534749 P30
N90-16217

ISOLATING CONTOUR INFORMATION FROM ARBITRARY IMAGES

Daniel J. Jobson
NASA Langley Research Center
Hampton, Virginia

Abstract

Aspects of natural vision (physiological and perceptual) serve as a basis for attempting the development of a general processing scheme for contour extraction. Contour information is assumed to be central to visual recognition skills. While the scheme must be regarded as highly preliminary, initial results do compare favorably with the visual perception of structure. The scheme pays special attention to the construction of a smallest scale circular difference-of-Gaussian (DOG) convolution, calibration of multiscale edge detection thresholds with the visual perception of grayscale boundaries, and contour/texture discrimination methods derived from fundamental assumptions of connectivity and the characteristics of printed text. Contour information is required to fall between a minimum connectivity limit and maximum regional spatial density limit at each scale. Results support the idea that contour information, in images possessing good image quality, is contained largely if not wholly in the highest two spatial frequency channels (centered at about 10 cyc/deg and 30 cyc/deg). Further, lower spatial frequency channels appear to play a major role only in contour extraction from images with serious global image defects.

Introduction

The goal of sophisticated machine vision capabilities requires that attention be paid to the semantic content of images together with intrinsic image characteristics such as contrast, noise, blur, and sampling. The visual task of recognition and interpretation is of paramount interest. Here the first step toward a more general approach to machine vision recognition is defined as a set of methods for transforming arbitrary images into contour or schematic line drawing information. This set of methods was fashioned from natural vision concepts (physiological and perceptual) coupled with image acquisition processes. Emphasis is placed on the smallest scales of the image. An abbreviated pyramid of circular difference-of-Gaussian (DOG) convolution operators forms the first stage of image processing. In particular the nontrivial problem of constructing a smallest scale DOG operator from discrete image samples is of special interest. Edge detection and contour extraction processing is then performed for each scale of operator. Multiscale contour information is then merged in an hierarchical manner with priority being given to the information from the smallest scale. Larger scale information is only added to spaces not already occupied by smaller scale information.

The set of methods can be viewed as a progression starting with the image and proceeding through "seeing" to significance. The processing scheme, which must be regarded as preliminary, is described together with the ideas behind the scheme. Results are given for a series of image processing experiments designed to provide a partial demonstration of the overall consistency of the scheme with the visual perception of structure in images. While conciseness of either the processing or the resulting information was not a major goal, it is noted that processing is reasonably simple and the contour information is intrinsically concise (and can be made more compact by the addition of coding schemes).

Construction and Resiliency of Smallest Scale Operator

Previous work by Huck et al. (Ref. 1) demonstrated that a well-behaved smallest scale DOG operator could be constructed for one case of a specific amount of image blur and a particular spacing of the square grid sampling lattice. Subsequently this work was extended to show that this well-behaved operator will result from the same set of weighting coefficients even for significant changes in blur or sampling lattice spacing (Ref. 2, Figs. 1 and 2). The two-dimensional convolution of these weighting coefficients with the image samples is then equivalent to applying a small DOG operator to the original scene radiance distributions. Less attention was paid to the construction of larger operators (in this case about 3 times and 6 times larger) and the larger functions used were merely formed from discrete values of the desired size of DOG function. The spatial spread of image edges after convolution was checked as a rough verification that the desired scale operator was achieved.

Multiscale Two-Dimensional Edge Detection and Representation

A fully two-dimensional edge detection method was found to be necessary (Ref. 2). Likewise an edge representation space magnified by a factor of two

over the original image space was also a requirement (Figs. 3, 4, and 5). Edge detection is based on zero-crossings (Ref. 3); however, the determination of thresholds for "zero" for edge detection could not be determined from fundamental considerations. In particular, noise-limited edge detection produced edge representations with a wealth of textural detail which seemed inconsistent with the subsequent goal of contour extraction. Therefore, the performance of visual perception was examined for guidance in determining multiscale contrast sensitivities. To this end, the perception of grayscales edges and bar patterns was examined.

The Disparity Between Grayscale and Bar Pattern Perception at Scales Larger than the Visual Acuity Limit

The perception of both grayscale and bar patterns (Fig. 6) seemed suitable to the calibration of contrast sensitivity versus image scale. One aspect of grayscale edge perception is noteworthy. For equal step intervals in the grayscale and decreasing angular size, a point is reached where almost all edges vanish at once. The exceptions are the lowest and highest steps which vanish at slightly larger angular sizes. Therefore for a particular angular size, edge detection in grayscales seems to be an almost linear process with constant threshold value.

While a consistent result for grayscale edges and bars was expected, actual results were quite different. A striking disparity occurred between the perception of grayscale edge and bar patterns at 3x and 6x the visual acuity limit (Fig. 7). This led to the use of the grayscale sensitivities for edge detection and the formation of a hypothesis that contour information exists as a higher contrast subset of information within the full range of visual phenomena (Fig. 8). It should be noted in these spatial frequency diagrams that higher contrast at a given scale is a necessary but not sufficient requirement for visual phenomena to be contour information. That is, some higher contrast phenomena may still prove to be textural or otherwise not relate to overall contour description of a scene. Contrast sensitivity versus scale must now be related to edge detection zero-crossing thresholds by considering noise, blur, edge contrast, and most importantly sampling effects.

Edge Detection Threshold-Calibration to Contrast Perception Considering Sampling, Noise and Blur

Sampled edge convolution signals exhibit considerable chatter compared to the characteristic analog signal (Fig. 3). Therefore, capture of extended edges in a test image at each scales' contrast sensitivity was calibrated for intrinsic sampling errors coupled with reasonable values of noise and blur. The existence of reasonably low noise (S/N $>$ 50) and modest blurring (Gaussian σ = 0.6 of the sampling lattice spacing) was checked. This calibration (Table 1) was performed in two stages. A set of convolution samples on extended edges at threshold contrast was used to make an estimate of edge detection threshold. Since this relatively small sample might not be highly accurate, some image processing experiments were performed. Edge detection thresholds were adjusted until the bulk of extended straight edges at minimum contrast were detected.

Contour Extraction Methods (Semantic Reference Points)

To this point only the "seeing" of edges has been considered. The isolation of contour information now necessitates a step beyond this into the semantic content of contour information. The question is namely - can we find a fundamental characteristic of contour information on which further processing can now be based? Contour information subjectively seems to always steer between "too little" and "too much", so we can look for ways to establish quantitative criteria for these subjective limits. Connectivity was investigated as a "too little" criterion while regional spatial density of events was used for a "too much" criterion. The following two approaches to minimum connectivity were investigated: 1) the minimum feature of printed text - the period, or more fundamentally 2) connectivity across the space occupied by the original DOG operator. The latter proved to be the most perceptually consistent. Printed text characteristics were also investigated to establish a region size and a maximum number of spatial events allowed. This hypothesis arose from the idea that printed text is engineered by man for possibly maximum information throughput. The resulting quantitative criteria are summarized in Table 2.

Results, Discussion, and Conclusions

As a partial demonstration of generality, the computational scheme is applied identically to diverse images. The original image is shown at the correct size to place each image pixel at about the visual acuity limit for a normal reading distance (Fig. 9). However, perceptual comparisons with these reproduced images are not particularly accurate because the contrast rendition of the original image cannot be maintained in publication. Only results for the 1x and 3x combined scales are shown since this appears to be sufficient for good quality images. Addition of 6x scale information appears to be unnecessary in this case and comes into play for images with global defects (weak contrast, severe blur or noise). The handling of defective images is the subject of an on-going investigation and seems to require a graceful shift to a pair of larger scales as one or more smaller scales produce insufficient information in some global sense.

In an overall sense, these results support the idea that a general scheme for contour extraction is possible and can be based mostly on a pairwise selection of two scales of edge detection and representation. These two scales should be the smallest two for most normal imagery and shift to pairs of successively larger scales only when globally defective images occur.

References

1. Huck, F. O.; Fales, C. L.; Halyo, N.; Samms, R. W.; and Stacy, K.: Image Gathering and Processing: Information and Fidelity. J. Opt. Soc. America, Vol. 2, No. 10, October 1985, pp. 1644-1666.

2. Jobson, D. J.: Spatial Vision Processes: From the Optical Image to the Symbolic Structures of Contour Information, NASA TP2838, November 1988.

3. Hildreth, E. C.: The Detection of Intensity Changes by Computer and Biological Vision Systems. Comput. Vis., Graph., and Image Process., Vol. 22, No. 1, April 1983, pp. 1-27.

TABLE 1. - EDGE DETECTION THRESHOLDS FOR THE THREE SMALLEST IMAGE SCALES (1X, 3X, AND 6X VISUAL ACUITY LIMIT)

SCALE	DESIRED CONTRAST THRESHOLD	ESTIMATED EDGE DETECTION THRESHOLD	ACTUAL EDGE DETECTION THRESHOLD
1X	50%	19%	16%
3X	15%	7.0%	5.5%
6X	5.5%	1.2%	1.2%

TABLE 2. - CONTOUR PROCESSING CRITERIA (IN MAGNIFIED EDGE REPRESENTATION SPACE)

SCALE	CONNECTIVITY	NUMBER OF EVENTS	REGION SIZE
1X	6	75	25 x 25
3X	18	260	50 x 50
6X	36	350	75 x 75

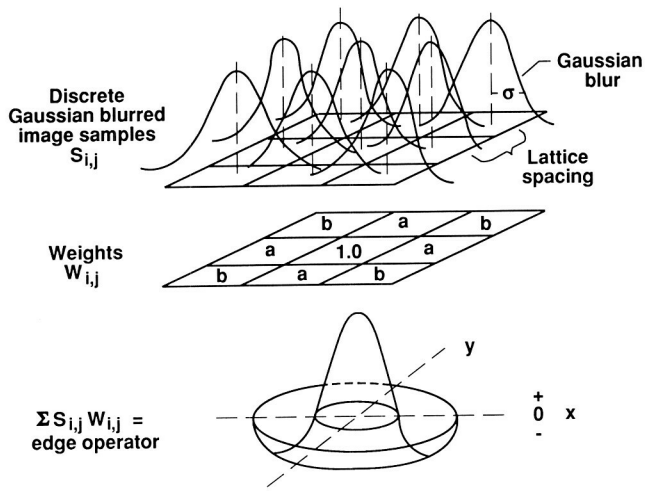


Figure 1. Construction of Smallest Scale Edge Operator for Square Lattice Image Space

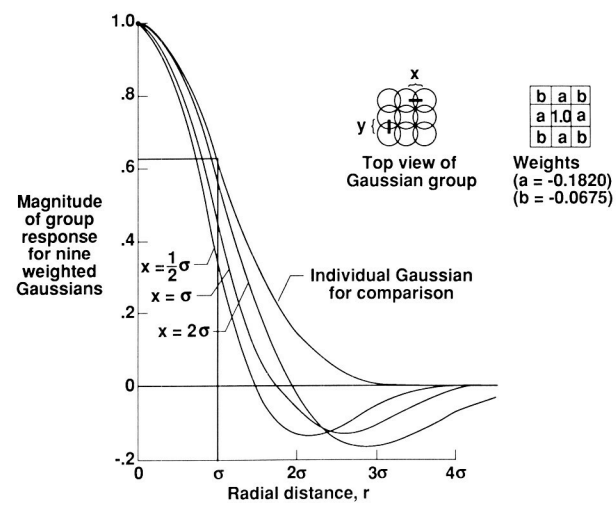


Figure 2. Resiliency of Operator to Variable Sampling Lattice Spacing (or Equivalently Variable Blur)

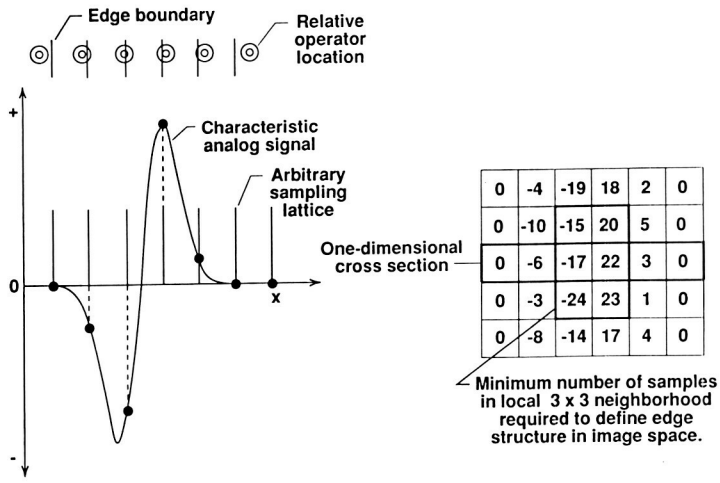


Figure 3. Discrete Samples of Convolution of Edge with DOG Operator

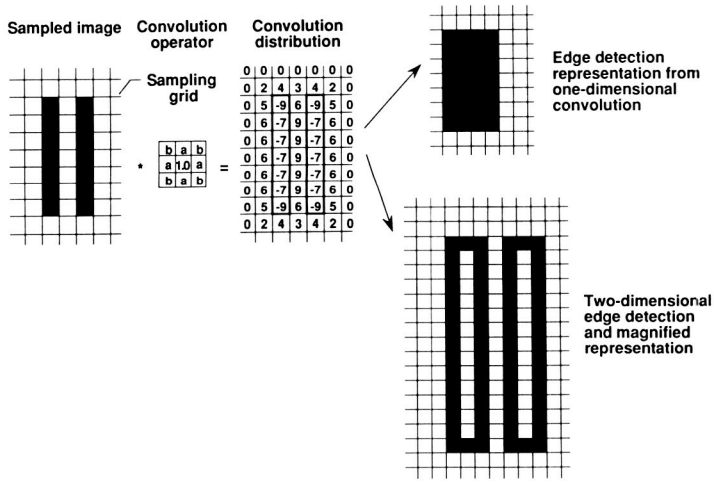


Figure 4. Illustration of the Requirement for a Magnified Edge Representation Space

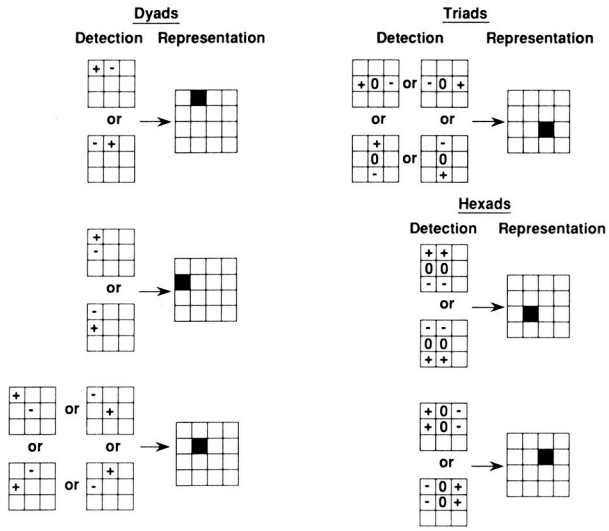


Figure 5. Zero-Crossing Comparisons Used in Edge Detection

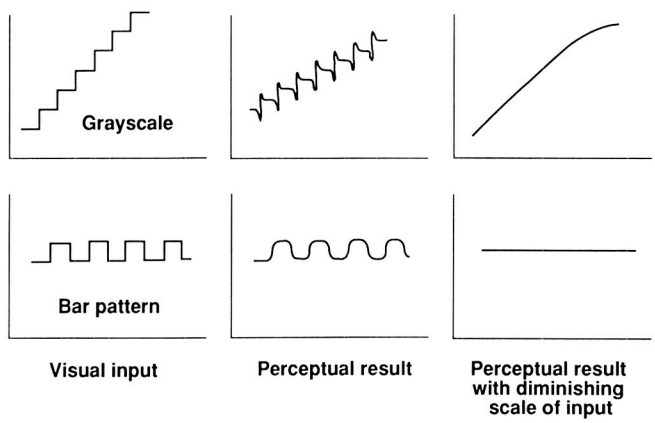


Figure 6. Perceptual Determination of Contrast Sensitivity Versus Image Scale

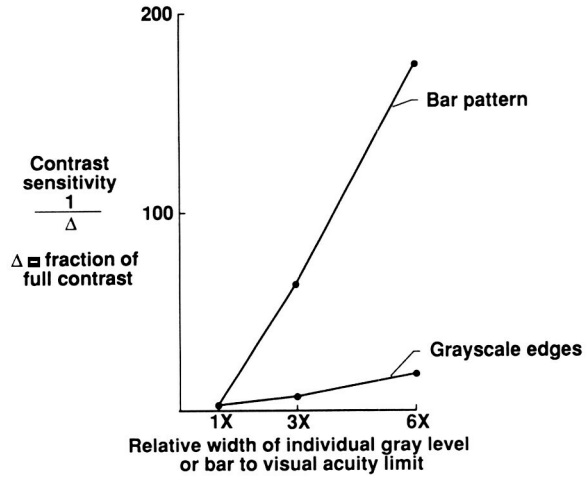


Figure 7. Disparity in Grayscale Edge and Bar Pattern Contrast Sensitivities for Scales Above Visual Acuity Limit

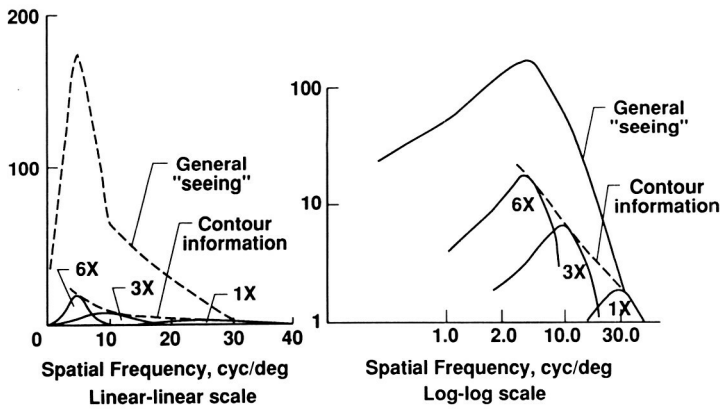
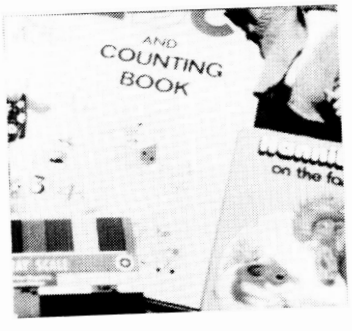


Figure 8. Hypothesis Regarding Contour Information

Figure 9. Image Processing Results

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

C-3



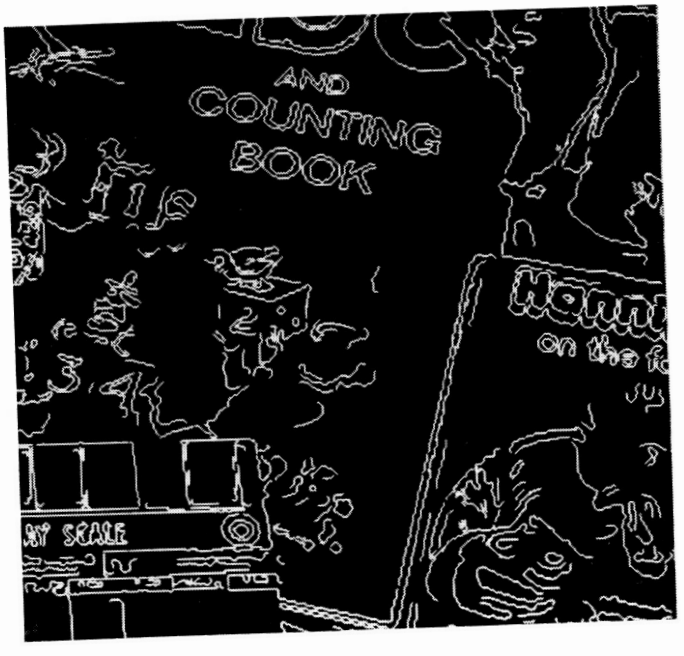
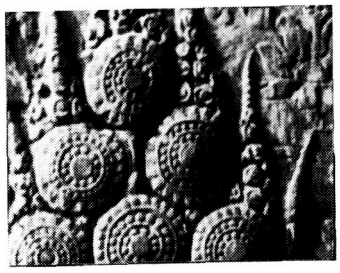
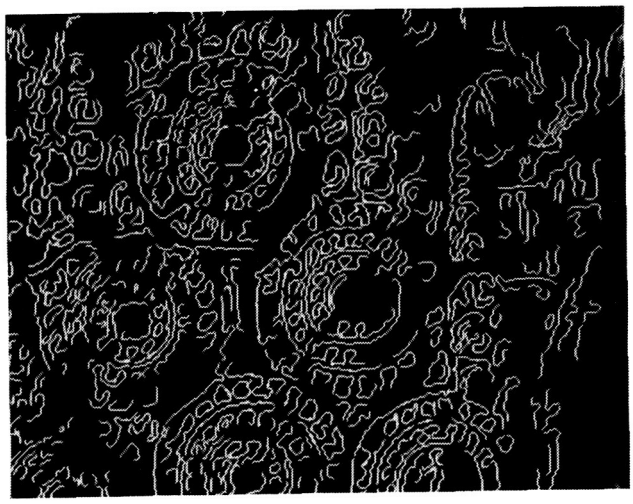
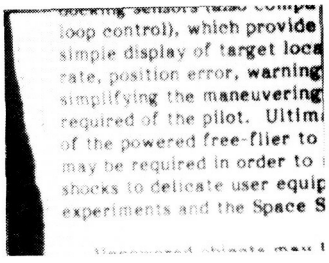


Figure 9. Continued

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH







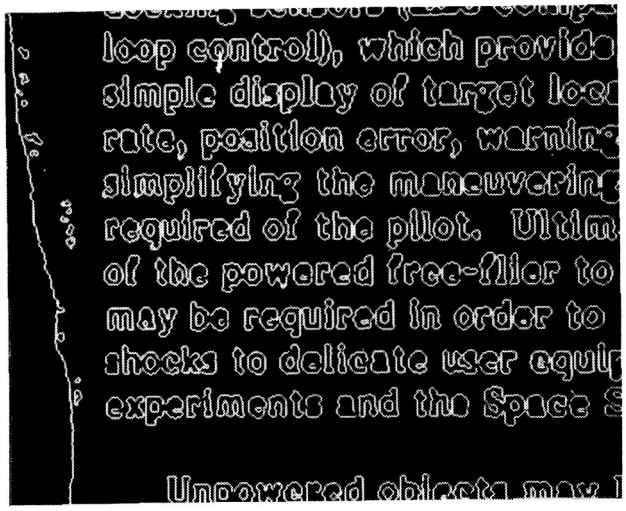
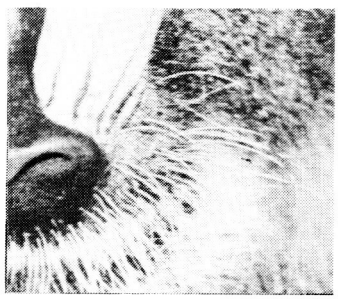


Figure 9. Continued

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH



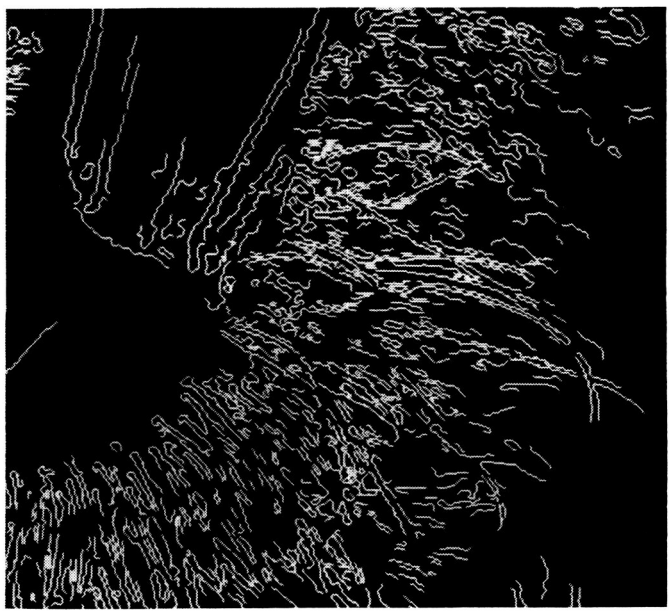


Figure 9. Continued

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

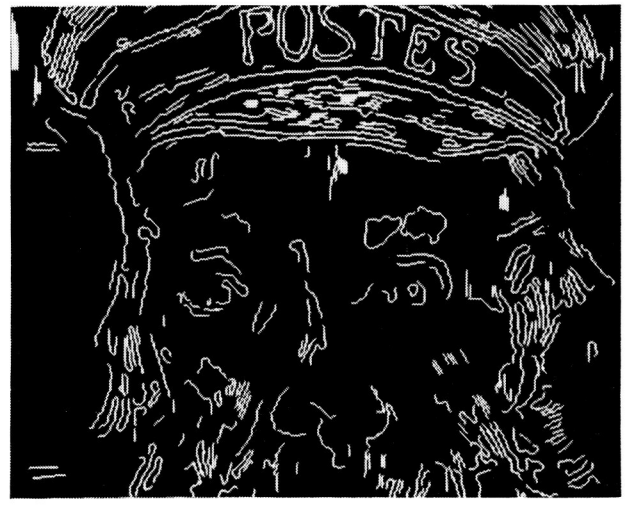


Figure 9. Concluded.

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

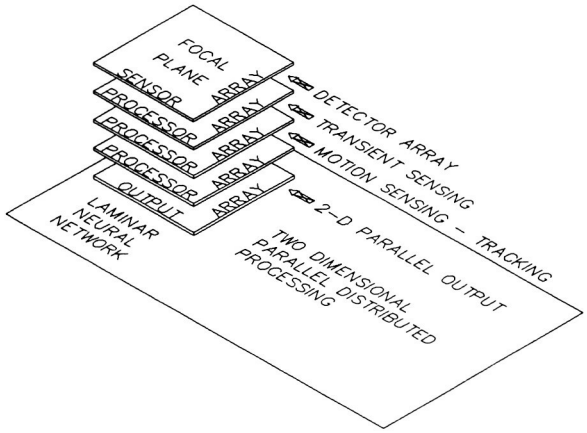


Figure 3: Illustrative laminar architecture showing stacked wafers in three dimensions.

Biological image processing strategies and algorithms are of interest to NASA because of their potential for practical application. Of special interest is an algorithm of Cornsweet.[9,10] This Intensity Dependent Spatial Summation (IDS) algorithm is correlated with a number of quantitative, empirical aspects of vision.[11] The spatial scale of the associated point spread function is intensity dependent. Lower intensities are associated with a broader range of spatial integration. There is an interesting similarity between intensity dependent spatial integration and spiketrain coding of intensity in which the integration time is longer for lower intensities. Extended spatial integration and extended temporal integration are both strategies tradeoffs appropriate to coping with low intensity signal-to-noise problems.[12]

The Cornsweet algorithm is of interest in connection with edge detection, the identification of contours of objects and the specification of an image in terms of reflectance ratios. The temporal analog of a reflectance discontinuity at an edge is a step function intensity transient. One vision-system-like mode of transient sensing has already been demonstrated in our approach.[13] Implementation of the Cornsweet algorithm is a more subtle and interesting problem than transient sensing, although some insights may emerge from the similarity between spatial and temporal integration.

A parallel asynchronous hardware implementation of the Cornsweet algorithm would represent an interesting application of our approach. The ultimate and most challenging application would be real time, high frame rate, high resolution image processing. Hardware implementa-

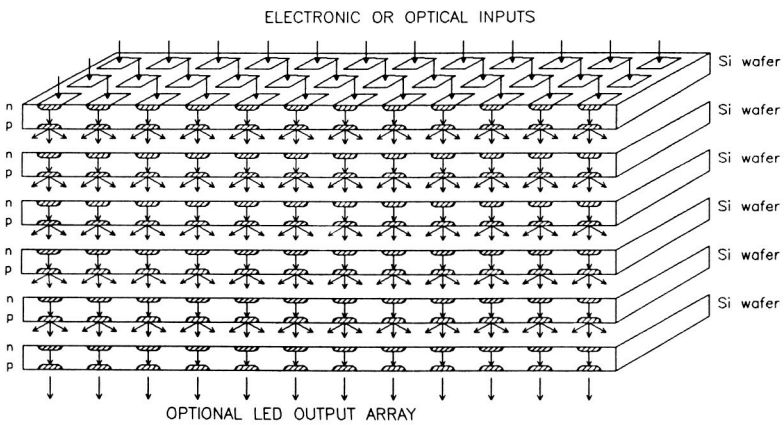


Figure 4: Schematic illustration (cross-sectional side view) of the signal flow pattern through a 2-D parallel asynchronous processor consisting of stacked silicon wafers. Parallel asynchronous fire-through is a key to propagation of pulsed signals through chips. Injection pulses are associated with current flow between the n- and p-layers.

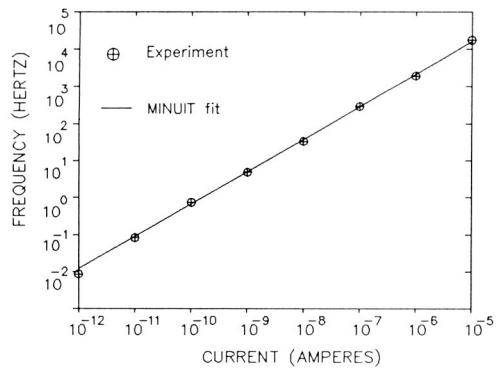


Figure 5: Experimental data points [6] and the calculated (MINUIT) fit. The output dynamic range is slightly less than the input dynamic range corresponding to sublinear current-to-frequency conversion. The model gives an extremely good fit which ranges across 7 decades.

tion of the Cornsweet algorithm is an unusually interesting area of research because the relevant retinal and neural mechanisms have not yet been identified.[1] On the other hand, a great deal is known about the connectivity of the retinal neural network, so that biological plausibility might be invoked as a broad, qualitative constraint on network architecture. For applications, it is of course not necessary to be unduly constrained by biological analogies and differences will surely appear in a silicon device approach. However, a point which is frequently made in connection with neural network research is that, at our present level of understanding, there is probably much to be gained from a reverse engineering analysis of high performance biological systems.

2 Devices For Parallel Asynchronous Processing

Previous studies[6] of current driven p^+-n-n^+ diodes led to the discovery of input current dynamic ranges up to 10^7. The corresponding output pulse rate range was sometimes less than the input range. See Fig.5. We have developed a model for spontaneous firing during current-to-frequency conversion (I-to-f conversion) and used the model to analyze the data shown in Fig. 5 using a program developed at CERN called MINUIT[14]. A key feature which is explained is that the slope of lnf vs lnI is not always unity. The data in Fig. 5 correspond to $f \propto I^{1-\epsilon}$. A simple picture with an equal amount of charge transfer in each impulse would explain $f \propto I$. However, more detailed device modeling was required to understand sublinear $f \propto I^{1-\epsilon}$ behavior.

2.1 Sensors and Sensor-Processor Interfacing

This section describes experimental work on sensors and sensor-processor interfacing. Results have been obtained for reverse biased p-i-n photodiodes which are useful in the visible, ultraviolet and near infrared regions and for infrared detectors which are useful in the far infrared region. The most dramatic results in terms of dynamic range came from visible light measurements with reverse biased p-i-n photodiodes where the dark current reduction associated with cooling led to the enormous dynamic range shown in Fig. 6.

[1] T. Cornsweet, private communication.

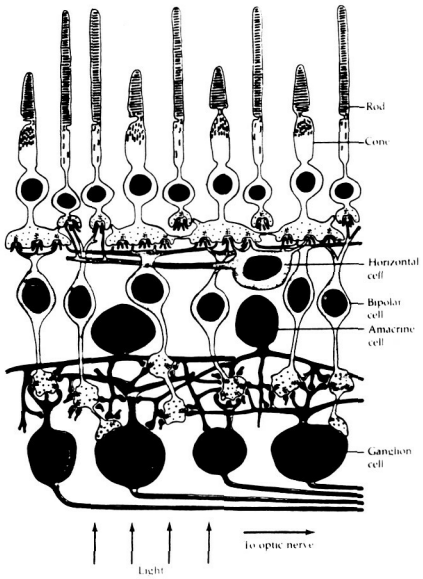


Figure 7: Schematic representation of the primate retina.[17] The light sensitive rods and cones pass signals to horizontal cells where lateral inhibition may occur. Signals then pass through bipolar cells to the highly interconnected plexiform layer. From the plexiform layer, signals are transmitted to the ganglion cells which connect to fibers of the optic nerve. Reproduced with the permission of Sinauer Assoc. Inc.

The relevant output from the IDS algorithm conveys information in the neighborhood of an edge, so it's not strictly one-dimensional. Besides edge contour information, there is additional information in the IDS algorithm output, namely, intensity-ratios or reflectance-ratios associated with the two regions on either side of the edge. If we assign one normalized reflectance to each 2-D subregion (plaquette) within a closed contour revealed by edge detection, then the IDS output data could be compressed into a "sketch" displaying 1-D edge contours (plaquette perimeters) and a set numbers (normalized reflectances), one for each 2-D plaquette.

2.5 Similarity with Image Processing in Natural Vision Systems

Our parallel asynchronous processing strategy, our neuronlike information coding and our intrinsically 2-D data flow all suggest a close analogy with natural vision systems. In addition, our approach preserves the geometrical relationship of neighboring channels as is the case in natural vision systems. A key aspect of processing in natural vision systems is lateral interaction between neighboring or nearby processing channels. It thus appears that our hardware approach is well suited to implementation of image processing schemes which parallel those of natural vision systems.

Lateral interaction between nearby processing channels is associated with vision system spatial filtering. Lateral interactions determine the receptive fields of neuron processing elements and the point-spread functions of individual photoreceptors. In natural vision systems, neurons mediate lateral interactions as shown in Fig. 7. In the retina chip of Mead and Mahowald, lateral interactions are incorporated via a resistive network.[16] However, no spiketrain generation and no intensity-to-frequency conversion occur as in the retinas of natural vision systems. See Fig. 8.

In our approach, neuronlike spiketrain generation is used. An artificial neuron circuit and the analogy with real, stereotypical neurons is illustrated in Fig. 9.[5]

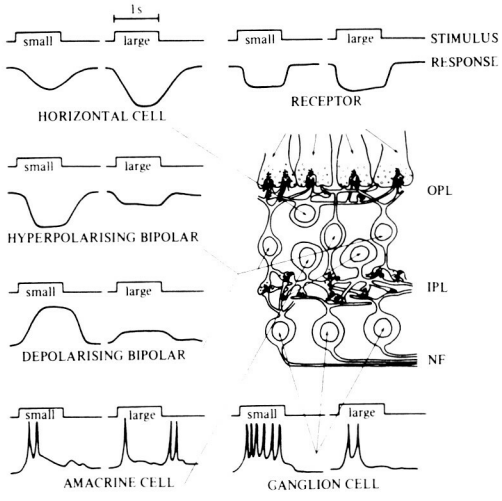


Figure 8: Waveforms recorded from various cells in the vertebrate retina when a small stimulus spot is shown on retina photoreceptors, and when a large spot that includes surrounding elements is used. The stimuli last about 1 second, and the responses are up to 30 mV in amplitude. OPL and IPL refer to outer and inner plexiform layers and NF refers to the nerve fibers. [18] Reproduced with the permission of Cambridge University Press.

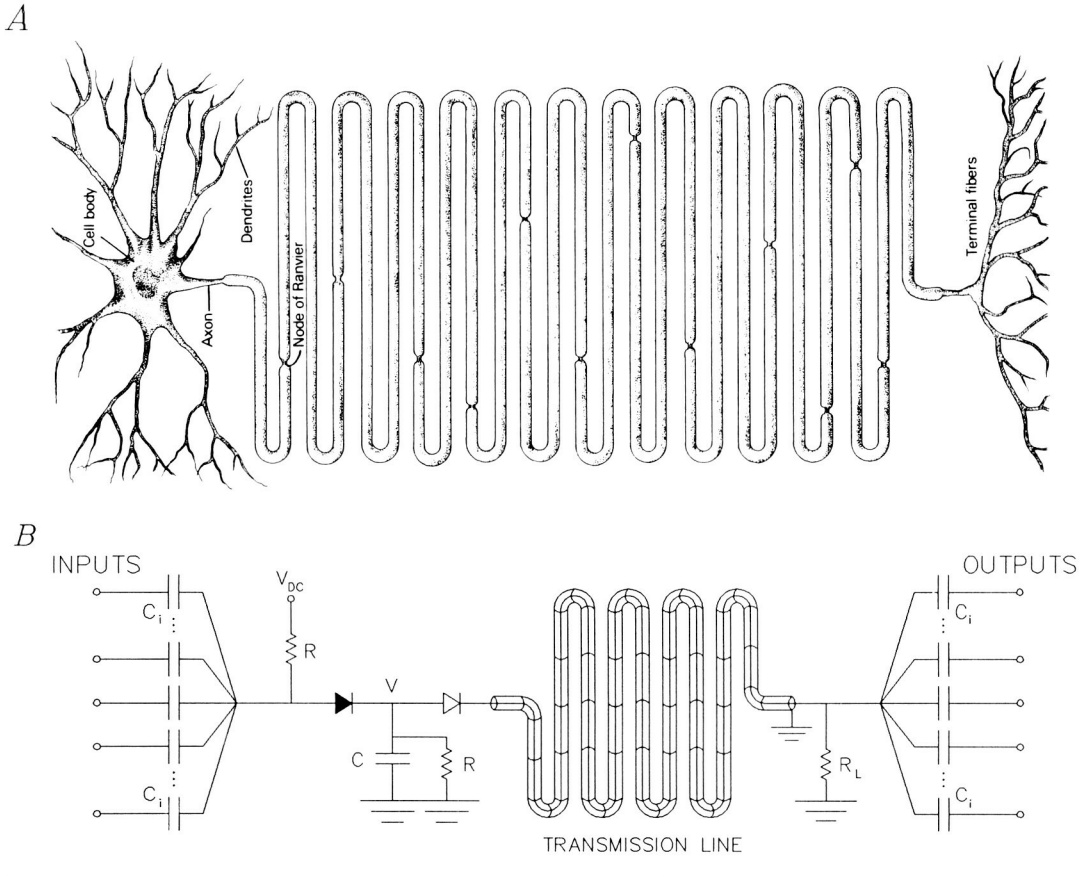


Figure 9: A): Features of a typical neuron from Kandel and Schwartz [19] and B): our artificial neuron, which exhibits the summation over synaptic inputs and fan-out. The input and output capacitive couplings are useful in conjunction with spiketrains. The darkened diode is a p-n junction device used for pulse height discrimination. The other diode is a p^+-n-n^+ diode used for spiketrain generation.

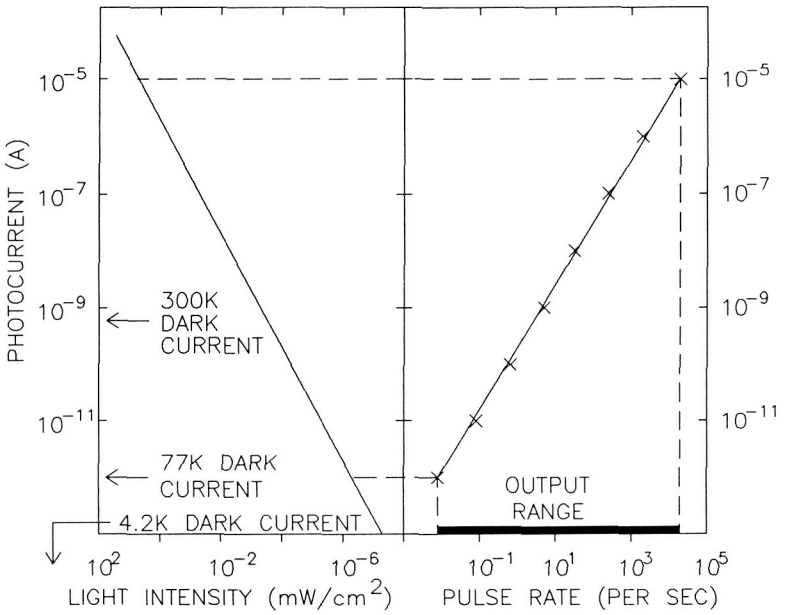


Figure 6: The graphs show that the photocurrent output of a reversed biased p-i-n photodiode (left graph) in response to visible light overlaps the large dynamic range of p^+-n-n^+ devices (right graph). This implies that such photodiodes could be directly interfaced to a parallel asynchronous processor based on p^+-n-n^+ devices. Such interfacing would preserve the high dynamic range.

2.2 Two-Dimensional Data Transfer

We have performed experiments in order to examine means of 2-D parallel data transfer without multiplexing. The idea is that neuronlike spiketrains could be used to drive arrays of LEDs for 2-D optical data transfer. The LED firing pattern could be recorded using a video camera or received by a photodiode array.

In our experiments, a p^+-n-n^+ diode drives a LED which is also inside a cryogenic environment. When a pulse from the p^+-n-n^+ diode goes above the threshold voltage of the LED, the LED starts conducting and emitting light. The LED inside the dewar can be viewed from outside the dewar. This is convenient and avoids a heat load. While the p^+-n-n^+ diode pulse is greater than the threshold the LED will be on. The pulse will decay according to the circuit parameters, i.e., the time constant. The speed of data transfer will be limited by the RC time constant. Optimal performance corresponds to dissipation of power in the LED rather than in the load resistor so that the RC decay is undesirable from the point of view of power considerations as well as avoidance of pile up at high pulse rates.

2.3 Ultralow Power Requirements

Massive processing tasks, operation in space and cooling for high performance (low dark current) operation are all factors which point to the benefits of low power operation. Von Neumann's estimate of the power consumption of the brain was 10-25 watts [15] which is remarkably small for a system with $\sim 10^{11}$ neurons, i.e. $\sim$ 100 picowatts/neuron. It has been argued that arrays of small p^+-n-n^+ diodes could offer comparably low (or even lower) power consumption.[5] Scaling down the device size will scale down the power requirements per device. For p^+-n-n^+ diodes, we have observed pulses with energy dissipation down to 4 picojoules/mm^2/pulse and a quiescent power dissipation of 10 picowatts/mm^2. Considering the thermodynamic efficiency of cooling, these numbers correspond to 290 picojoules/mm^2/pulse and 710 picowatts/mm^2 at room temperature. For comparison, we note that the retina chip of Mead and Mahowald[16]

has a power dissipation of 4 microwatts/mm^2.

The range of p^+-n-n^+ diode action potential pulse heights observed to date is from about 20 millivolts to 50 volts with the low end of the range corresponding to the low power figures reported here. The low end of this range of pulse heights is comparable to action potential pulse heights of real neurons. Our device physics modeling of spiketrain generation could lead to further power reductions if necessary. Low power dissipation would permit substantial processing to be performed at or just behind a focal plane array of detectors which are normally cooled to achieve high performance.

This electronic approach is remarkably well suited to neural network emulation and parallel asynchronous processing. Such hardware offers the possibility of 2-D parallel image processing in conjunction with image acquisition in much the same way as image acquisition and early processing are performed in natural vision systems. This is of interest because it is generally acknowledged that many image processing tasks are performed by natural vision systems with noteworthy speed, even in comparison with the fastest available systems employing conventional electronics.

We have identified certain image processing algorithms (IDS and pyramid) as being (A) especially well suited to our 2-D parallel approach and (B) of special relevance to potential NASA applications. The ultimate system which could emerge from our research would be a real time, high resolution, high dynamic range, low power integrated (single package) focal plane array-2-D parallel processor. The processor would be hard-wired to implement particular algorithms. Successive processing levels could perform a succession of processing tasks. For example, one might want to perform further parallel processing on the output of an IDS algorithm stage.

2.4 Parallel Processing Speed and Data Compression

In a fully parallel processing system the bandwidth per processing channel can be of the order of the bandwidth required per pixel. By contrast, serial systems introduce bottlenecks and require higher processing speeds which scale with the array size.

Standard planar semiconductor technology dictates that signals be transmitted to the edge of chips where the 2-dimensional input image data flow confronts a 1-D perimeter bottleneck. The number of detectors per preamplifier is a measure of the chip level bottleneck. Conventional approaches thus require increased electronics to reduce bottlenecks. The devices discussed here require **no preamplifiers** so we are able to go all the way to 2-D parallel processing and eliminate bottlenecks without introducing a 2-D array of preamplifierss or even one preamplifier. For conventional systems, there would be higher power requirements proportional to the number of preamplifiers. This would be disadvantageous in a cryogenic environment especially in NASA space applications. (Note that conventional CCD cameras are cooled to achieve their best performance.)

The dimensionality of desired patterns at the output plane of a 2-D signal processor provide a rough qualitative measure of the degree of data compression which is possible:

Detected Features	*Input → Output Dimensionality*
Set of point targets	2D → 0D
Set of edges	2D → 1D
Full optical data flow	2D → 2D