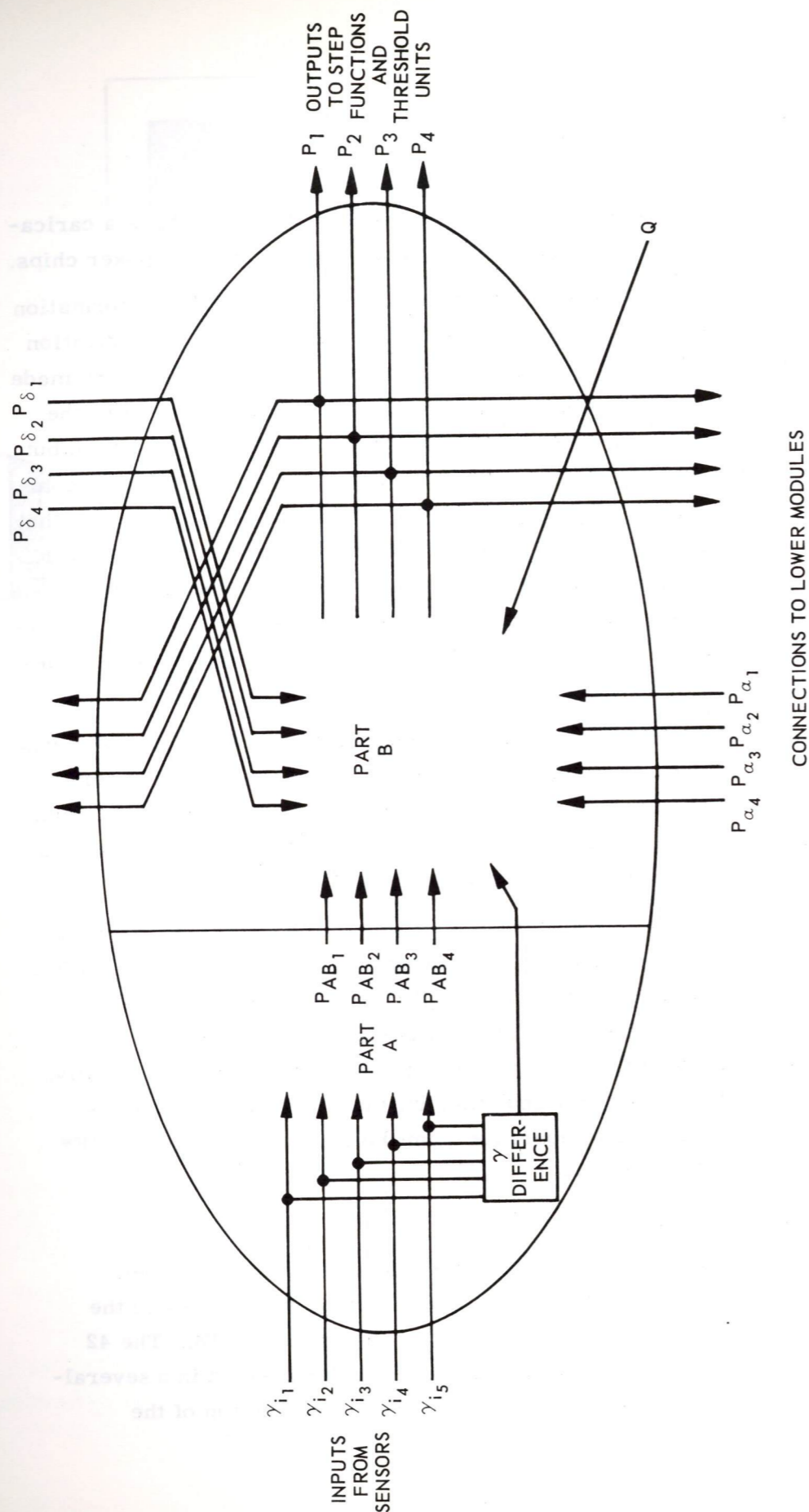
"Its important features are that it must handle a vast amount of highly correlated input information and arrive at one of a small number of mutually exclusive acts in a dozen time steps or so and with minimal equipment. The crucial information as to what act the organism has selected is distributed over its input lines.

"In this paper we are concerned only with that part of the RF which makes the decision to commit the organism, henceforth denoted RF*. RF* is RF minus everything on the RF's input side (the dorsal-lateral RF) and output side (the ventral-lateral RF, basal ganglia, etc.), all of the reflexes along the neuraxis that are handled locally, and all of the respiratory and other rhythmic operational aspects which are functionally separable from the decisionary task." (RF* is called \overline{RF} in Ref. 11.)

4.3 Model of RF*

"Figure 17 shows a schematic of a decision and control model analogous to the reticular formation. Each of the modules in the figure is a hybrid computer of the type shown in Fig. 18. The 42 input lines, γ_i , to the modules in Fig. 17 are connected in a several-to-all, but not all-to-all, fashion to the modules. The top of the

CONNECTIONS TO HIGHER MODULES



CONNECTIONS TO LOWER MODULES

Fig. 18. Input and output connections to parts A and B of a module.

drawing (T) and the bottom (B) correspond roughly to the diencephalic and high cervical regions, respectively, of higher vertebrates. The white-boundaried half-box represents the environment of the organism being modelled. The environment is established by factors, σ_i , of the set Σ . The S_i represent exteroceptive, interoceptive, and internuncial systems which sense this environment and feed highly (but nontrivially) correlated inputs, γ_i , directly into the model. For convenience, all σ_i and γ_i lines are binary.

"There are twelve logic modules in the simulated RF*. The sets of lines Ω_i (only Ω_7 is marked) indicate the act preferred by module M_i . There are four lines in each set because the model is at present capable of four acts. The four lines carry the four components of a probability vector.

"For clarity, all types of connection lines (Γ, Ω , ascending and descending lines) are shown only for module M_7 , whereas these types of connection are actually made to all regular modules. Each M_i receives inputs from several but not all S_j , and each S_j feeds several but not all M_i .

"Each module (Fig. 18) computes directly on the input information that it receives and makes a best guess as to what the corresponding act should probably be. After the initial guess, the modules communicate their decisions to each other over low capacity information channels. Then each module takes the information (from all other modules) and combines it in a nonlinear fashion with the information coming directly into it to arrive at a mixed guess as to what act should be performed. This is in turn communicated to the subset of modules to which it is connected above and below. Thus we have decomposed the module shown in Fig. 18 into two parts. The A part operates on the module's input information. The B part operates on information from above, below and from the a part.

"The A part with five binary input variables and four analog output variable outputs, is a nonlinear probability transformation network."

Quoting Ref. 14, "The B part receives 4-component probability vectors P_δ from the above, P_α from below and P_{AB} from the A part. The j^{th} component of each probability vector is the probability, computed by the module of origin, that model RF*'s present γ input signal configuration demands act number j . The B part also receives the difference between γ inputs, called " γ difference" in Fig. 18, and the overall decoupling Q shown being formed in Fig. 17. The B part yields the module's determination of the probability that the model RF*'s present γ input signal configuration demands act number j .

"The design of the modules was straightforward. The main problems concerned the way in which the computation converged to produce consensus. This consensus is achieved, as illustrated in Fig. 17, by first determining at point s if the j^{th} component of the probability vector P from each module exceeds 0.5. If it does, a 1 is passed on to the threshold element T . There, if 60% of the input connections are 1's, act j is decided upon. Note that each element T in Fig. 17 receives 10 inputs. For clarity of the drawing connections are shown only to T_3 . The threshold elements T are crude models of the muscle systems of animal which receive many inputs, decode them and act.

"The model has been successfully simulated on the M. I. T. Instrumentation Laboratory Honeywell Computer in collaboration with Jay Blum, E. Craighill and D. Peterson. The model converged to the correct act in each of about 50 test cases, and always in from 5 to 25 time steps.

"We are now concentrating on the functional design of a considerably enriched model which can handle conditioning and extinction in a satisfactory time domain sense."

4.5 Visual-Center

Walter Pitts once pointed out the consequences of having sequences of widely connected layers such as exist in a brain. It is that there will exist "stimulus" equivalences in the output — and the classing of stimuli

into "similar" and "dissimilar" may not necessarily follow those which would be natural to us. Nor could we from analyzing the connectivities in the net determine what these classifications would be.

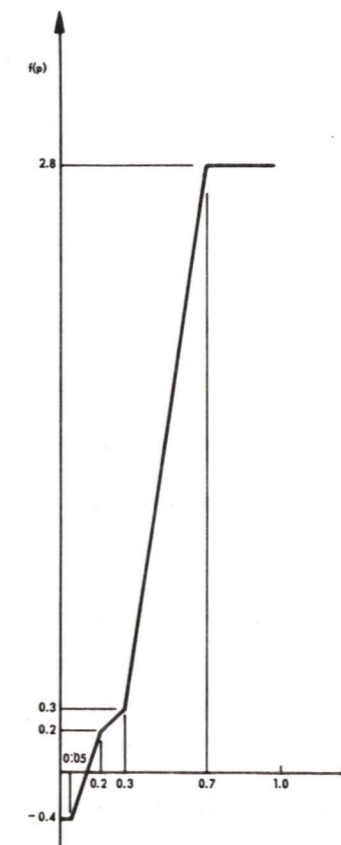


Fig. 19. The $f(p)$ function.

We intend, however, to superpose on our "eye" other layered nets so as to come, by patch and try, to a kind of excerpting of forms from the image as classes in the output. We have no idea yet of our chances for success. But the method must be tried.

4.6 Application of RF* to Missions to Mars

As Dr. Kilmer puts it, ⁽¹⁵⁾ "Our next task will be to sufficiently enrich the present model of RF* to enable it to exhibit several types of learning. This will require the insertion of delay chains between modules and the addition of adjustable decision circuitry in the a part of each module."

A task beyond this will be to tie the visual center to the model of RF* through far more channels than are shown in Fig. 17. Other sense modalities will be needed such as the sense of touch. Dr. Kilmer suggests that on Mars the model of RF* may have to decide among the following acts:

1. Rest
2. Locomote to _____
3. Turn around
4. Preprogrammed operation mode I
5. Preprogrammed operation mode II
6. Preprogrammed operation mode III
7. Maintenance
8. Communication to space vehicle (landed) I
9. Communication to space vehicle (landed) II
10. Righting (after overturn)

For example, if operation mode I were a visual experiment, it would not be switched on in darkness or a very high wind. For another example, if operation mode II were a soil assay experiment, only conditions of warmth and low wind velocities might trigger it.

5. THE NEED TO CARRY KNOWLEDGE ALONG

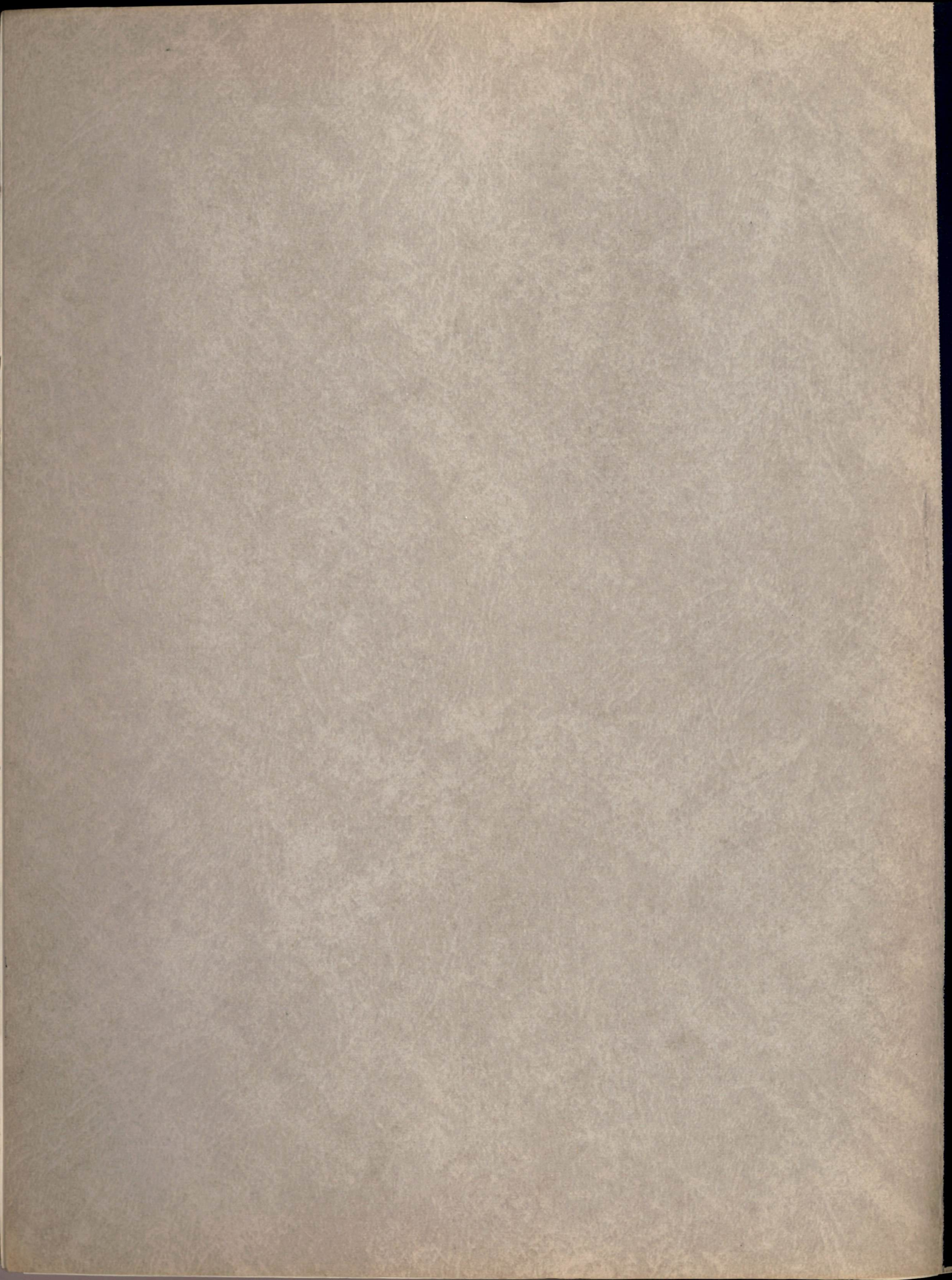
To be useful, such a device has to know something, as Lettvin, Maturana, Pitts, and McCulloch concluded when they started to study a frog's eye. What a frog knows in its retina is described above.

To look for what scientists need to learn about the surface of Mars, our device must know what the scientists are looking for. We shall be happy to respect their wishes.

LIST OF REFERENCES

1. Lettvin, J. Y. , Maturana, H. R. , McCulloch, W. S. , and Pitts, W. H. , What the Frog's Eye Tells the Frog's Brain, Proc. of the IRE, p. 1940, November 1959.
2. Maturana, H. R. , Lettvin, J. Y. , McCulloch, W. S. , and Pitts, W. H. , Anatomy and Physiology of Vision in the Frog (Rana Pipiens), Journal of General Physiology, pp. 129-175, July 1960.
3. Lettvin, J. Y. , Maturana, H. R. , Pitts, W. H. , and McCulloch, W. S. , Two Remarks on the Visual System of the Frog, Sensory Communication, W. Rosenblith, ed. , pp. 757-776, 1961.
4. Sutro, L. L. , editor, Advanced Sensor and Control System Studies, 1964 to October 1965, R-519, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, p. 16, January 1966.
5. Moreno-Diaz, R. , An Analytical Model of the Group 2 Ganglion Cell in the Frog's Retina, E-1858, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, October 1965.
6. Ramon y Cajal, S. , Die Retina der Wirbelthiere, Tafel II, Fig. 6 Wiesbaden, Verlag von J. F. Bergmann, 1894.
7. Schypperheyn, J. J. , Contrast Detection in Frog's Retina, Acta Physiol. Pharmacol. Neerlandica 13, pp. 231-277, 1965.
8. Hartline, H. K. , The Response of Single Optic Nerve Fibres of the Vertebrate Eye to Illumination of the Retina, Amer. J. Physiol. , Vol. 121, pp. 400-415, February 1938.
9. Sutro, L. L. , editor, op. cit. , pp. 10-38.
10. Westinghouse Electric Corporation Aerospace Division, Baltimore, Md. , Advanced Molecular Systems Technology, Volume 1 Systems Applications, pp. 4-11 to 4-16, November 1964.
11. Sutro, L. L. , editor, op. cit. , pp. 84-95.

12. Kilmer, W. L., McCulloch, W.S., and Blum, J., Towards a Theory of the Reticular Formation, 1966 IEEE National Convention, Cybernetics Session, March 14, 1966, New York City, New York, as modified for this paper.
13. Scheibel, M. and A., On Neural Mechanisms for Self-Knowledge and Command, Mitre Report SS-3, First Congress on the Information Sciences, Mitre Corporation, Bedford, Massachusetts, 1962.
14. Kilmer, W. L., Summary of Research Progress, Research Laboratory of Electronics Quarterly Progress Report, July 15, 1966, as modified for this paper.
15. Kilmer, W. L., Section in Advanced Sensor and Control System Studies, April to June 1966, R-548, Louis Sutro, Editor, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, (in preparation).



R-545

INFORMATION PROCESSING AND
DATA COMPRESSION FOR
EXOBIOLGY MISSIONS

by

Louis L. Sutro

Revised June 1966

INSTRUMENTATION
LABORATORY •

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Cambridge 39. Mass.

The pulse repetition frequency of the output along the ganglion cell axon is proportional to F (Hypothesis 3i).

4.4 Evaluation of the Constants and Thresholds

4.4.1 Summary

The documented results of physiological experiments on the frog retina are used here to determine the values of the various constants and thresholds of the model. First are described the spatio-temporal properties of an object and its motion that correspond to the numbers of active fibres. Then successively are shown how k_2 depends on the size of the dark disc that maximizes the response; how an adaptation is necessary for the threshold of A_{net} ; how R_2 may be evaluated from the width of the widest straight band that can excite the cell; how k_1^* is restricted by the result that the cell responds only to centripetally moving images and how the relative influence of Level 3 predicts the preferential response to dark rather than bright images.

4.4.2 Geometric Significance of n_{d1} , N_{D1} , n_{b1} , and N_{B1}

The bipolar cells are randomly distributed (Hypothesis 2g) in such a manner that, given a line of constant length in the retina, the number of cells intersected is independent of the direction and position of the line. Consider the sharp edge of an object in contrast with its background, that is moving through the retinal field. The number n_{d1} of off-bipolar cells activated at any instant is proportional to the length of image dimming within the circle of radius $R_1^{(*)}$.

$$n_{d1} \propto d_1 \quad (4.23)$$

where d_1 is the length of the edge causing dimming – or "length of dimming" – within the RRF.

Similarly, n_{d2} is proportional to the "length of dimming" within the circle of radius R_2 . Thus, we write

$$k_2 n_{d2} r = K_2 d_2 \quad [\text{See note ** below}] \quad (4.24)$$

(*) n_{d1} is also a function of image contrast and velocity, but is independent of them within limits determined by photoreceptor and bipolar cell parameters. Image contrast and velocity are assumed constant in this report.

(**) In subsection 4.3 variables are the number of bipolar cells that fire under different conditions; lower case letters express the constants: c_1^* , k_1^* , c_2 and k_2 . In subsection 4.4 variables are the length of the edge and the area scanned; upper case letters express the constants: C_1^* , K_1^* , C_2 , and K_2 .

PLEASE INSERT THIS CORRECTED PAGE IN YOUR COPY OF
M.I.T. INSTRUMENTATION LAB. REPORT E-1858, "AN ANALYTICAL
MODEL OF THE GROUP 2 GANGLION CELL IN THE FROG'S RETINA",
BY ROBERTO MORENO-DIAZ, OCTOBER, 1965

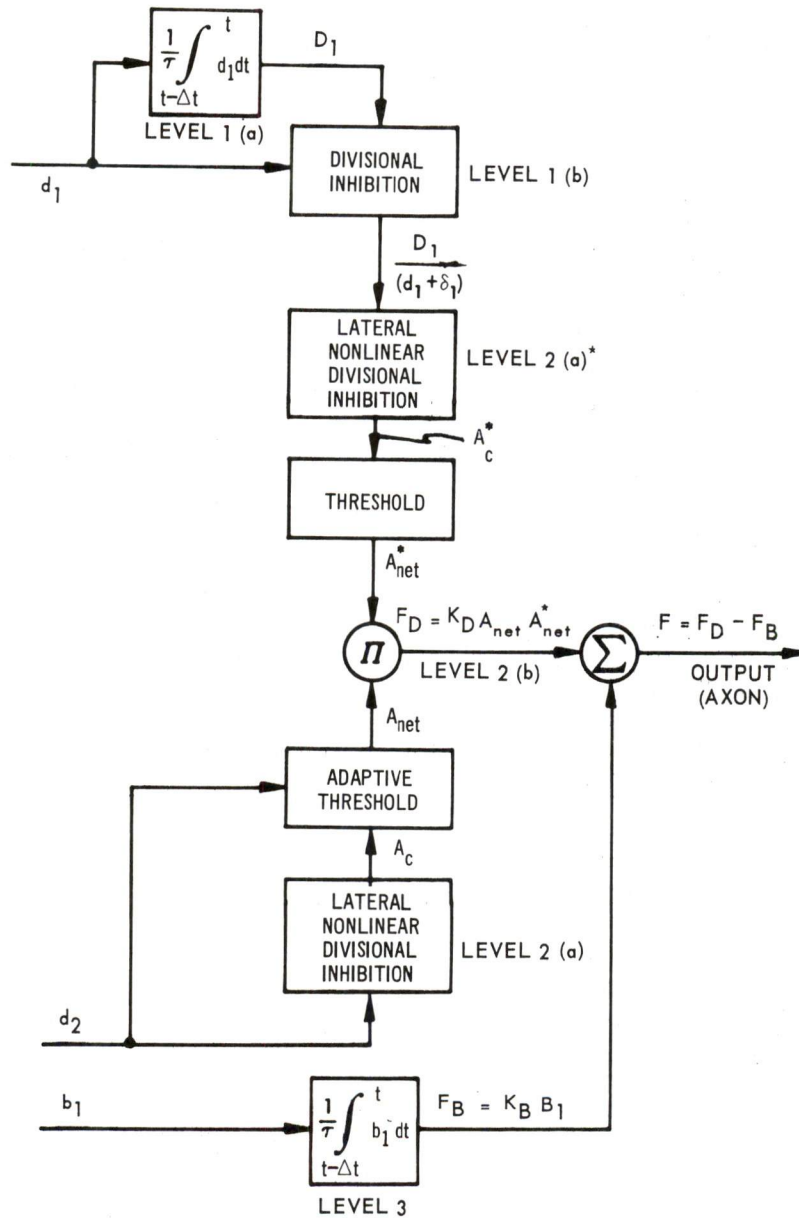


Fig. 16 Equivalent block diagram of Group 2 ganglion cell.

R-545

**INFORMATION PROCESSING AND
DATA COMPRESSION FOR
EXOBIOLGY MISSIONS**

by

Louis L. Sutro

Revised June 1966

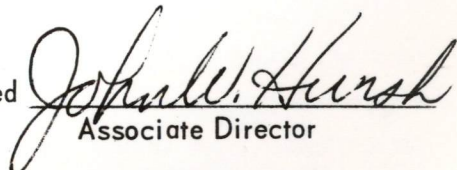
A report of work performed from
October 1965 through March 1966
under NASA Contract Number
NSR-22-009-138.

Presented at the 12th Annual Meeting
of the American Astronautical Society,
Disneyland Hotel Convention Center,
Anaheim, California, May 23-25, 1966

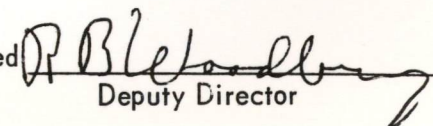
**INSTRUMENTATION LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS**

*Prepared for publication by Jackson & Moreland
Division of United Engineers and Constructors, Inc.*

Approved


Associate Director

Approved


Deputy Director

ACKNOWLEDGEMENT

This report was prepared under the auspices of DSR Project 55-257, sponsored by the Biosciences Division of the National Aeronautics and Space Administration through Contract Number NSR-22-009-138.

Publication of this report does not constitute approval by NASA of the findings or conclusions contained herein. It is published for the exchange and stimulation of ideas.

Biological help was provided by Jerome Y. Lettvin and Warren S. McCulloch, engineering help by William Kilmer, Roberto Moreno-Diaz, Jerome Krasner and Richard Catchpole.

ABSTRACT

Techniques being developed for missions to Mars are part of a larger effort to develop a robot capable of carrying out programs and of self-programming within certain limits. Three development efforts at the Instrumentation Laboratory lead in this direction: Design of a layered processor like that in a frog's retina shows how a model of animal vision may reduce data for transmission to Earth and permit response to the environment. Design of a stereoscopic system suggests how a pair of layered processors, when fed data from two television cameras, may find the positions of objects. Design of a decision and control system like that of animals suggests how a system of this kind may decide in what direction a Mars lander should move, when it should perform each experiment and what it should report.

TABLE OF CONTENTS

<u>Section</u>		<u>Page</u>
1	INTRODUCTION	1
2	THE LAYERED PROCESSING OF A VISUAL SYSTEM	2
	2.1 The Structure of a Visual System	2
	2.2 Properties of a Visual System	2
	2.3 Design of the Model (Ref. 5)	5
	2.4 Computation in Levels 1 and 2(a) *	9
	2.5 Computation in Level 2(a)	10
	2.6 Computation in Level 3	12
	2.7 Instrumentation with Available Components	13
	2.8 Instrumentation with Integrated Circuits	15
	2.9 Application to Missions to Mars	15
3	STEREOSCOPIC SYSTEMS	16
	3.1 Why Stereoscopic?	16
	3.2 Camera-Counter Chain C	17
	3.3 Camera-Computer Chain D	18
	3.4 Application to Missions to Mars	19
4	DECISION AND CONTROL PROCESSORS	21
	4.1 General	21
	4.2 Reticular Formation	21
	4.3 Model of RF*	25
	4.4 Visual-Center	28
	4.5 Application of RF* to Missions to Mars	29
5	THE NEED TO CARRY KNOWLEDGE ALONG	30
	LIST OF REFERENCES	31

INFORMATION PROCESSING AND DATA COMPRESSION
FOR EXOBIولوجY MISSIONS

1. INTRODUCTION

In designing interplanetary probes such as that we plan to put on Mars we are faced with the problem of exploring after the vehicle has landed. We are attempting to devise a self-moving probe capable of exercising judgments in a certain sense and handling a variety of contingencies. It would broadcast the results of its experiments rather than be a simple extension of our senses and movements into the alien world.

It is clear that technology supplies us increasingly reliable parts ever smaller and able to work on little power so that computers become more and more compact. At the same time the art of using these devices expands greatly so that they have ceased to be the novelties for study and have become, instead, common tools. A number of workers, concerned with this art (including us), are attempting to make a true robot, capable of self-programming within certain limits, one that is capable of handling a large sequence of choices, correcting itself when it makes an error in judgment. That is not to say that we are so rash as to claim that we can build an animal or an animal intelligence, but it would be intelligence of a sort, and one that would permit communication with us.

Each experiment proposed for Mars requires both a sensor system, designed for the problem, and a data processing system designed to bridge the gap between the sensor system and the way our minds work. We propose to subject the data of the experiments to a decision system similar to ours. What comes out of this robot must then comply with our criteria of interest and intelligibility.

Can this be done in time for a Mars landing in the 1970's? The only way to find out is to try.

This paper describes three development programs under way at the Instrumentation Laboratory. One is the design of the layered processing of a visual system. The second is the design of a stereoscopic system. The third is a decision and control system.

How would a robot reduce data? We propose that the robot carry with it measures of the usefulness to man of the data it acquires, so that it will send back only data which satisfies those measures. For each of the development programs to be described such measures will be suggested. But a more useful set of measures will have to come from you.

2. THE LAYERED PROCESSING OF A VISUAL SYSTEM

2.1 The Structure of a Visual System

Dr. Warren McCulloch refers to an eye as a "layered computer". By this he means that all the bipolar cells process the data they receive from the photoreceptor layer, passing it along to the ganglion cell layer. The latter processes the output of the bipolar-cell layer passing along the results of its computation to the layered structures of the brain.

We want to model the human visual system, but it comprises far more cerebral layers than we can model now. For that reason we sought an animal with fewer layers in its visual system but equal or greater complexity per layer. We found this in the visual system of the frog.

2.2 Properties of a Frog's Visual System

The species studied, the common leopard frog or *Rana pipiens*, responds to one color and to four pattern aspects of objects: edges, dark convex moving edge (bug), any time varying visual event, and dimming.

The retina of the frog has three layers of nerve cells: photoreceptors, bipolar cells, and ganglion cells, as shown schematically in Fig. 1.

Each cell type is uniformly distributed over the retina.

The photoreceptor layer consists of several types of rods and cones.

An individual photoreceptor apparently may have two output parameters.

The first appears to be a logarithmic function of illumination, related to

the bleaching rate of a pigment. A second parameter appears to be a function of the history of illumination, related to the amount of bleached pigment.

Bipolar cells are so called because their two ends are somewhat alike. Horizontal cells make interconnections among bipolar cells in the layer of photoreceptor outputs. There are interconnections formed by amacrine cells in the layers of connections between bipolars and ganglion cells.

Ganglion cells appear to have five distinguishable dendritic arbors, that is, tree-like branchings. Three of these arbors are shown as O, P, and Q in Fig. 2. On the basis of the connections of bipolar cells to these arbors, Lettvin et al. were able to assign one of five functions to each arbor form;^(1, 2, 3) four concerning the visual pattern and one concerning color. The four pattern outputs go to the tectum (Fig. 1 right) while the color output goes to the geniculate body (not shown). The first four cell groups produce outputs representing:

1. Edge detection
2. Bug detection
3. Event detection (any stimulus change)
4. Dimming detection.

Of the approximately 5×10^5 ganglion cells (Fig. 1) approximately half are bug detectors. Since this has proved to be the most difficult ganglion cell to model,^(4,5) we have concentrated on it.

Figure 3 portrays vertical sections of two bipolar cells and one multi-level E-shaped ganglion cell, as drawn by Ramon y Cajal for the frog.⁽⁶⁾ In the latter there are three dendritic levels. Lettvin et al identified this structure as a bug detector. Note that arborization exists between the two outer dendritic levels.

Two types of bipolar cells will be represented in the model. Their designation "on" and "off", is that of Schwyperheyn.⁽⁷⁾ The distinction between "on" and "off" cells was previously applied by Hartline to ganglion cells.⁽⁸⁾

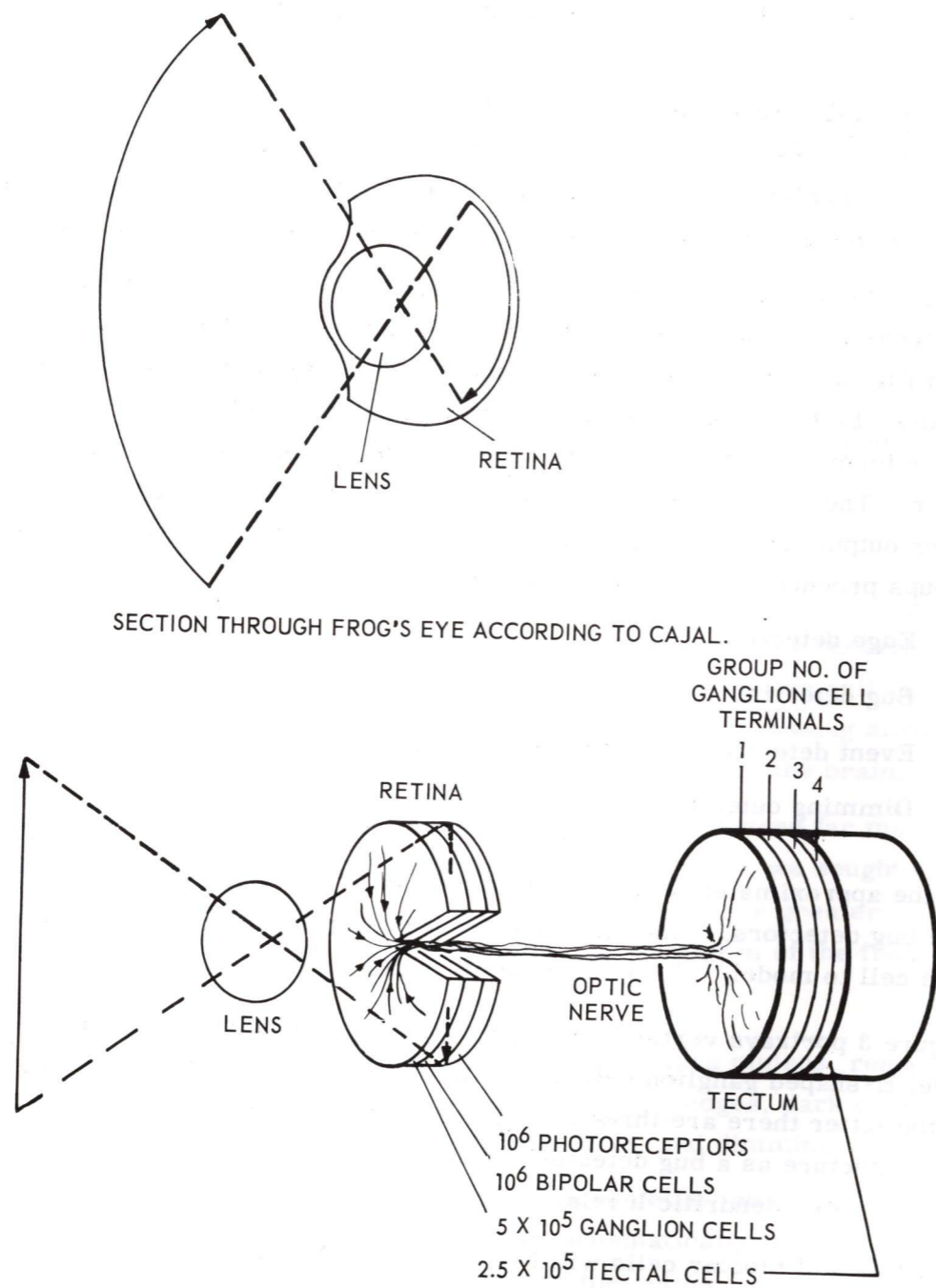


Fig. 1. Schematic of frog visual system.

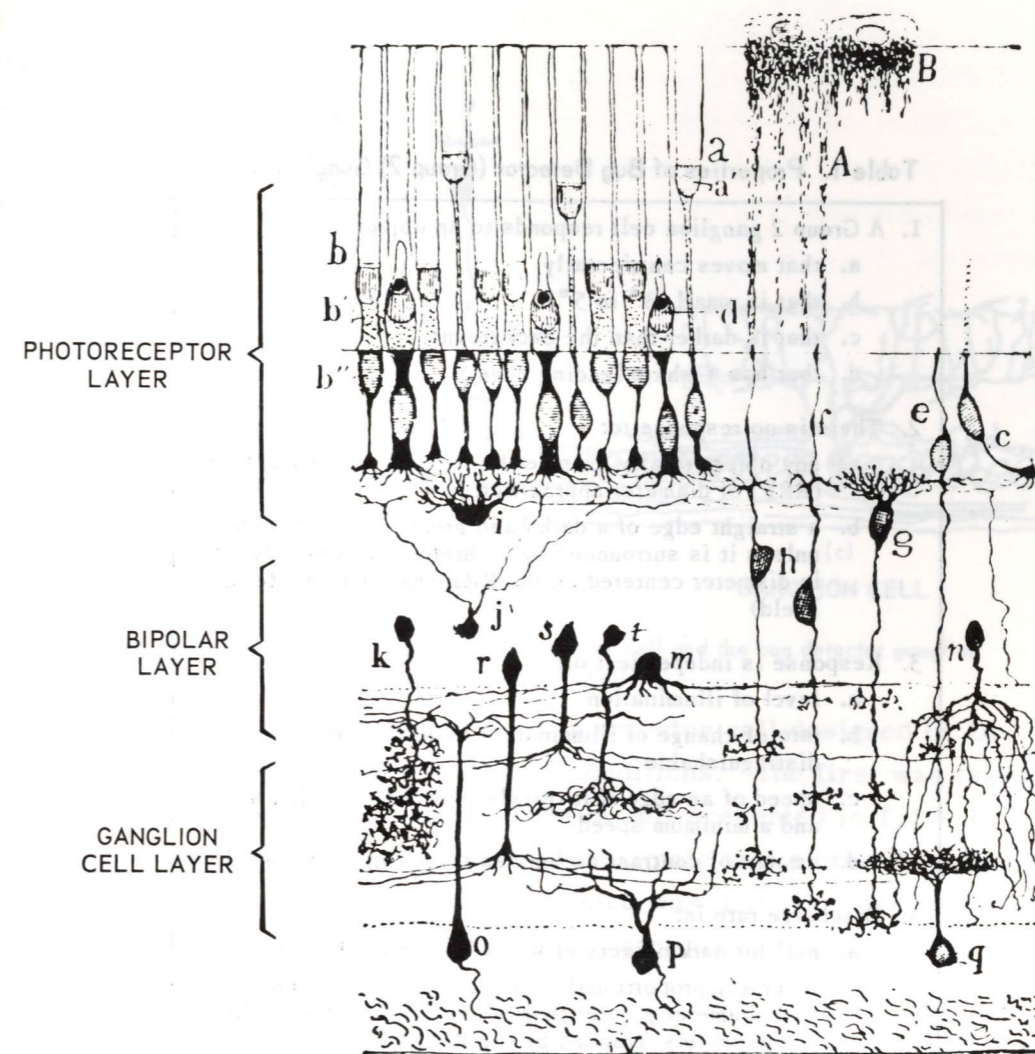


Fig. 2. The retina of frog in Colgi stain - highly schematic but showing spatial arrangement: g and h are bipolar cells; i is a horizontal cell; k, r, s, t, and m are amacrine cells; o, p, and q are ganglion cells. This is the off-hand diagram used for illustration in Ramon y Cajal's *Histologic du Systeme Nerveux* (Paris: Maloine, 1909-1911).

2.3 Design of the Model (Ref. 5)

Essentially, the bug detector cell responds to a small dark object which moves centripetally into its responsive field. In the list of the cell's properties given in Table 1, the smallness of the object is item 1b, darkness item 1c, centripetal movement item 1a. Note that the discharge rate of the cell (items 4a to 4g) depends upon the size and location of the object, both measured in degrees or minutes of arc at the lens of the frog's eye. Five minutes of arc are imaged on one photoreceptor.

Table 1. Properties of Bug Detector (Group 2) Ganglion Cell.

1. A Group 2 ganglion cell responds to an object :
 - a. that moves centripetally
 - b. that is small (3° to 5°)
 - c. that is darker than the background
 - d. that has a sharp leading edge
2. There is no response to:
 - a. any object totally outside the responsive retinal field (RRF) of diameter approximately 4°
 - b. a straight edge of a dark band greater than 2° wide unless it is surrounded by a shield approximately 4° in diameter centered on the RRF (Responsive Retinal Field)
3. Response is independent of:
 - a. level of illumination
 - b. rate of change of illumination as long as this is distinguishable
 - c. speed of an edge if it travels between a maximum and a minimum speed
 - d. amount of contrast as long as this is distinguishable
4. Discharge rate is:
 - a. null for dark objects of width less than about $3'$
 - b. inversely proportional to the convexity of an image with diameter greater than $3'$ but less than one half of the angle subtended by the RRF
 - c. maximal (approx. 40 pulses per second) for an image subtending half the angle subtended by the RRF
 - d. proportional to the convexity of images larger than one half of the RRF angle
 - e. feeble if the object is larger than RRF
 - f. greater to movement broken into several steps than to continuous movement
 - g. feeble in response to light spots
5. The response is maintained for several seconds if the object stops in the RRF

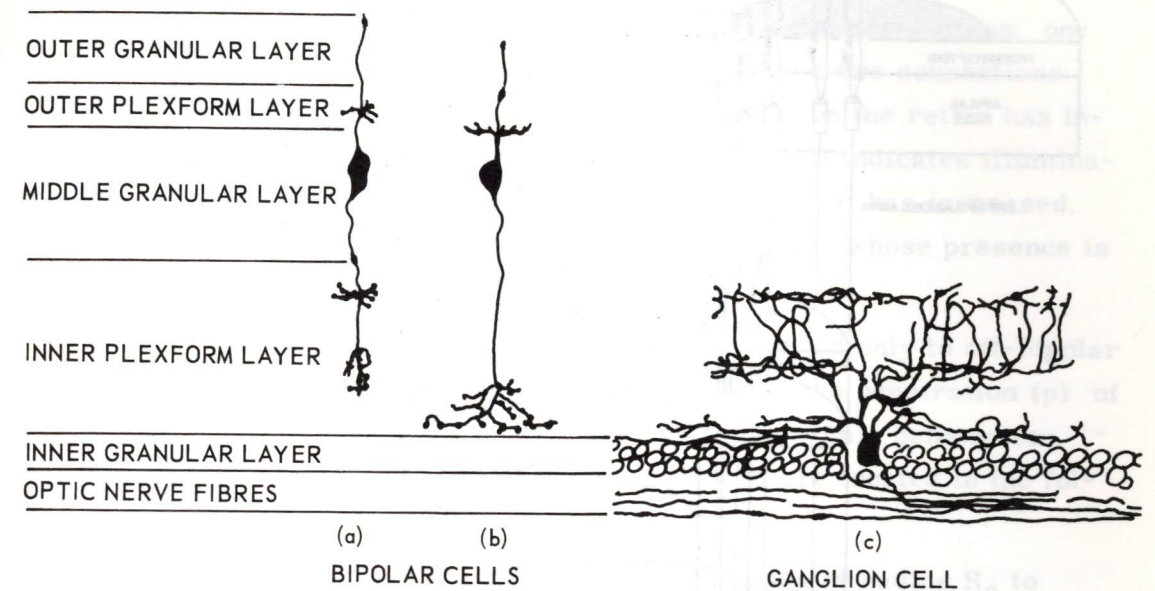


Fig. 3. Retinal sections of two types of bipolar cell and the bug detector ganglion cell in a frog retina (after Ramon y Cajal).

Figure 4 diagrams a model of the bug detector cell designed by Dr. Roberto Moreno-Diaz to meet three conditions. The first was that it fit the experimental data on this type of cell presented in Table 1. The second was that its structure approximate that pictured by Cajal and identified by Lettvin et al. The third was that all of the $1/4$ million cells of this type be capable of being built with existing hardware.

Dr. Warren McCulloch feels that in a model of perception, feedback plays an important role. Since this model employs no feedback it cannot be considered an attempt to solve the general problem of perception. It is rather a scheme to match experimental results of physiology with known anatomy, within the restrictions mentioned above.

At the top of Fig. 4, the image of a bug (shaded) is moving toward the center of concentric areas, of radii R_1 and R_2 . The area of radius R_1 (called responsive retinal field or RRF in Table 1) includes approximately 2000 photoreceptor cells, three of which are represented by small cylinders. The larger area represents approximately 20,000 cells; five of which are represented by the three same cylinders and two additional ones.

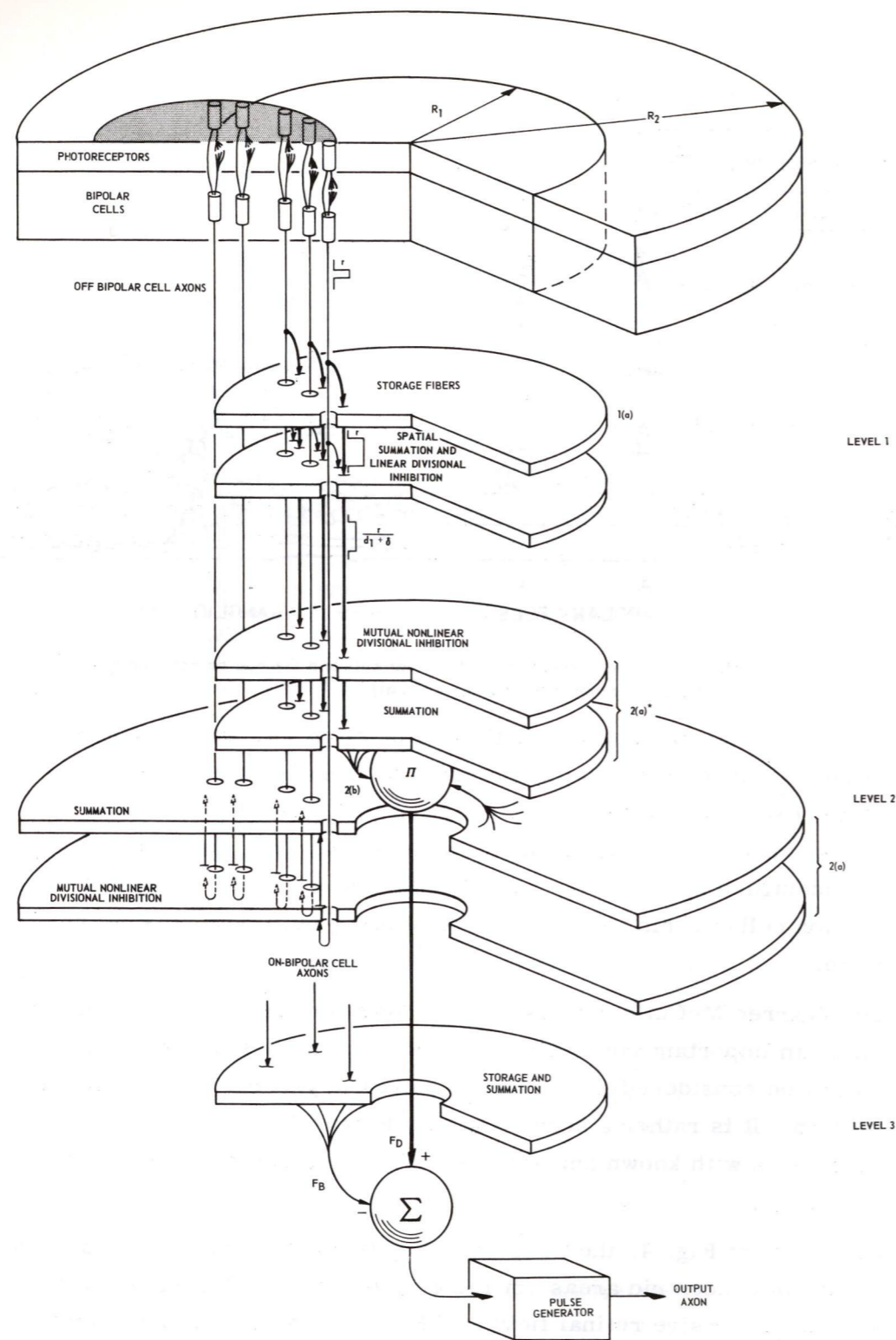


Fig. 4. Diagram of models of photoreceptors, bipolar cells and bug detector ganglion cell.

Between each photoreceptor and bipolar cell are two connections, one fast and one slow. By comparing inputs from these two connections, each bipolar cell determines whether illumination on the retina has increased or decreased. Firing of an off-bipolar cell indicates illumination has decreased, firing of an on-bipolar cell that it has increased. Only off-bipolar cells are shown. On-bipolar cells, whose presence is assumed, connect to level 3.

Levels 1 and 2(a)* of the ganglion cell model respond only to off-bipolar cells in the area of radius R_1 (RRF) to determine the penetration (p) of the bug into this area and the nonlinear response to this penetration. Computations performed in this layer and a half are treated in the following subsection.

Level 2(a) responds to off-bipolar cells in the area of radius R_2 to determine the convexity of the leading edge of the bug. This computation is treated in the second subsection following.

Level 3 responds only to on-bipolar cells in the area of radius R_1 (RRF) to determine, with previous computations, that the motion is centripetal. This computation is described in the third subsection following.

Activity of a bipolar cell is represented by a pulse of amplitude r , as shown in Fig. 4. This pulse is applied through a broad curving branch to level 1(a), through a fine curving branch to level 1(b) and through the lowest slab of level 2(a), to which it returns.

2.4 Computation in Levels 1 and 2(a) *

The pulse of amplitude r that passes through the broad curving branch of the off-bipolar cell axon is stretched in level 1(a), as shown in Fig. 4. While such neural storage usually has an exponential decay, it is here given a square waveform for simplicity. In level 1(b) the amplitude r is divided by the number of photoreceptors in the leading edge of the image, d_1 , plus a constant δ . The second branch — the fine one — makes possible the spatial summation that forms this sum d_1 .

In the first slab of level 2, the advancing pulses interact to inhibit each other in a mutual nonlinear divisional fashion suggested by Schuyperheyn.⁽⁷⁾ The stretched pulses, representing the length of time since

the image moved into the circle of radius R_1 , together indicate the area darkened, D_1 . By the formula for this kind of inhibition,⁽⁵⁾ the amplitude of the signal on the j^{th} wide line in the upper slab of layer 2(a)* is

$$A_j^* = c_1^* \frac{r}{d_1 + \delta} e^{-k_1^* \frac{rD_1}{d_1 + \delta}}$$

The asterisks or stars are applied to the symbols of this layer to distinguish them from the symbols of layer 2(a). The quotient $D_1/(d_1 + \delta)$, which can be approximated D_1/d_1 , is an area divided by the length of the leading edge. It equals approximately the penetration p of this area into the circle of radius R_1 (RRF).

The second slab of level 2(a)* forms a sum A_c^* of all the active A_j^* signals. There are D_1 of these.

$$A_c^* = c_1^* \frac{rD_1}{d_1 + \delta} e^{-k_1^* \frac{rD_1}{d_1 + \delta}}$$

$$\approx c_1^* r p e^{-k_1^* r p}$$

Employing the experimental evidence of Table 1, $k_1^* r$ can be evaluated, leading to

$$A_c^* = c_1^* r p e^{-\frac{p}{0.26R_1}}$$

which is plotted in Fig. 5. The constant c_1^* can be evaluated by a method indicated in Ref. 5.

2.5 Computation in Level 2(a)

The sum of the outputs of the off-bipolar cells, in the area of radius R_2 , in the last time period, is the "length of dimming" d_2 . The signal passed from the lower to the upper slab of layer 2(a) on the j^{th} line is

$$A_j = c_2 r e^{-k_2 d_2 r}$$

where c_2 and k_2 are constants. The interaction represented by this formula is again the "mutual nonlinear divisional inhibition" of J. J. Schypperheyn.

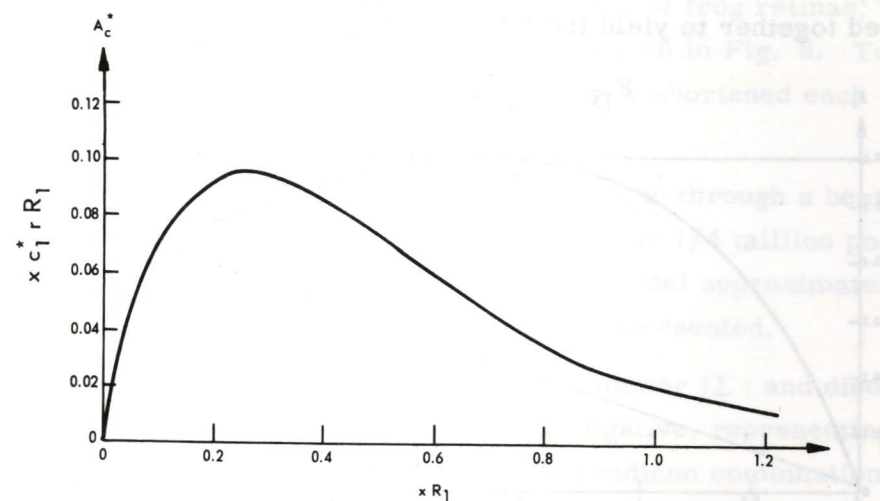


Fig. 5. A_c^* as a function of penetration p .

The upper of the two slabs of layer 2(a) sums the signals on the lines rising from the lower slab and compares the sum to a threshold θ . The sum is represented by A_{net} and can be computed by summing the A_j ,

$$A_c = \sum A_j,$$

and subtracting the threshold

$$A_{\text{net}} = A_c - \theta$$

Figure 6 presents a graph of the sum of the signals, A_c , rising from the lower slab. Figure 7 presents the difference A_{net} . The response is thus made dependent on the size of the object. Experiments on the frog indicate that the smallest object detected is about 3' (Table 1, 4b) of arc wide, that there is an optimum response to objects whose image area is about one half the area of radius R_1 and response falls to zero for objects whose image size is approximately this area, i. e., $d_1 \approx \pi R_1$.

At level 2(b), represented by a sphere, A_{net} , A^* and a constant are multiplied together to yield the function of dimming, F_D :

$$F_D = K_D A_{net} \cdot A^*$$

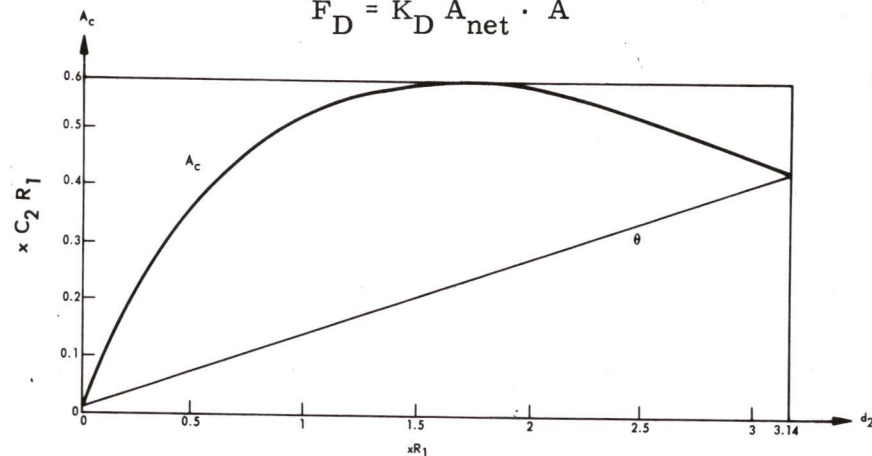


Fig. 6. A_c and θ as functions of the length of dimming d_2 .

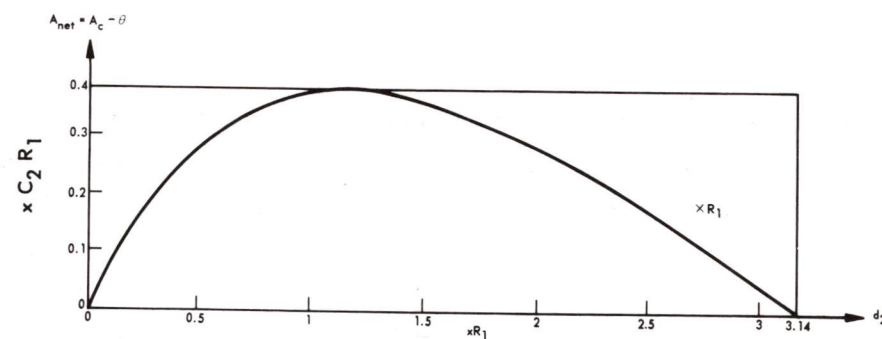


Fig. 7. A_{net} as a function of the length of dimming d_2 .

2.6 Computation in Level 3

The outputs of on-bipolar cells report brightening as the image of the bug advances. Level 3 stretches the pulses from on-bipolar cells, and sums them to give the function of brightening F_B . The summer at the bottom of the diagram takes F_D in a positive sense and F_B in a negative sense to get their difference. The pulse generator emits a frequency proportional to this difference. Thus, only as the dark image of a bug moves toward the center of the field of radius R_1 is there an output. By choice of constants, the pulse rate can be made maximum for an image that subtends half the angle subtended by the area of radius R_1 . This is the condition of maximum response of the living cell.

2.7 Instrumentation with Available Components

Using elements of previous designs of models of frog retinas,⁽⁴⁾ Dr. Moreno-Diaz devised the hardware model shown in Fig. 8. To permit placing the entire design on one page, we have shortened each vidicon into a wide flat bottle.

The two vidicons at the top view the same scene through a beam splitter. The raster of each tube comprises 500×500 or 1/4 million positions. If this resolution is maintained through the model approximately 1/4 million bug detector ganglion cells will be represented.

The bipolar cells are represented by the summer (Σ) and diodes that follow it. If the output of the summer is negative, representing dimming, the output goes to the cathode ray tube and vidicon combination immediately below it. If the output is positive, representing brightening, it goes to the CRT-vidicon in the lower part of the illustration.

The model of the bug detector ganglion cell is divided into three levels as in Fig. 4. In level 1 of Fig. 8, dimming is mapped on the face of the cathode ray tube. Fibre optics communicate the slowly fading image to a vidicon. The combination of tubes, commercially available as a scan conversion tube, performs temporal summation.

The output of the vidicon goes to a pulse generator (PG) which delivers its output to a bank of shift registers which map 50 lines of the vidicon raster. Attached to the shift registers and representative of the area of radius R_1 in Fig. 4 is a summer (Σ) which determines the part of the area of radius R_1 that has been dimmed, D_1 . The bank of shift registers work this way: As the first position of the first line of the raster is scanned, a 1 or 0 is fed to the first shift register. As successive positions are scanned, successive 1's or 0's are fed to the first shift register. At the start of the second line of the vidicon raster, the 1's and 0's in the first shift register are fed to the left end of the second shift register. After 50 lines have been scanned 1's and 0's representing the dimming in these lines will be loaded in the bank of 50 shift registers. From then on the registers overflow but always contain 50 lines.

The length of dimming d_1 is computed in the second bank of shift registers and combined with D_1 in the box labelled \div . The quotient D_1/d_1 is the

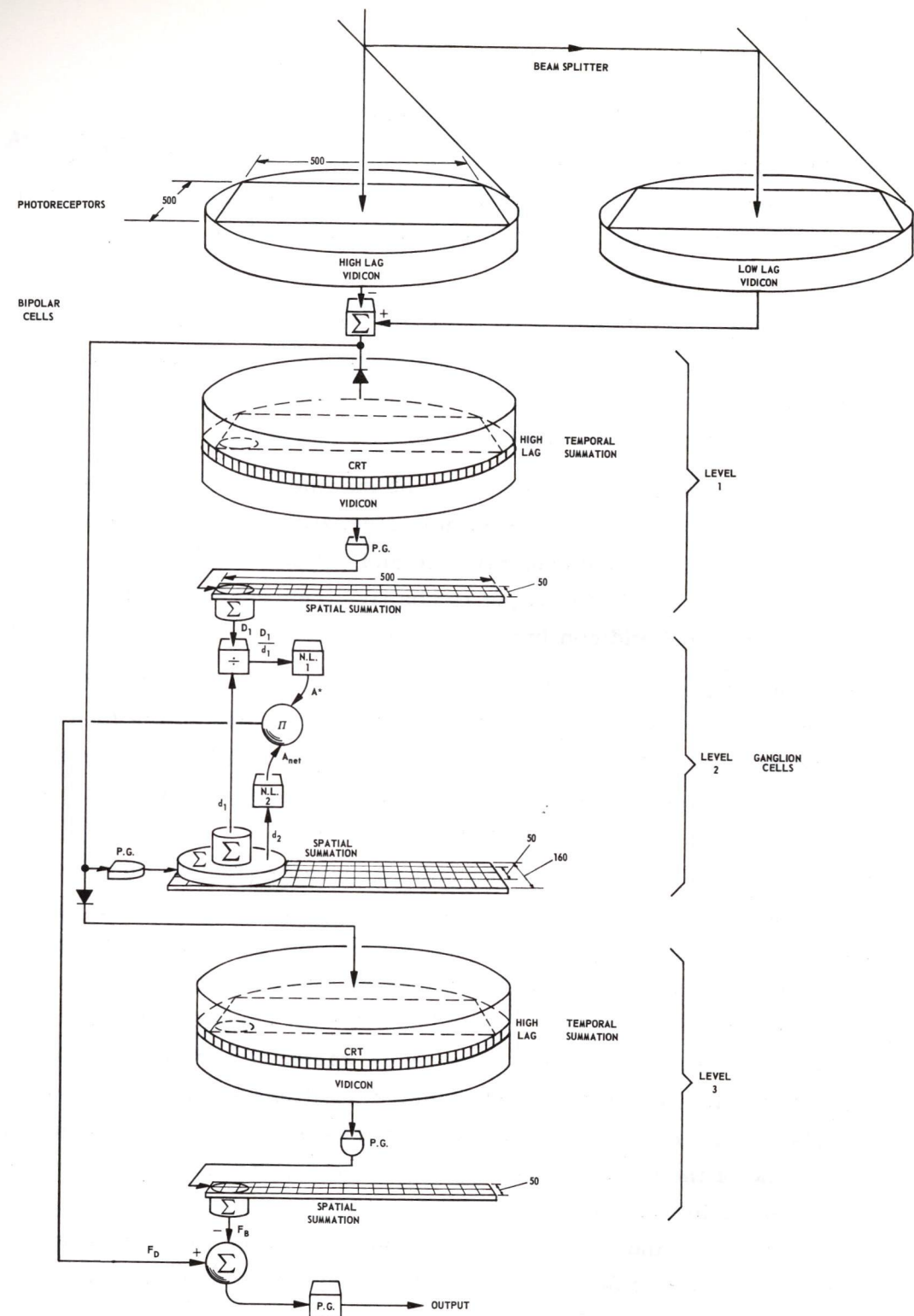


Fig. 8. Possible instrumentation of 1/4 million Moreno-Diaz models of the bug detector ganglion cell.

penetration p of the image of the bug into the area of radius R_1 . The box labelled "N. L. 1" performs the nonlinear transformation of Fig. 5.

The second shift register and its large-area summer (Σ) computes the length of dimming d_2 . This is fed to Nonlinear Circuit 2 which performs according to Fig. 7. The analog multiplier labelled π combines A_{net} , A^* and a constant to form F_D .

In level 3 a second combination of cathode ray tube and vidicon determine the areas brightened in the whole retina. A third bank of shift registers and summer determine the brightened area B_1 for each cell. F_B is proportional to B_1 . The final summer subtracts F_B from F_D . The final pulse generator yields a pulse frequency proportional to this difference.

2.8 Instrumentation with Integrated Circuits

We are now redesigning the system of Fig. 8 to employ integrated circuit elements instead of vidicons and cathode ray tubes. Figure 9 shows an array of 50×50 photo transistors developed by Westinghouse.⁽¹⁰⁾ That company has since built four times as many units in the same area.

We are on the way, we think, of devising a way of using this kind of array. From our point of view, it has two difficulties. One is the difference in response characteristics between phototransistors. The other is defects in phototransistors, i. e., some may not function at all. In a system we are designing the output of each phototransistor will be compared to its own previous output. Thus discrepancies between phototransistors should not enter the computation.

2.9 Application to a Mission to Mars

Design of a layered processor to recognize a bug of a particular size and travelling in a particular direction indicates that other recognition systems can be designed both to recognize predetermined objects and objects newly experienced. Such recognition is a reduction in data, because it permits reporting simply the presence of the objects. The system about to be described should also permit reporting the location of an object.

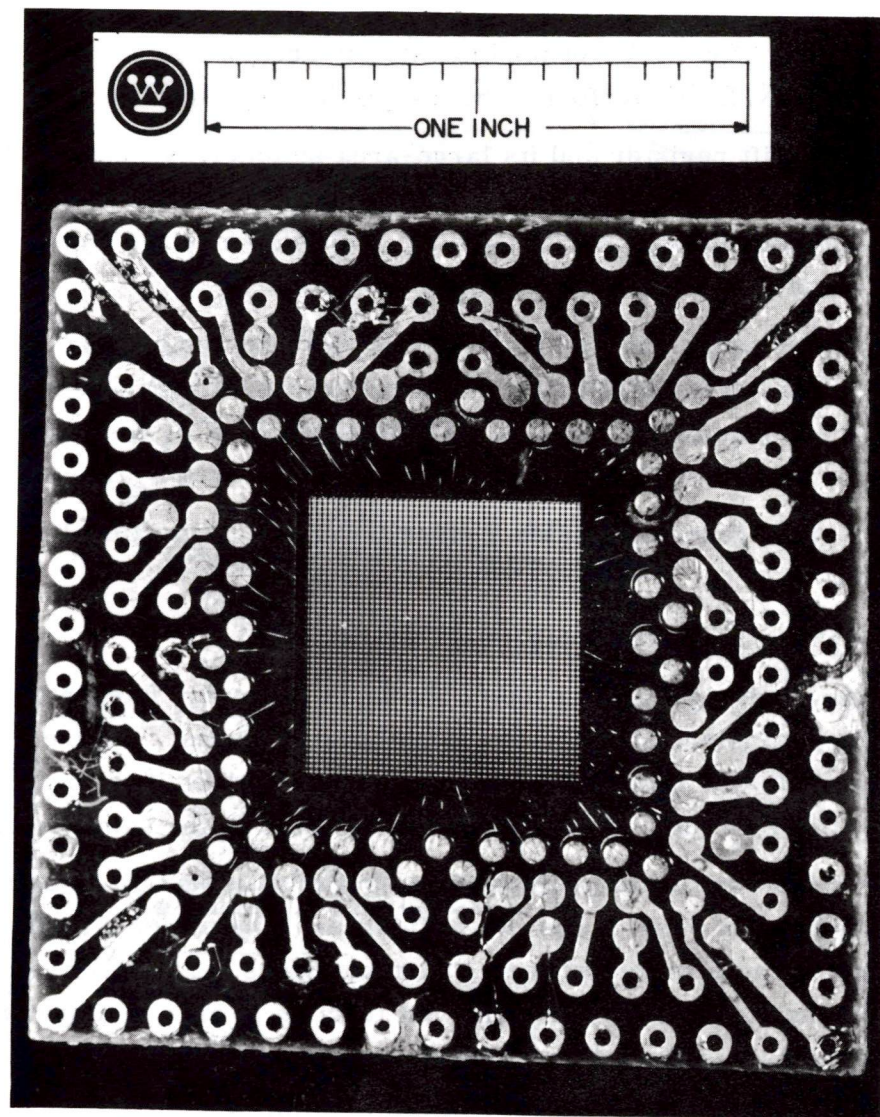


Fig. 9. Array of 50×50 phototransistors developed by Westinghouse.

3. STEREOSCOPIC SYSTEMS

3.1 Why Stereoscopic?

One of the tasks a robot may have to perform is to discover separate objects so as to avoid them, if moving, or describe them in detail if they have properties of interest. Several strategies are possible in the use of a pair of television cameras coupled to a computer. One that we are considering is diagrammed in Fig. 10.

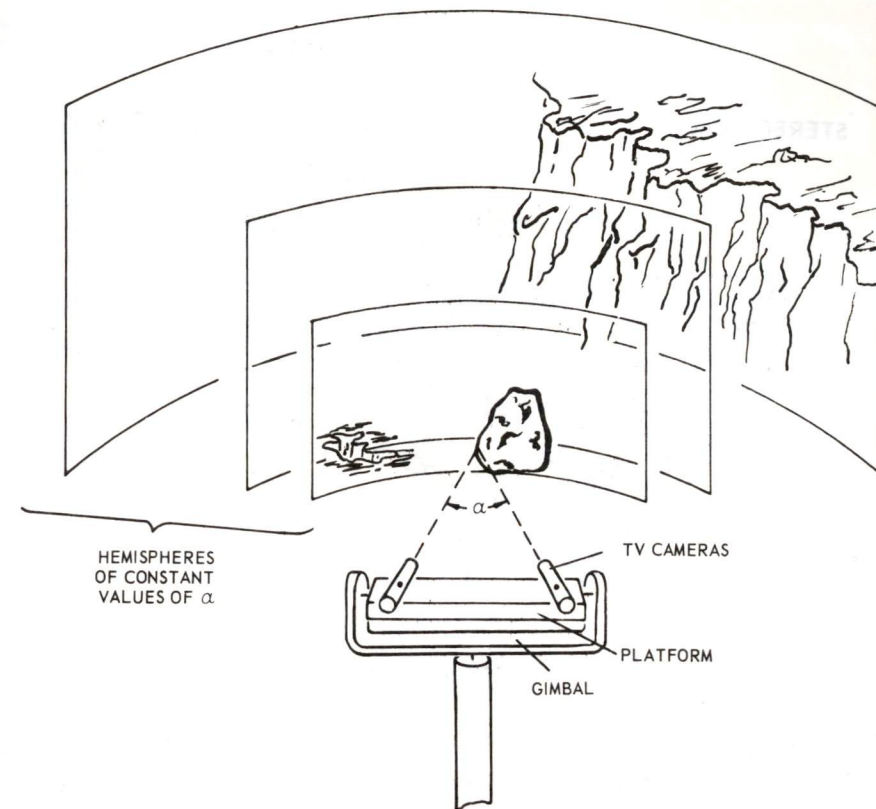


Fig. 10. A strategy to discriminate objects by a stereoscopic pair of TV cameras and a computer.

Here the cameras have mechanically scanned the innermost of three hemispheres until they have encountered the edge of a rock. The hemisphere scanned is that for which the angle of convergence of the two cameras is the value of the angle α shown. For a smaller value of α , the hole could be encountered. For a still smaller value, the edge of the cliff could be encountered. The problem is how to bring about the encounter.

3.2 Camera-Counter Chain C

To carry out initial experiments, we have attached a beam splitter to the front of a television camera, as shown in the illustration of our Model C camera computer chain (Fig. 11). The geometry of the two views is shown in Fig. 12. The prism of the previous illustration is here omitted for simplicity and the single camera shown as two. The left camera looks out through the lower cone, the right through the

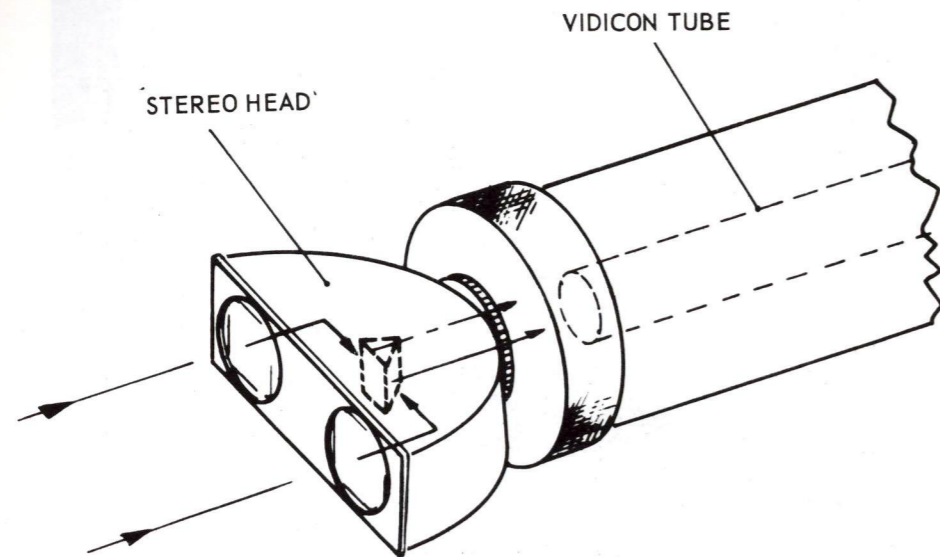


Fig. 11. Front end of Model C camera computer chain.

upper cone, the two cones overlapping some distance in front of the cameras. The point P on the edge of the right cone is near the center of the lower cone. It is imaged at P'_R on the right half of the camera, at P'_L on the left half of the camera.

Left and right TV frames have been drawn in front of the cameras to illustrate the method of computation. To search for an object such as that with the angle of convergence α shown in Fig. 11, windows will be advanced from left to right across the frames. Computation then can be limited, at any instant, to interpretation of the data in the two windows.

3.3 Camera-Computer Chain D

A constraint on camera-computer chain C is that it employ a lens of average focal length. Thus each camera sees neither a full field of view nor the narrow field of view that provides high resolution. To obviate this difficulty, camera computer chain D is being designed. (See Fig. 13.) It has interchangeable lenses and each camera is mounted to swivel. In addition, the cameras are mounted on a platform that nods.

Figure 14 shows a possible configuration of cameras and computers to visually explore a scene and report, according to the strategy described

above, first the presence of free-standing objects then details about one or a few of these objects.

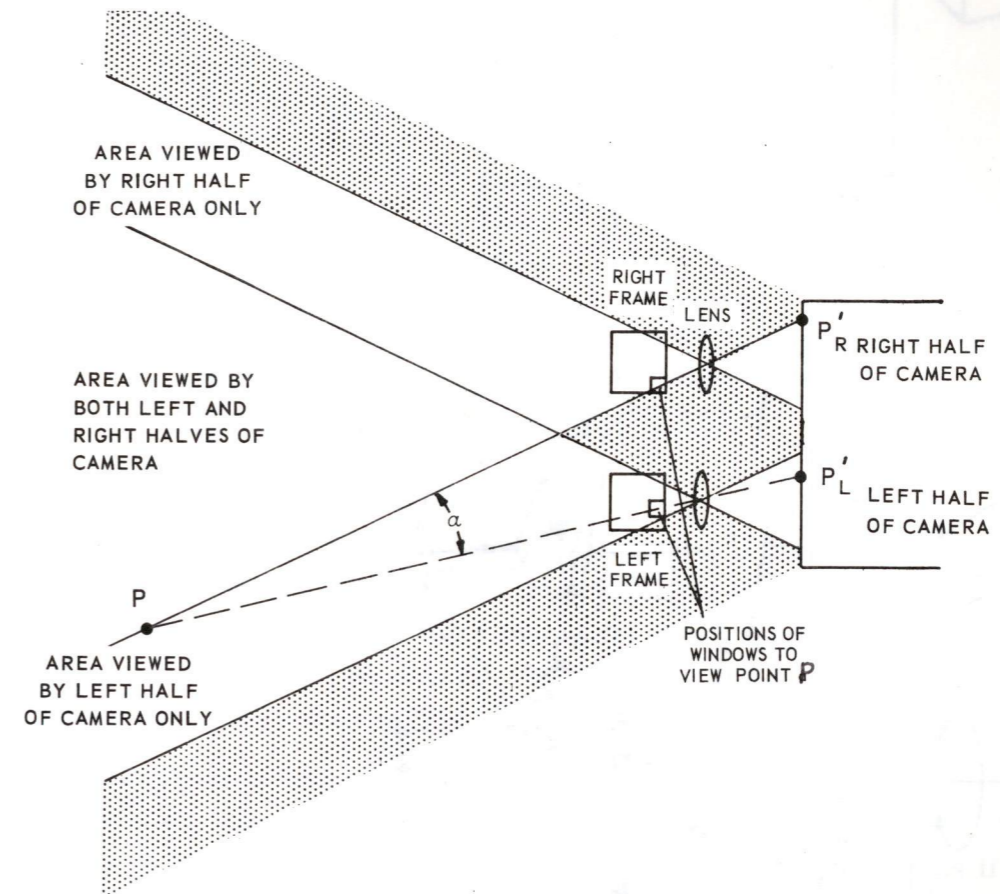


Fig. 12. Schematic of front end of Model C camera computer chain.

3.4 Application to Missions to Mars

If several free-standing objects are detected, how is one to be selected for more detailed examination? An animal could use another sense, such as the sense of touch. We are planning senses other than that of vision. In the following section we refer to these as sense modalities. The problem then arises: How should a robot combine the inputs from different sense modalities to decide what to do?

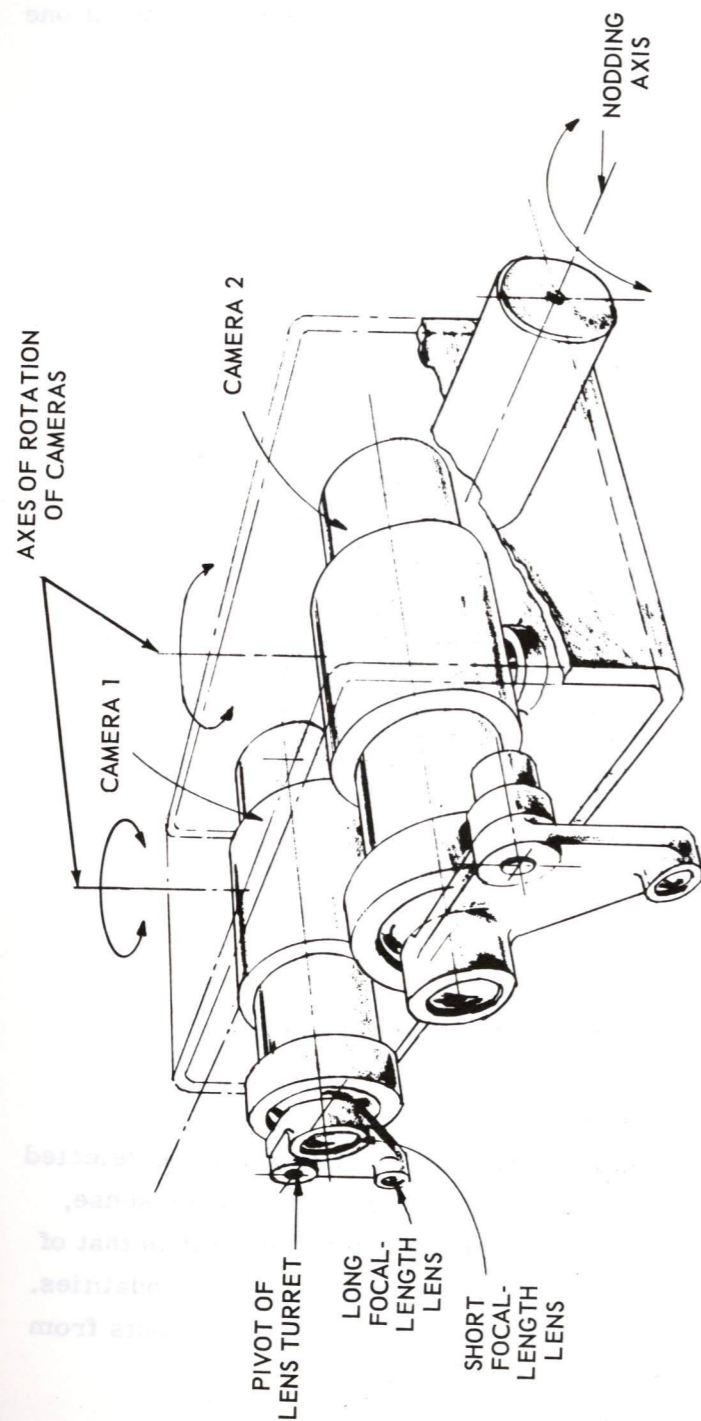


Fig. 13. Front end of Model D camera computer chain.

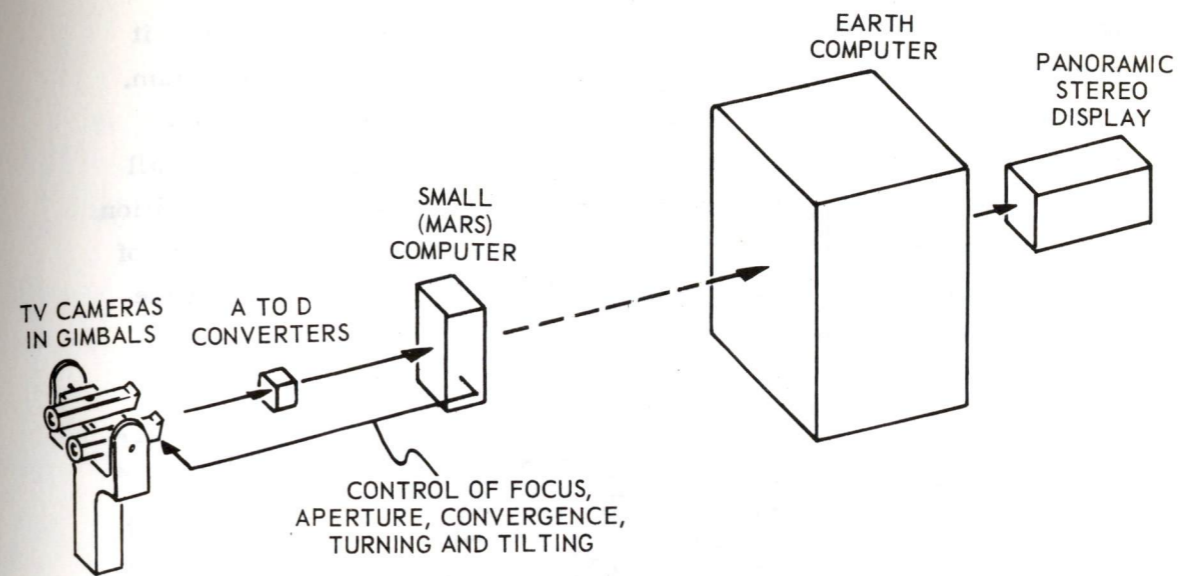


Fig. 14. Block diagram of Mars-Earth camera computer chain.

4. DECISION AND CONTROL PROCESSORS

4.1 General

As we enter into the field of layered processing of information we are in strange territory. There is no mathematical theory of nonlinear integral transforms. There is no theory for handling purely informational (as opposed to power) feedback. Group theory is concerned with better defined cases than what we are handling.

But, if layered computation seems the method used by all brains, it may be one of the necessary conditions for doing what an animal does. There is no way by which sequential operations map parallel operations if the elements in a layer are also mutually connected, as seems the case in most of brain. Thus we have had no recourse but to start simulating and trying before we adduce general design principles.

Two ventures in computer design beside the design of eyes will now be presented. First is the model of reticular formation designed by Drs. McCulloch and Kilmer.⁽¹¹⁾ Second is a processor for the output of our "eye" (described under "Visual Center" below).

4.2 Reticular Formation

The reticular formation (RF) integrates the sensory-motor and autonomic-nervous relations so as to permit an organism to function as a unit instead

of a mere collection of organs. The core of all nervous systems, it extends throughout the length of the spinal cord and into the cranium. (See Fig. 15.) It receives inputs from all of the separate sensory systems, auditory, visual, olfactory, vestibular, etc., and from all of the housekeeping systems involving visceral regulation, respiration, cardiovascular control, etc. Figure 16 shows the wide flat spread of the reticular cells to which the inputs connect and between which the reticular cells interconnect.

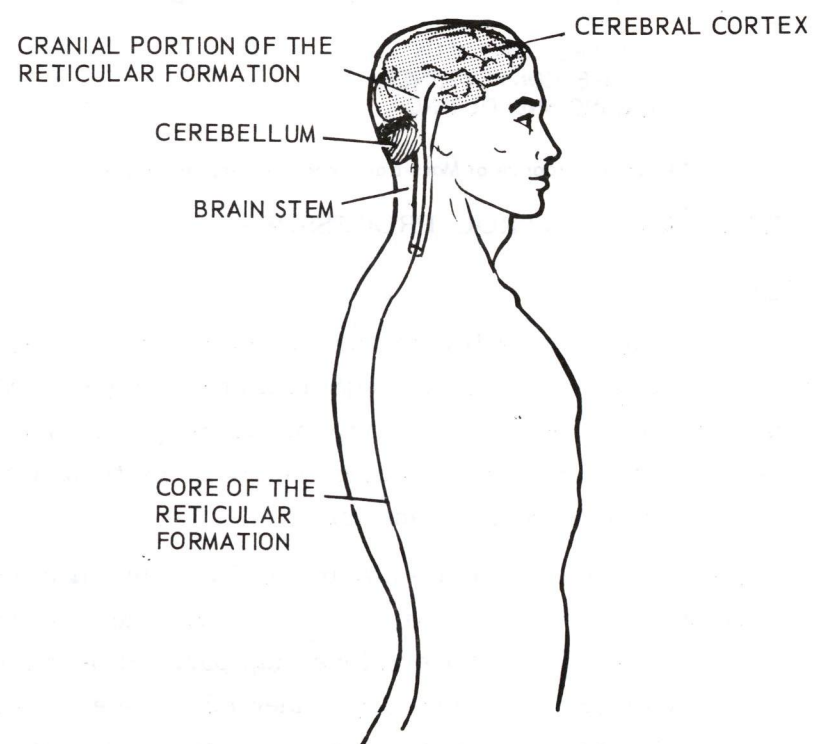


Fig. 15. Core of the reticular formation in the cranium, brain stem and spinal cord of man.

As Kilmer et al⁽¹²⁾ say, "The Scheibels⁽¹³⁾ have so far done the most definitive neuroanatomy that we have on the reticular formation. In their milestone report of 1962 they caricatured its anatomical structure in the brain stem by comparing it to a stack of poker chips. In each chip region the dendritic processes of RF neurons ramify in the plane of the chip face, often covering nearly half of the face area. This causes a very large degree of overlap and intermingling among dendrites of

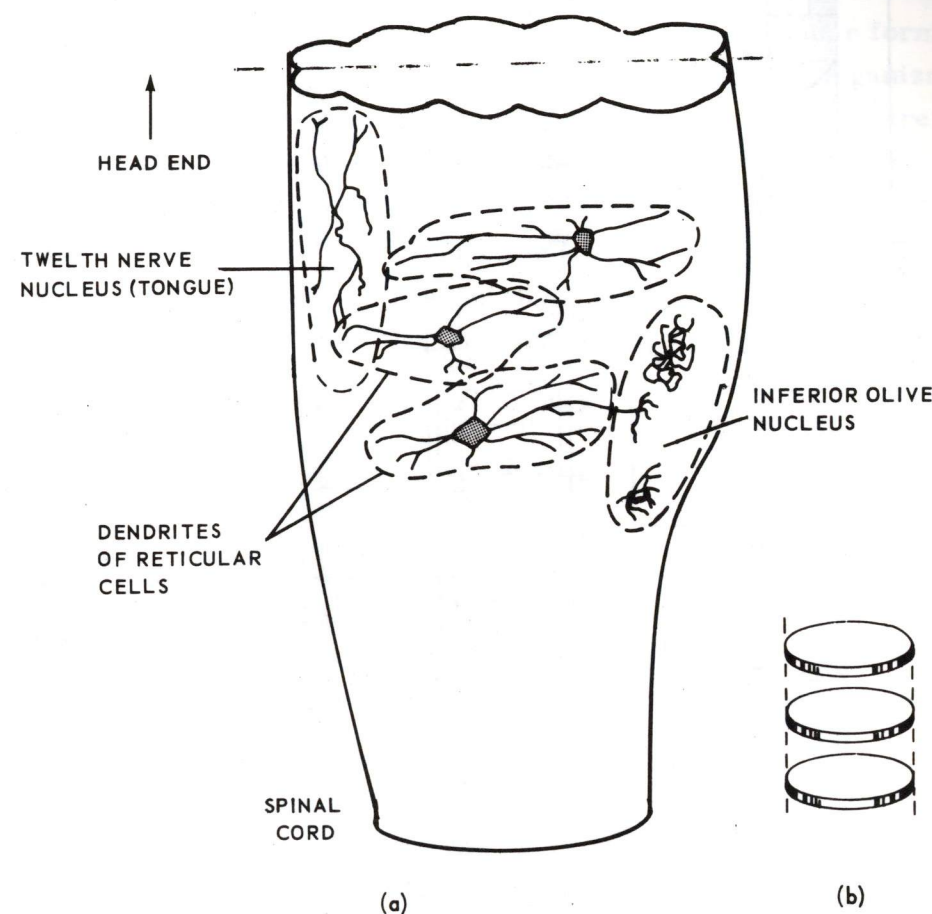
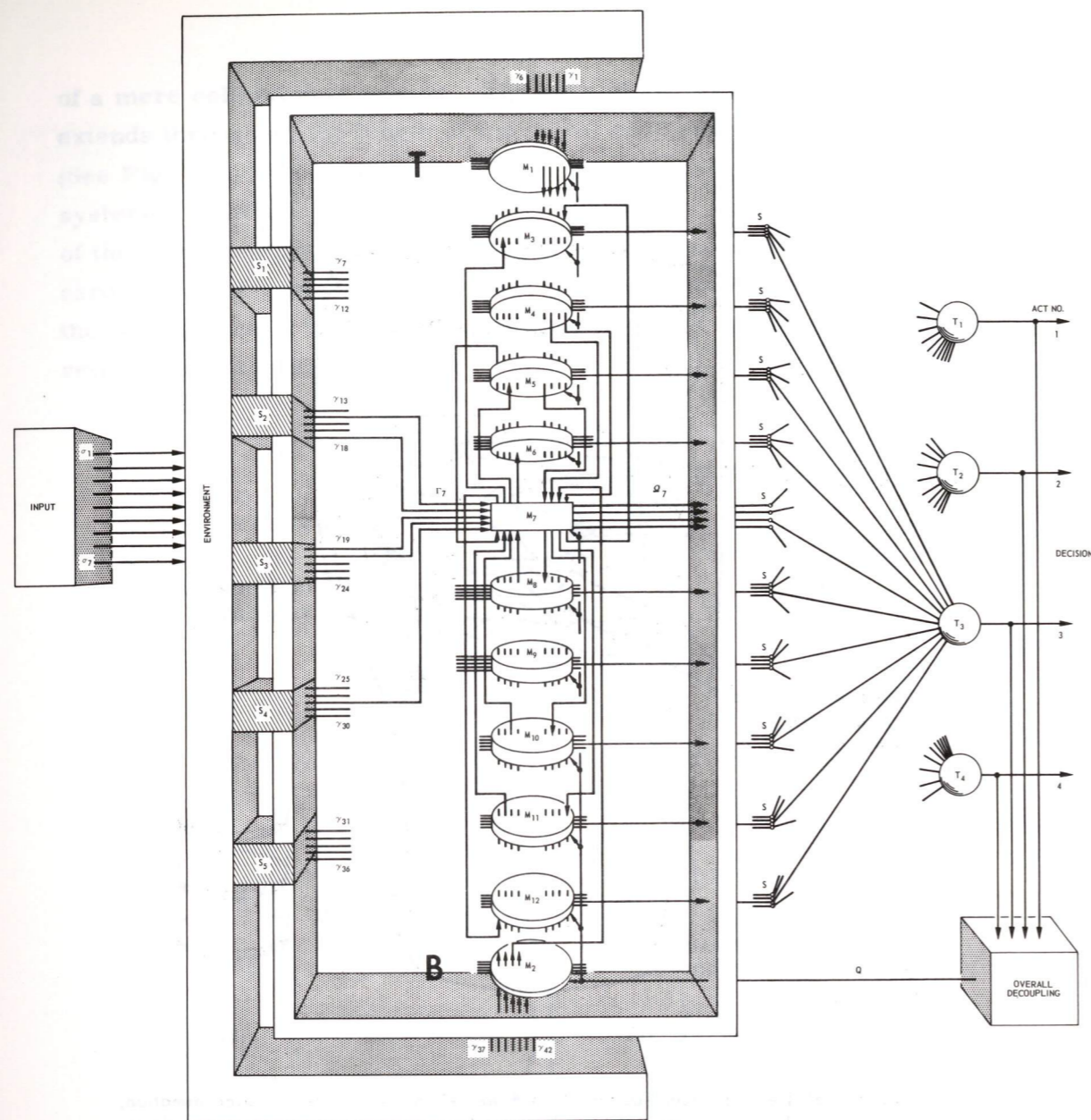


Fig. 16. (a) Dendritic ramifications from three cell bodies in the reticular formation, from two in the inferior olive and from one in the twelfth nerve nucleus (kitten brain, Scheibel, 1962)

(b) Poker chip caricature of reticular cell dendritic ramifications in (a)

nearby neurons, as shown schematically for just the brain stem region in Fig. 16. (This is very similar to Scheibels' Figure 1 in reference 13.) . . . Since as many as half dozen or more input systems may synapse on a single RF neuron, and each RF input nucleus and fibre tract in general feeds very many RF cross-sectional levels, the Scheibels suggest that the RF might tolerate considerable puddling of information at each of its cross-sectional levels, but demand somewhat



Legend (from left to right)

σ_i	factor of the environment	M_i	module i
S_i	sense	Ω_i	four components of the probability vector from module i
γ_i	input	s	step function
Γ_i	inputs to module i	T_i	threshold element

Fig. 17. Schematic of the computer-simulated RF*.

greater informational rigor between levels. Figure 16(b) is a caricature of the dendritic fields, comparing them to a stack of poker chips.

"In its non-specific control of sensory inputs, the reticular formation is analogous to an admiral-of-the-fleet, committing the organization under its command to one act, trusting that fine perceptions are made at their centers of specialty and are accurate. As a computer, the reticular formation is seen to be broad in its domain of command but exceedingly shallow in any specific area under its command. It commits the organism to an act which is a function of the information that has been played upon it in the last fraction of a second or so. After commanding the organism to act, e.g., fight, run, swallow, vomit, sleep, copulate, etc., it sends out control directives to all of the specialized centers of the central nervous system, tuning their activities to this task.

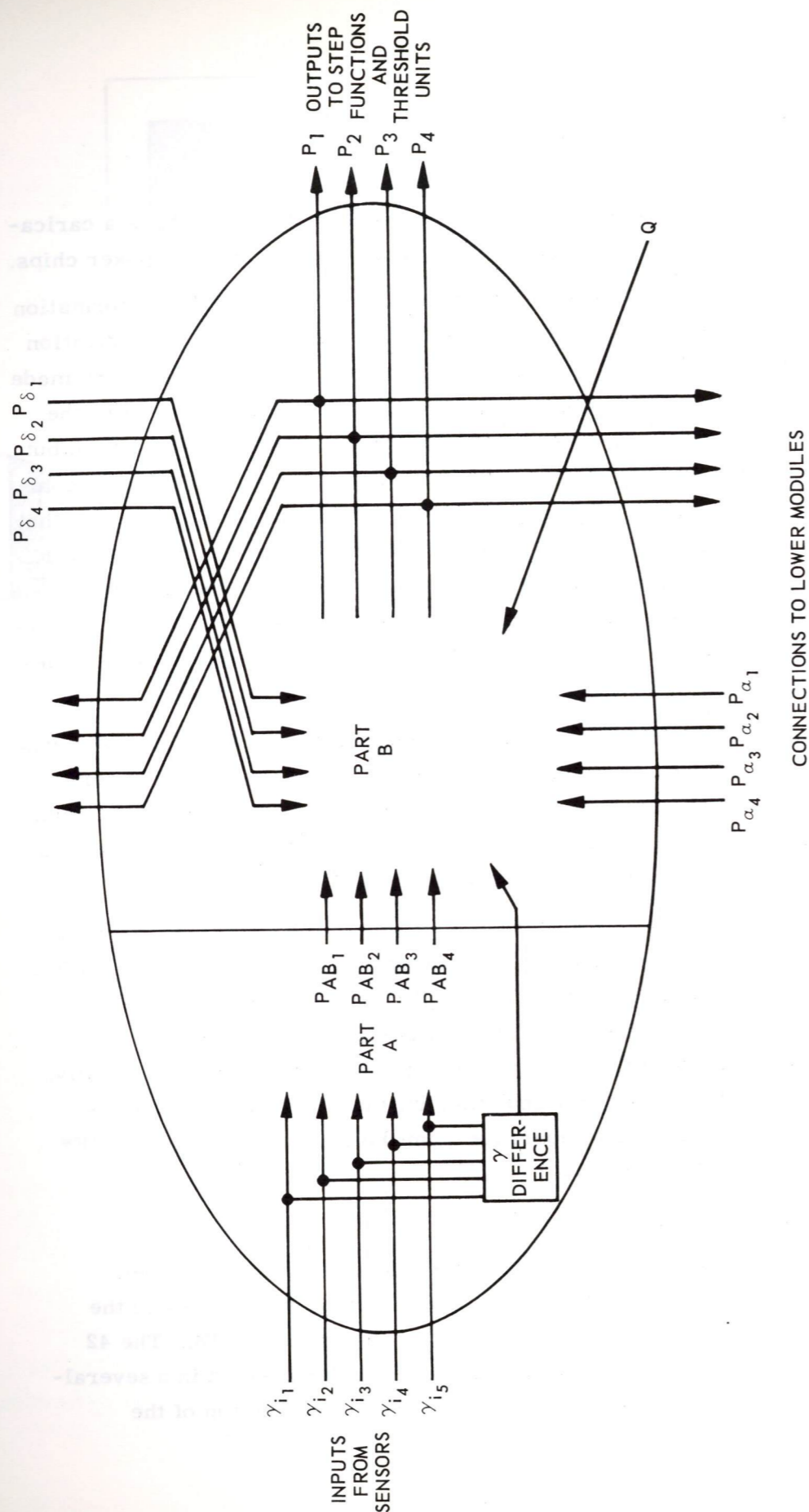
"Its important features are that it must handle a vast amount of highly correlated input information and arrive at one of a small number of mutually exclusive acts in a dozen time steps or so and with minimal equipment. The crucial information as to what act the organism has selected is distributed over its input lines.

"In this paper we are concerned only with that part of the RF which makes the decision to commit the organism, henceforth denoted RF*. RF* is RF minus everything on the RF's input side (the dorsal-lateral RF) and output side (the ventral-lateral RF, basal ganglia, etc.), all of the reflexes along the neuraxis that are handled locally, and all of the respiratory and other rhythmic operational aspects which are functionally separable from the decisionary task." (RF* is called \overline{RF} in Ref. 11.)

4.3 Model of RF*

"Figure 17 shows a schematic of a decision and control model analogous to the reticular formation. Each of the modules in the figure is a hybrid computer of the type shown in Fig. 18. The 42 input lines, γ_i , to the modules in Fig. 17 are connected in a several-to-all, but not all-to-all, fashion to the modules. The top of the

CONNECTIONS TO HIGHER MODULES



CONNECTIONS TO LOWER MODULES

Fig. 18. Input and output connections to parts A and B of a module.

drawing (T) and the bottom (B) correspond roughly to the diencephalic and high cervical regions, respectively, of higher vertebrates. The white-boundaried half-box represents the environment of the organism being modelled. The environment is established by factors, σ_i , of the set Σ . The S_i represent exteroceptive, interoceptive, and internuncial systems which sense this environment and feed highly (but nontrivially) correlated inputs, γ_i , directly into the model. For convenience, all σ_i and γ_i lines are binary.

"There are twelve logic modules in the simulated RF*. The sets of lines Ω_i (only Ω_7 is marked) indicate the act preferred by module M_i . There are four lines in each set because the model is at present capable of four acts. The four lines carry the four components of a probability vector.

"For clarity, all types of connection lines (Γ, Ω , ascending and descending lines) are shown only for module M_7 , whereas these types of connection are actually made to all regular modules. Each M_i receives inputs from several but not all S_j , and each S_j feeds several but not all M_i .

"Each module (Fig. 18) computes directly on the input information that it receives and makes a best guess as to what the corresponding act should probably be. After the initial guess, the modules communicate their decisions to each other over low capacity information channels. Then each module takes the information (from all other modules) and combines it in a nonlinear fashion with the information coming directly into it to arrive at a mixed guess as to what act should be performed. This is in turn communicated to the subset of modules to which it is connected above and below. Thus we have decomposed the module shown in Fig. 18 into two parts. The A part operates on the module's input information. The B part operates on information from above, below and from the a part.

"The A part with five binary input variables and four analog output variable outputs, is a nonlinear probability transformation network."

Quoting Ref. 14, "The B part receives 4-component probability vectors P_δ from the above, P_α from below and P_{AB} from the A part. The j^{th} component of each probability vector is the probability, computed by the module of origin, that model RF*'s present γ input signal configuration demands act number j . The B part also receives the difference between γ inputs, called " γ difference" in Fig. 18, and the overall decoupling Q shown being formed in Fig. 17. The B part yields the module's determination of the probability that the model RF*'s present γ input signal configuration demands act number j .

"The design of the modules was straightforward. The main problems concerned the way in which the computation converged to produce consensus. This consensus is achieved, as illustrated in Fig. 17, by first determining at point s if the j^{th} component of the probability vector P from each module exceeds 0.5. If it does, a 1 is passed on to the threshold element T . There, if 60% of the input connections are 1's, act j is decided upon. Note that each element T in Fig. 17 receives 10 inputs. For clarity of the drawing connections are shown only to T_3 . The threshold elements T are crude models of the muscle systems of animal which receive many inputs, decode them and act.

"The model has been successfully simulated on the M. I. T. Instrumentation Laboratory Honeywell Computer in collaboration with Jay Blum, E. Craighill and D. Peterson. The model converged to the correct act in each of about 50 test cases, and always in from 5 to 25 time steps.

"We are now concentrating on the functional design of a considerably enriched model which can handle conditioning and extinction in a satisfactory time domain sense."

4.5 Visual-Center

Walter Pitts once pointed out the consequences of having sequences of widely connected layers such as exist in a brain. It is that there will exist "stimulus" equivalences in the output — and the classing of stimuli

into "similar" and "dissimilar" may not necessarily follow those which would be natural to us. Nor could we from analyzing the connectivities in the net determine what these classifications would be.

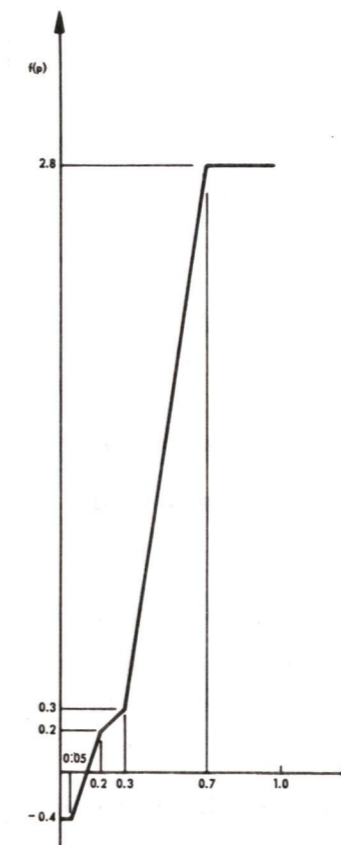


Fig. 19. The $f(p)$ function.

We intend, however, to superpose on our "eye" other layered nets so as to come, by patch and try, to a kind of excerpting of forms from the image as classes in the output. We have no idea yet of our chances for success. But the method must be tried.

4.6 Application of RF* to Missions to Mars

As Dr. Kilmer puts it, ⁽¹⁵⁾ "Our next task will be to sufficiently enrich the present model of RF* to enable it to exhibit several types of learning. This will require the insertion of delay chains between modules and the addition of adjustable decision circuitry in the a part of each module."

A task beyond this will be to tie the visual center to the model of RF* through far more channels than are shown in Fig. 17. Other sense modalities will be needed such as the sense of touch. Dr. Kilmer suggests that on Mars the model of RF* may have to decide among the following acts:

1. Rest
2. Locomote to _____
3. Turn around
4. Preprogrammed operation mode I
5. Preprogrammed operation mode II
6. Preprogrammed operation mode III
7. Maintenance
8. Communication to space vehicle (landed) I
9. Communication to space vehicle (landed) II
10. Righting (after overturn)

For example, if operation mode I were a visual experiment, it would not be switched on in darkness or a very high wind. For another example, if operation mode II were a soil assay experiment, only conditions of warmth and low wind velocities might trigger it.

5. THE NEED TO CARRY KNOWLEDGE ALONG

To be useful, such a device has to know something, as Lettvin, Maturana, Pitts, and McCulloch concluded when they started to study a frog's eye. What a frog knows in its retina is described above.

To look for what scientists need to learn about the surface of Mars, our device must know what the scientists are looking for. We shall be happy to respect their wishes.

LIST OF REFERENCES

1. Lettvin, J. Y. , Maturana, H. R. , McCulloch, W. S. , and Pitts, W. H. , What the Frog's Eye Tells the Frog's Brain, Proc. of the IRE, p. 1940, November 1959.
2. Maturana, H. R. , Lettvin, J. Y. , McCulloch, W. S. , and Pitts, W. H. , Anatomy and Physiology of Vision in the Frog (Rana Pipiens), Journal of General Physiology, pp. 129-175, July 1960.
3. Lettvin, J. Y. , Maturana, H. R. , Pitts, W. H. , and McCulloch, W. S. , Two Remarks on the Visual System of the Frog, Sensory Communication, W. Rosenblith, ed. , pp. 757-776, 1961.
4. Sutro, L. L. , editor, Advanced Sensor and Control System Studies, 1964 to October 1965, R-519, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, p. 16, January 1966.
5. Moreno-Diaz, R. , An Analytical Model of the Group 2 Ganglion Cell in the Frog's Retina, E-1858, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, October 1965.
6. Ramon y Cajal, S. , Die Retina der Wirbelthiere, Tafel II, Fig. 6 Wiesbaden, Verlag von J. F. Bergmann, 1894.
7. Schypperheyn, J. J. , Contrast Detection in Frog's Retina, Acta Physiol. Pharmacol. Neerlandica 13, pp. 231-277, 1965.
8. Hartline, H. K. , The Response of Single Optic Nerve Fibres of the Vertebrate Eye to Illumination of the Retina, Amer. J. Physiol. , Vol. 121, pp. 400-415, February 1938.
9. Sutro, L. L. , editor, op. cit. , pp. 10-38.
10. Westinghouse Electric Corporation Aerospace Division, Baltimore, Md. , Advanced Molecular Systems Technology, Volume 1 Systems Applications, pp. 4-11 to 4-16, November 1964.
11. Sutro, L. L. , editor, op. cit. , pp. 84-95.

12. Kilmer, W. L., McCulloch, W.S., and Blum, J., Towards a Theory of the Reticular Formation, 1966 IEEE National Convention, Cybernetics Session, March 14, 1966, New York City, New York, as modified for this paper.
13. Scheibel, M. and A., On Neural Mechanisms for Self-Knowledge and Command, Mitre Report SS-3, First Congress on the Information Sciences, Mitre Corporation, Bedford, Massachusetts, 1962.
14. Kilmer, W. L., Summary of Research Progress, Research Laboratory of Electronics Quarterly Progress Report, July 15, 1966, as modified for this paper.
15. Kilmer, W. L., Section in Advanced Sensor and Control System Studies, April to June 1966, R-548, Louis Sutro, Editor, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, (in preparation).

