AN ACCOMMODATING EDGE FOLLOWER

by

Karl K. Pingle and Jay M. Tenenbaum*

Computer Science Department
Stanford University
Stanford, California USA

We describe an operational program that locates objects in a television image and traces their edges. The program accommodates the television camera, maximizing dynamic range during acquisition and sensitivity during tracing, to obtain the most appropriate image for each phase. If the trace routine loses an edge, various heuristics diagnose the difficulty and tune both the camera and software to recover contrast at the point of difficulty. Experimental evidence of the effectiveness of accommodation is provided.

1. INTRODUCTION

At IFIPS '68, the Stanford Artificial Intelligence Project described a system which used a mechanical arm to stack cubes that had been located by an edge tracing program (1). This edge follower could easily find the outlines of white cubes on a black table, but was prone to error in less carefully controlled environments. Our studies of its inadequacies have stimulated the development of a more powerful edge follower, which overcomes most of the limitations of the old one.

This program is currently the initial stage of visual processing in the Stanford hand-eye system (2). It has demonstrated an ability to track weak edges under adverse lighting conditions

2. HARDWARE

The edge follower uses a standard vidicon television camera, modified to provide computer control of orientation (a pan-tilt head), focal length (a lens turret), color filter, focus, and target voltage. The lens iris is set manually. The pan-tilt head, lens turret, and focus motor drive are standard manufacturer options that were augmented with feedback pots and control logic to allow computer servoing. Three color filters and one neutral density filter are mounted on a "filter wheel" located between the lens turret and vidicon. Any filter can be selected in 1/5 second. The camera's "Auto-target" circuit was modified to allow computer selection of 64 discrete target voltages between 0 and 50 volts. From 1/A to 1/2 second (about 10 television frames) are required before an image stabilizes after a change of target voltage. The auto-target originally provided automatic exposure control by servoing the target voltage to maintain a uniform average signal current. The modified circuit will not allow the computer selected target voltage to exceed this internal reference, thus protecting the vidicon from programming errors. The internal voltage can also be used to set the target voltage when the computer has no knowledge about the scene.

The vidicon provides a 333 x 256 array of intensity samples in 1/60 second. The capacity of the computer's high speed data channel (also used by the swapping disk) is 24 million bits/second. Thus, without local buffering each sample can be encoded as a 4 bit number. For flexibility the width of this 16 level quantization window can be adjusted within one milli-second to trade off dynamic range and resolution. For maximum dynamic range, the 16 quantization levels are spread over the 1 volt working range of the video amplifier, while for maximum resolution, they are concentrated into a 1/8 volt sub-window. Intensities in this selected range can thus be resolved to the equivalent of 128 levels (7 bits) over the maximum lv. range.

3. BASIC ALGORITHM

The basic algorithm, used by both the old and new edge follower, begins by scanning in a coarse horizontal raster for an intensity discontinuity. An operator is applied between the points of discontinuity to determine the exact center and direction of the edge. The operator is then moved a short distance along the edge to find the next edge point. If nothing is detected, the operator is scanned back over symmetrical arcs towards the previous edge point until the edge is reacquired. The image coordinates of the new edge point are added to an ordered list representing the object's topology and the operator again moved along the current direction of the edge. The program normally terminates when it returns to the first point it saw. Straight lines, fit through the edge points, provide the line drawings used for scene analysis. A detailed description of the old edge follower and its line fitting algorithm can be found in (3).

4. LIMITATIONS OF PREVIOUS SYSTEM

The old edge follower worked well only with highly contrasting objects, viewed under strong

*This research was supported by the Advanced research Projects Agency of the Department of Defense under Contract No. SD-183. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency of the U.S. Government.

**Now with the Lockheed Palo Alto Research Laboratory.

uniform illumination. Its major limitation was that it processed only one image, obtained by manually tuning the camera to maximize average contrast over the field of view. Unfortunately, with 16 levels of quantization, it is seldom possible to combine adequate intensity resolution for low contrast edges with enough dynamic range to see all edges. Thus, in the computer's view of a typical hand-eye scene (figure 1), the interior detail of the left-hand cube (labeled B) is cut off by inadequate dynamic range while the rear exterior edge of the wedge (labeled A) is not resolved by the available quantization density.

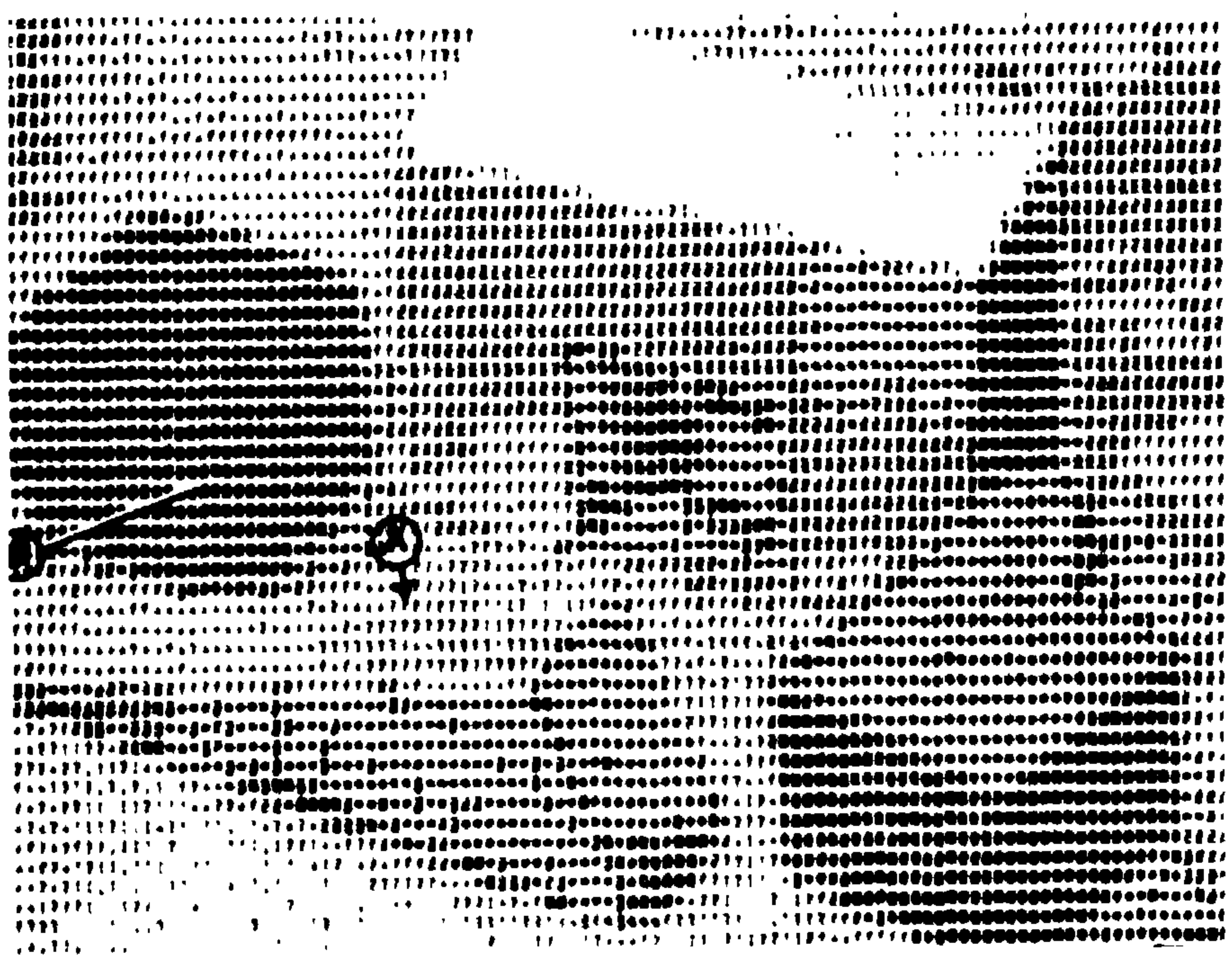

Figure 1 Typical Hand-eye Workspace
(Three levels of overprinting were used to obtain simulated grey tones with conventional line printer characters. About 8 of the actual 16 levels are visible.)

In the past, we overcame these limitations by tuning the camera and lighting to emphasize those objects actually involved in a task. Figure 2, for example, shows the effects of intentionally narrowing the quantization window to enhance the contour of the wedge (lower left corner) at the expense of other features in the scene.

There are severe drawbacks to this approach. First, an autonomous robot should decide for itself when an image is inadequate and how it can be improved. Many crucial aspects of perception are masked by using an image which has already been filtered by human recognition to emphasize selected objects. Secondly, the effectiveness of this method is restricted to tasks in which all objects are of similar brightness or of sufficiently high contrast that all can be seen in a single image. The difficulty of seeing both the exterior and interior of an object further limits applicability to scenes simple enough to be interpreted solely on the basis of object outlines.

Figure 2 Accommodation for Contour of Wedge

The old edge follower and its data structure were only designed to handle simple closed contours. Its one level of processing could not recover from the simplest of failures, such as a trace that looped around a small shadow or sharp corner. If a closed outline was not found, or if line fitting did not produce a reasonable number of corners, the object was purged from the data structure and the global scan resumed. Occasionally, such an object would be reacquired at a different point on its contour and traced successfully.

The new edge follower has a better operator, a more general data structure and heuristics to help it recover from trouble. The largest single improvement in performance, however, comes from the edge follower's ability to 'tune' the camera for optimum sensitivity at a particular point in the image where an edge is expected but not seen.

5. CAMERA ACCOMMODATION

Figures 1 and 2 graphically demonstrate the importance of camera adjustments for successful edge following; what is seen depends strongly on how the parameters of the sensory channel are tuned. Unless adequate edge information is available, sophisticated processing cannot help.

We have already described the critical limitation on intensity resolution imposed by channel bandwidth. Similar conflicts are found in each dimension of the camera's response. For example, the combination of spatial resolution and field of view available in any image, which is strictly constrained by the number of available raster samples, will provide coarse coverage of a wide field of view or high resolution in a small area.

Since no single image can contain all of the information available in a scene, the sensory channel must be tuned to concentrate all available resolution on that information currently required by a particular perceptual task. This requires that the parameters of the visual sensor be controlled, not by a person, but directly by the perceptual process (6).

We name this concept "Accommodation", a term borrowed from physiology where it specifically denotes the role of the lens in focusing the human eye. In man, there are may examples of accommodation in this wider sense, for example, the use of neural adaption and an adjustable iris to achieve vision over an extraordinary range of brightness. Since the performance of present artificial sensors falls considerably below the standard established by nature, the need to accommodate them is correspondingly more acute.

The accommodation requirements of edge acquisition are quite different from those of tracing; during the coarse raster scan, the computer usually ha6 no preferences about where to look, and no expectations about what it will find. To facilitate a thorough search, the quantization window is opened as wide as possible, sacrificing resolution to obtain wide dynamic range. On the other hand, once an edge is acquired, the tracing program knows precisely what it is looking for and where to find it; if the edge is lost, the camera should be tuned to emphasize intensity discontinuities in the immediate vicinity of where it was last seen.

The appropriate accommodations for this tracing mode were deduced by examining the effects of each accommodation parameter on contrast and signal/noise, the primary requisities for successful edge detection.

Target voltage

The effect of target voltage (ET) on the transfer characteristics of a vidicon camera is expressed by Equation 1.

$$VS=K*L^{g}*(ET)^{\alpha}, \quad g=.65, \quad \alpha=1.42 \quad [1]$$

VS is the output signal voltage corresponding to a light flux of L foot candles incident on the face plate of a vidicon. This surprisingly obscure equation was inferred from more commonly available vidicon specifications (6). The values of g and α were obtained by fitting Equation 1 to data observed with our RCA-8507A vidicon. ET bounds the range of illumination that will yield signals within the one volt window of the quantizer. It also determines the minimum edge contrast, ∂L, that can be resolved by a given quantization density.

The principal noise in normal operation is statistical fluctation in the bias current of the preamplifier's input stage. Since this noise is independent of target voltage and light level, the signal/noise of weak edges is maximized by raising target voltage until the signal, corresponding to the brighter side of the edge, is just inside the one volt limit of the quantization window. This target voltage is known as E_{sat}

The target voltage may be limited, in practice, by the possibility of damaging the vidicon. The autotarget circuit provides a maximum safe target voltage (E_{dam}) which is also an inverse function of light level. Although E_{sat} is always less than E_{dam} in uniform illumination, E_{dam}, imposed because of a scene highlight, may be lower than the E_{sat} for a lowlieht in which we are interested. An empirical rule for maximizing detection confidence is thus to set target voltage to the minimum of E_{sat} and E_{dam}.

Quantization Window

Edge following requires a quantization window that is narrow enough to resolve weak intensity discontinuities yet wide enough to encompass the range of intensities found over the length of a contour. To satisfy these incompatible demands, the quantization window is adjusted to bracket the range of intensities currently covered by the operator. This optimizes contrast where an edge is expected.

Focus

Defocus increases the transition width of an edge, reducing the gradient (i.e., the difference of intensities at a fixed separation) calculated at the transition. Focus is suspect when, in spite of adequate contrast across a suspected boundary, there is no distinct gradient peak. When focusing is required, the vidicon is moved relative to a fixed lens, until a focus criterion is maximized. Since focusing is intended to increase the local gradient, this function is the natural criterion by which to measure image sharpness (6), (7).

Focal length

The wider the field of view, the less need for re-orienting the camera to track a long edge. Since spatial resolution usually is not a limiting factor in detecting step-type edges, a short (1") lens is commonly used to obtain a broad view. (The computer can switch the turret to a longer lens if one is needed to see textural detail.)

Iris

Shc_ noise in the vidicon's dark current, usually a second order disturbance, increases faster than the signal with target voltage and can be significant when a weak edge is sought in a dark region. A wide iris minimizes this noise by reducing the target voltage required to obtain a full signal from the available light.

A wide iris also decreases depth of field, which could increase the amount of re-focusing required along an edge. Fortunately, since the depth of field of a f/1.4 one inch lens includes the camera's entire field of view when looking obliquely down from above the hand-eye table, the iris can be left wide open. Focusing is necessary only when camera orientation changes.

Color filters

Color filters can restore the contrast between differently colored surfaces, that is often lost in a grey scale image. Since the colors are usually unknown during edge following, the filters must be randomly tried to find the best one. To avoid this expense, the edge follower uses only the clear filter which maximizes available light.

Effectiveness of accommodation

The edge follower was instructed to look for edges in the interior of the black cube at the left of Figure 1. Figure 3 was the image it used while tracing the diagonal interior edges between the top and front faces of this cube. This image shows the practical effects of raising target voltage and narrowing the quantization window to enhance the contrast in a specific local region, namely the immediate vicinity of the left diagonal interior edge.

We learned that contrast is optimized when the bright side of an edge is almost saturated. That the top of the cube is not lighter is therefore an indication that the target voltage was limited by E_{dam} to prevent damage to the vidicon by the brighter background. Because of this constraint, the vertical interior edge, which (without benefit of direct overhead illumination on one side) is both darker and of lower contrast than the diagonals, was not detected by the edge follower.

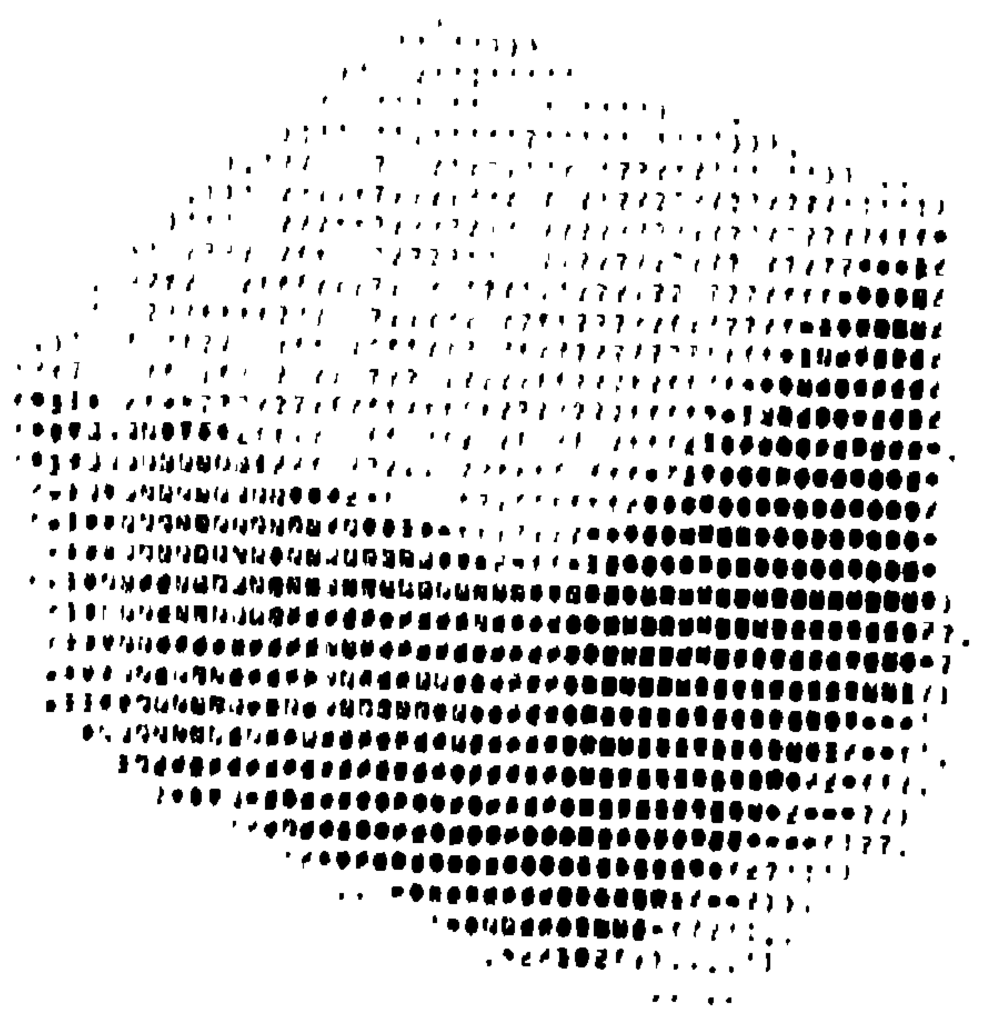

Figure 3 Accommodation for Diagonal Interior Edges

Although focus, focal length, iris, and color filters play passive roles in edge following, they are actively involved in the accommodation strategies of other perceptual skills utilized by our hand-eye system (6). One such program is the edge verifier which utilizes all of these accommodations in an exhaustive effort to find a specific edge. When requested to find the missing vertical edge, this program eliminated the limiting highlight by using a longer lens to restrict the field of view to the vicinity of the edge. The target voltage was then raised to obtain sufficient contrast (figure 4). Note that the top of the cube has become fully saturated and now limits target voltage.

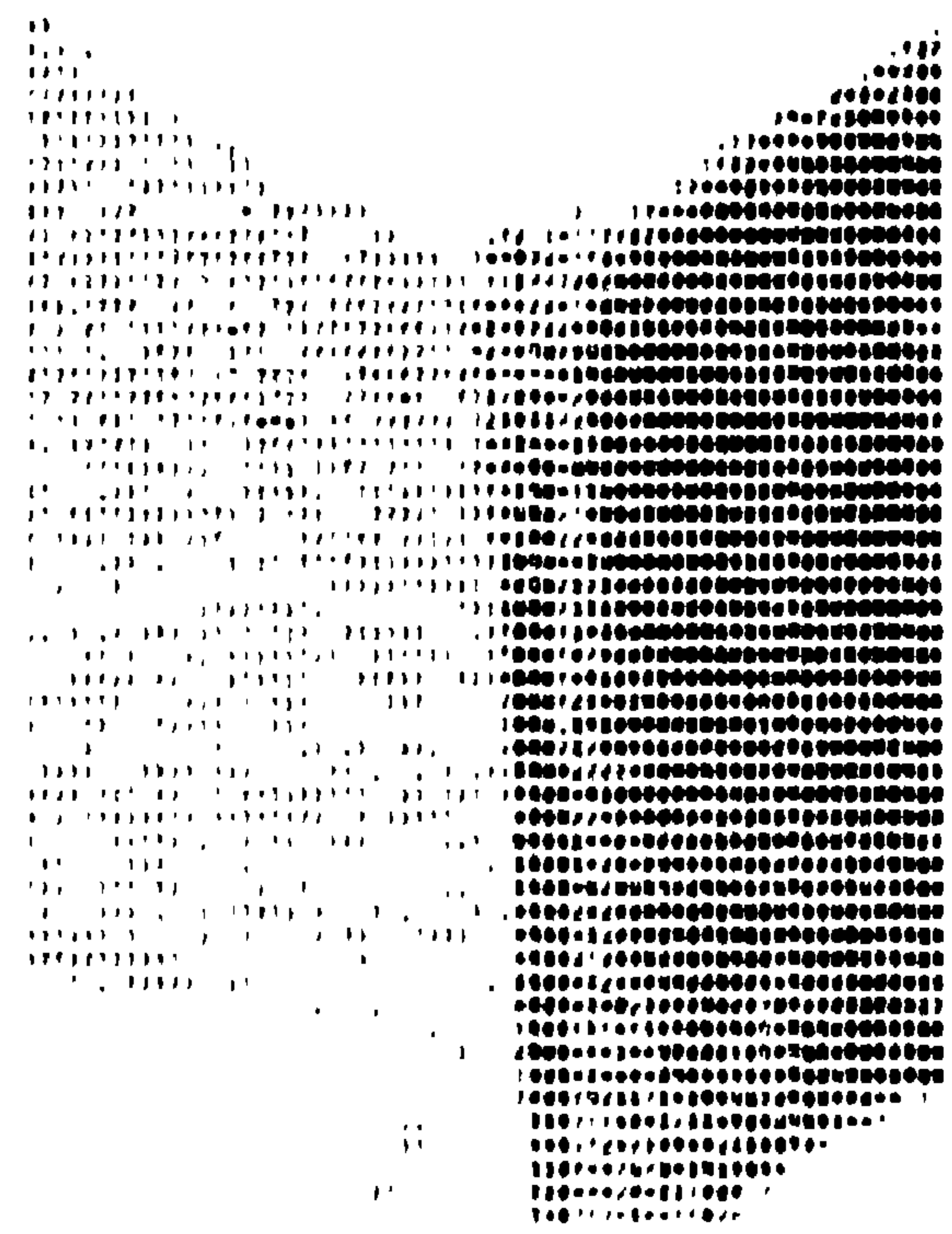

Figure 4 Vertical Interior Edge Viewed with 3" Lens

6. THE NEW EDGE FOLLOWER

The new edge follower begins by accommodating for the coarse initial scan. A wide quantization window and short lens provide broad coverage of the scene. The resulting low spatial and intensity resolution makes the system partial to strong edges and helps minimize the chance of detecting irrelevant fine detail, such as specks of dust.

A goal oriented strategy program, interested in a specific area of the table or in a specifically colored object, could bias the search with a more selective accommodation window. A longer lens or color filter would restrict the class of intensity discontinuities that could be detected during the coarse scan.

The accommodation routine monitors the sampled intensities, adjusting the target voltage whenever necessary to keep the values within the dynamic range of the camera and quantizer. If no edges are found, the scan is repeated with the quantization window and target voltage adjusted for maximum intensity resolution at each sample point.

After acquisition, but before beginning the trace, it is prudent to verify that the discontinuity was not the result of random noise or an isolated surface anomaly. An effective requirement is that the discontinuity have a sufficient spatial extent; i.e. that a significant difference also be observed in the average of intensities obtained from small neighborhoods about the original detection points. Other requirements, such as a desired color, can also be checked at this time.

To prepare for tracing, the accommodation is changed to maximize intensity resolution at the point of acquisition. This involves peaking the target voltage, narrowing the quantization window, and checking focus, as discussed earlier.

The normal trace progresses by applying an edge operator at a point extrapolated along the edge. The edge follower uses a powerful regional operator (4) which makes a global judgement of the best edge in a circular neighborhood (selected to encompass between 32 and 177 raster points).

This operator has several advantages over the 3 x 3 gradient operator used in the old system. It can detect weak edges (or those with ill defined gradients) in the presence of substantial noise, allowing it to take full advantage of the sensitivity gained with accommodation. It also determines the center of an edge with more accuracy, resulting in slightly better line fitting. Since the operator responds to edges over a sizeable area, scanning is not needed as often to recover the edge when it changes direction.

The camera accommodation need not be changed during the trace, unless the edge is lost. The edge follower then repeats the accommodation sequence it evoked at the start of the. trace to optimize accommodation at the operator's present position. The failure of accommodation is thus the ultimate criterion used to reject a suspected edge point; the success of the operator, on the other hand, is the criterion by which and for which accommodation is optimized.

If accommodation fails, the program resorts to the recovery heuristics described later. (One might also consider trying more costly accommodations, such as changing color filters.)

7. THE INTERNAL DATA STRUCTURE

As it traces objects, the edge follower builds a data structure to describe their topology. The data structure is generated by a list processing language (5) that resembles CORAL (8) and is created by assembly language macro statements which set up calls to list processing subroutines. The basic language unit is a block of contiguous core containing status bits, data, and pointers to other blocks. The parameters for each block type are given to the processor, along with a symbolic name, at the start of the program. Copies of these blocks are then generated as needed.

For each object, an OBJECT block is created and linked to a circular list, or ring, of SEGMENT blocks representing the edge segments found for that object. Each SEGMENT block points at a ring of POINT blocks, containing the coordinates of the individual edge points. The SEGMENT block of any segment that is not a closed outline is marked. The POINT blocks at each end of the segment are also marked to indicate the cause of the termination; lost edge, out of the camera's field of view, or intersected the middle of another edge.

The edge follower must be able to rapidly determine whether an edge point has been seen before and, if so, which object it belongs to. The world model divides the scene into a square grid. Associated with each square is a pointer to a ring of WORLD blocks, one for each object with an edge passing through the square. The WORLD block points at a ring linking all the POINT blocks with coordinates in the area.

When a trace intersects a previously seen edge point, the data structure Is consulted. If the edge point is the opposite end of the current segment, a closed outline has been traced. If the current segment intersects a previously traced one at an endpoint, the two are merged. Otherwise, the current segment is added to the ring linking the other segment with an object. Thus, all edges need not be traced at the same time, but can be added to the data structure of the appropriate object as found. Whenever a new segment is added to or merged with an object which has already had lines fit to it, the operation is repeated to obtain a better, more complete fit.

When the high level programs no longer require detailed edge data, they can instruct the edge follower to 'compact' the data structure by deleting the POINT blocks and associated SEGMENT blocks for subsequent garbage collection. These are replaced by a ring of CORNER blocks containing the coordinates of each corner of the associated object and the equations of the edges connecting them. The WORLD blocks are flagged and linked to these CORNER blocks. Thereafter, when the object is encountered while scanning, the scan line is extended until it intersects the furthest edge of the object and the scan continues from that point.

8. HEURISTICS

The old edge follower would often fail to complete a contour because small sections were slightly noisy or of marginal contrast. The heuristics described in this section combat the most common of these failure modes. The new edge follower attempts to diagnose the nature of its current difficulty, and recover intelligently.

A patch of noise or shadow can cause a trace to loop back in the direction from which it came. The program maintains a stack containing the coordinates of the last few points found. If one of these is encountered again, the program tries to resume its original direction, using a larger increment to jump over the bad area. If the edge has not turned a corner, it is often re-acquired on the far side.

If the loop persists, the trace returns to the point of acquisition and tries to close the contour in the opposite direction. If it gets close enough to the failure point the outline is considered closed. Otherwise, high level programs can try to extend the ends until they meet These heuristics are also used when edges are simply lost and when they intersect other edges or boundaries of the image.

When tracing is finished, a fine horizontal and vertical raster scan is performed with careful accommodation to find additional edges near unconnected ends of a segment. The tracing routine is called to pursue any that are found.

Occasionally an object is traced in such a way that although the entire outline is found, the ends of the lines fit to the edge points do not meet (figure 5). A post-processor tries to correct such errors to obtain closed outlines for simpler scene analysis. If the corner closest to each unconnected endpoint is close enough, the open line is extended toward the corner. If this extension intersects an extension of a line belonging to the corner sufficiently close to the corner, the unconnected end is attached to the corner in the data structure.

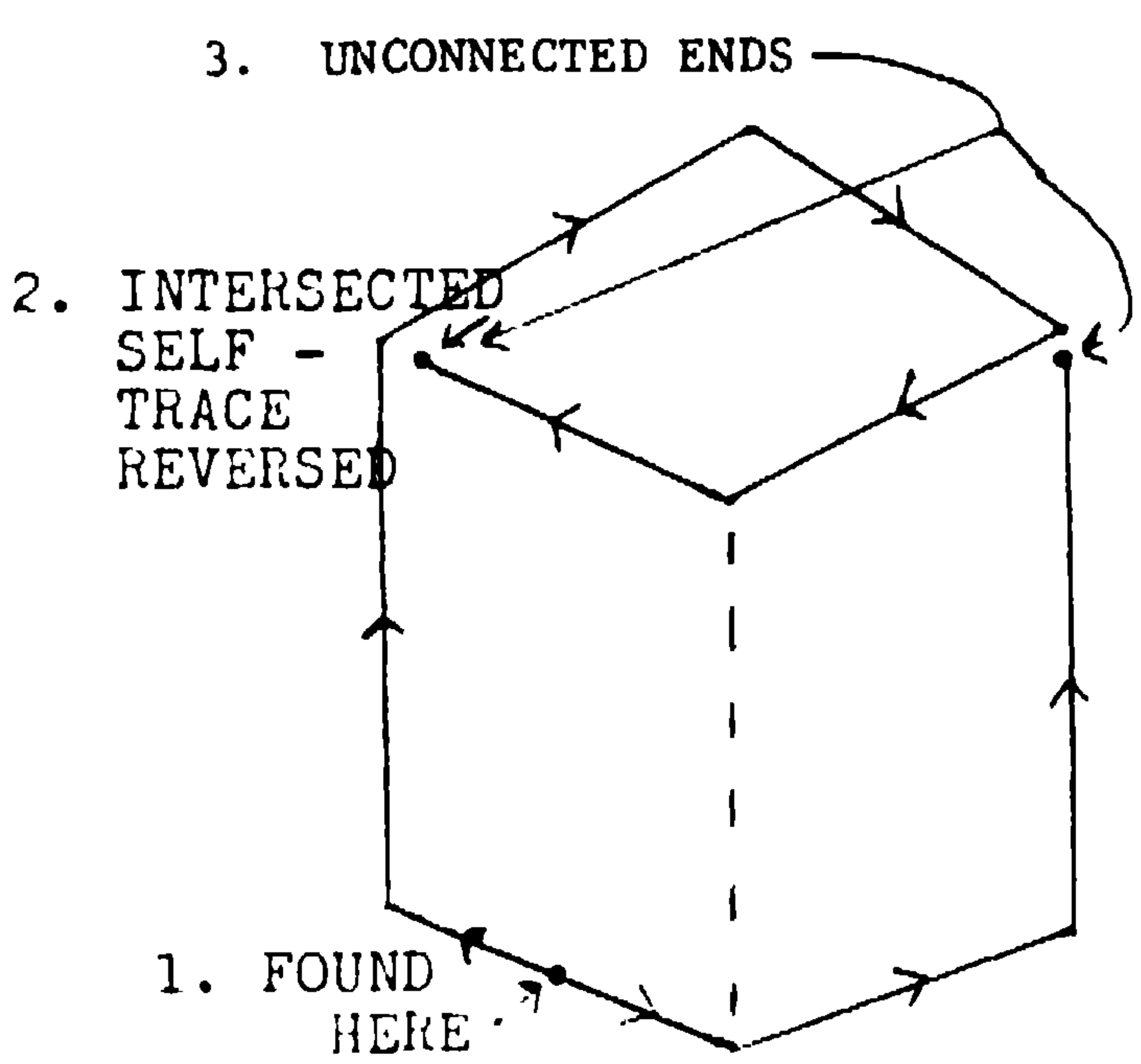

Figure 5 A Complete but Unclosed Outline

To determine whether a set of lines constitutes a closed outline, the program picks up the endpoints of a line and tries to find one other line that meets each endpoint. If both are not found, the chosen line is deleted. This test is then repeated on all remaining lines. When a complete pass through the set of lines produces no new deletions, the process terminates. Any remaining lines form a closed outline which is indicated by links added to the data structure.

Once an outline is found, high level strategy determines whether to look for other near-by edges which may have missed. The edge follower can be instructed to look carefully for new edges in a specified area.

9. DISCUSSION

The performance of an edge follower can be improved by applying more sophisticated processing to a given image or by accommodating to obtain a more appropriate image. Most researchers have followed the former course, relying on sophisticated processing to cope with inadequate images.

These efforts have ranged from digital spatial filtering techniques for image enhancement (9) to syntax directed analyses which utilize high level context to infer missing detail (10, 11). Of particular relevance is the work of Griffith (12) and Binford (13) at MIT who have developed pre-processors which organize an array of locally probable intensity discontinuities into a meaningful line drawing, and of Fennema and Brice (14) at SRI, who have abandoned edges in favor of regions, because missed edges frequently prevented them from finding properly closed regions. The problem that almost all of these workers attacked, was to analyze a 'given' image. This image was usually made available to the program only as a fixed array of intensities that had been previously digitized. (Some programs did have direct access to a camera, but since the parameters weren't controlled by the computer, this was equivalent to using a fixed image).

We feel that because of the inherent limitations of a single image, the acquisition of information should be treated as an integral part of the perceptual process. Thus, the edge follower optimizes accommodation when it does not see an edge where one was expected. In effect, many images are used, each best for a particular part of the scene.

By using the loss of an edge as a pragmatic indication to accommodate, the computer is utilizing what it already knows to see more. This strategy is economical because accommodation is optimized only at a few problem points. The machine time involved is, in many cases, considerably less than would be necessary to achieve comparable performance with "smarter" tracing algorithms. Furthermore, the expectation provided by edge continuity significantly reduces the attendant risk of responding to isolated noise.

These considerations are equally applicable to sensors like image dissectors and photo multiplier (flying spot) scanners, which are not dynamic range limited. Since these devices measure brightness by the time taken to observe a fixed number of photons, intensity resolution can be traded off against measurement time by changing the required count. For efficiency, this count should be kept low, except when resolution is needed to recover a lost edge. A fixed count is likely to result in too much or too little resolution at any particular point.

The accommodating edge follower has demonstrated that it can detect low contrast edges over a wide range of brightnesses, routinely finding most of the interior and exterior edges in typical hand-eye scenes (e.g. figure 1). However, it is misleading to quantitatively assess the worth of accommodation as one might a new edge operator, in terms of a numerical value like the signal/noise of the weakest edge that can be detected. By this criterion, the new edge follower does not perform significantly better than the old one did when its accommodation happened to be best for the particular edge. The worth of accommodation should instead be judged in the qualitative sense that performance is no longer dependent on manual intervention and on the environmental constrainst previously required to balance dynamic range and resolution in a single image; in figures 3 and 4 it is easy to find interior edges that are impossible to detect in figure 1. Accommodation attacks the fundamental limitation of image inadequacy rather than the secondary problems caused by it.

We view the significance of accommodation less as a means of overcoming sensor limitations than as a way to match the capabilities of a sensor with the information requirements of a specific task. The success of this edge follower can be attributed to a careful evaluation of how the characteristics of a vidicon sensor could be used most effectively in this application. Similar success has been found applying this accommodation philosophy to perceptual functions as diverse as range finding and color recognition (6).

REFERENCES

1. K. K. Pingle, J. A. Singer, and W. W. Wichman, "Computer Control of a Mechanical Arm Through Visual Input", Proceedings IFIPS '68, pp. 1563-1569, North Holland, 1969.

2. J. M. Tenenbaum et al, "A Laboratory for Hand-Eye Research", Proceedings, IFIPS '71.

3. K. K. Pingle, "Visual Perception by a Computer", Automatic Interpretation and Classification of Images, pp. 277-284, Academic Press, 1969.

4. M. Hueckel, "An Operator Which Locates Edges in Digitized Pictures", JACM Vol. 18, No. 1, Jan. 1971, pp. 113-125.

5. K. K. Pingle, "A List Processing Language for Picture Processing, Stanford Artificial Intelligence Project Operating Note, 1968.

6. J. M. Tenenbaum, "Accommodation in Computer Vision", Stanford Artificial Intelligence Project Memo, AIM-134, 1970

7. B. Horn, "Focusing", MIT Project MAC, Artificial Intelligence Memo No. 160, May, 1968.

8. W. R. Sutherland, "On-line Graphical Specification of Computer Procedures", Tech. Report 405, Lincoln Laboratory, MIT, Lexington Massachusetts, May, 1966.

9. A. Rosenfeld, "Picture Processing by Computer", Computing Surveys, Vol. 1, No. 3, Sept. 1969, pp. 147-176.

10. W. Miller and A. Shaw, "Linguistic Methods In Picture Processing-A Survey", Proc. FJCC, pp. 279-290, Dec. 1968.

11. M. Fischler, "Machine Perception and Description of Pictorial Data", Proc. 1st International Joint Conference on Artificial Intelligence, Washington, 1969.

12. A. K. Griffith, "Computer Recognition of Prismatic Solids", Ph.D. Thesis, Dept. of Mathematics, MIT, June, 1970.

13. T. 0. Binford, "A Visual Preprocessor", MIT Project MAC, Artificial Intelligence Memo, 1970.

14. C. Brice and C. Fennema, "Scene Analysis Using Regions", Artificial Intelligence, Vol. 1, No. 3, Fall, 1970.