ENGINEERING

USCIPI REPORT 660

UNIVERSITY OF SOUTHERN CALIFORNIA

SEMIANNUAL TECHNICAL REPORT

William K. Pratt
Project Director

Covering Research Activity During the Period
1 September 1975 through 31 March 1976

31 March 1976

Image Processing Institute
University of Southern California
University Park
Los Angeles, California 90007

Sponsored by
Advanced Research Projects Agency
Contract No. F-33615-76-C-1203
ARPA Order No. 3119

IPI

IMAGE PROCESSING INSTITUTE

Page 244 is not missing but was mismumbered see Mr. John Kmak.

JRS
2/9/77

USCIPI REPORT 660

SEMIANNUAL TECHNICAL REPORT

Covering Research Activity During the Period
1 September 1975 to 31 March 1976

William K. Pratt
Project Director
(213) 746-2694

Image Processing Institute
University of Southern California
University Park
Los Angeles, California 90007

31 March 1976

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Eastern Test Range under Contract No. F-33615-76-C-1203, ARPA Order No. 3119

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Image Processing Institute
University of Southern California, University Park
Los Angeles, California 90007

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

IMAGE PROCESSING RESEARCH.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Semiannual 1 Sep 75 - 31 Mar 76

5. AUTHOR(S) (First name, middle initial, last name)

William K. Pratt (Project Director)

6. REPORT DATE

31 Mar 76

7a. TOTAL NO. OF PAGES

263

7b. NO. OF REFS

105

8a. CONTRACT OR GRANT NO.

F33615-76-C-1203

b. PROJECT NO.

ARPA Order 3119

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

USCIPI 660

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for release; distribution unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

13. ABSTRACT

This technical report summarizes the image understanding and image processing research activities performed by the University of Southern California during the period of 1 September 1975 to 31 March 1976 under Contract No. F-33615-76-C-1203 with the Advanced Research Projects Agency Information Processing Techniques Office.

The research program has as its primary purpose the development of techniques and systems for efficiently processing, transmitting, and analyzing visual images and two dimensional data arrays. Three tasks are reported: Image Understanding Projects; Image Processing Projects; Smart Sensor Projects. The Image Understanding Projects involve the development of generalized processing systems for analyzing images and extracting salient information. The Image Processing Projects include research on novel image coding methods, image restoration, vision modelling and nonlinear two dimensional optical filtering techniques. The Smart Sensors Projects comprise investigations of electronic and optical processing methods integrated with imaging sensors to perform feature and symbolic extraction.

* *

14. Key words: Image Processing, Digital Image Processing, Image Coding, Image Enhancement, Image Restoration, Color Image Processing, Knowledge Bases, Artificial Intelligence, Scene Analysis, Image Understanding, Edge Detection, Texture Analysis.

DD FORM 1 NOV 65 1473

391 141

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

ABSTRACT

This technical report summarizes the image understanding and image processing research activities performed by the University of Southern California during the period of 1 September 1975 to 31 March 1976 under Contract No. F-33615-76-C-1203 with the Advanced Research Projects Agency Information Processing Techniques Office.

The research program has as its primary purpose the development of techniques and systems for efficiently processing, transmitting, and analyzing visual images and two dimensional data arrays. Three tasks are reported: Image Understanding Projects; Image Processing Projects; Smart Sensor Projects. The Image Understanding Projects involve the development of generalized processing systems for analyzing images and extracting salient information. The Image Processing Projects include research on novel image coding methods, image restoration, vision modelling, and nonlinear two dimensional optical filtering techniques. The Smart Sensors Projects comprise investigations of electronic and optical processing methods integrated with imaging sensors to perform feature and symbolic extraction.

PROJECT PARTICIPANTS

Project Director

William K. Pratt

Research Staff

Harry C. Andrews
Werner Frei
Nasser E. Nahi
Ramakant Nevatia
Guner S. Robinson
Erica Rounds
Alexander A. Sawchuk

Support Staff

Reiner Lenzen
Toyone Mayeda
James M. Pepin
Ray Schmidt
Joyce Seguy
Dennis Smith
Florence B. Tebbets

Students

Ikram Abdou
Behnam Ashjari
Ben Britt
Marilyn Chan
Stephen Dashiell
Faramarz Davarian
Greg Finn
Chung-Kai Hsueh
Michael Huhns
Steve Hou
Mohammad Jahanshahi
Scott Johnston
Johnny S. Lee
Simon Lopez-Mora
Dennis McCaughey
Clanton Mancill
Lee Martin
David Mehrle
Firouz Naderi
David Nagai
Clay Olmstead
Javad Peyrovian
Jin Soh
Tai Sung

Table of Contents

1. Research Project Overview

This report describes the progress and results of the University of Southern California image understanding and image processing research study for the period of 1 September 1975 to 31 March 1976. The research study has been subdivided into three projects:

Image Understanding Projects

Image Processing Projects

Smart Sensors Projects

The image understanding projects involve research directed toward the goal of developing generalized processing systems capable of analyzing images and extracting salient information. Specific studies include feature extraction, symbolic description, interpretation, and systems analysis. The image processing projects include research on novel image coding techniques based upon results of the image understanding study, advanced image restoration methods, vision modelling in support of image understanding, and studies of nonlinear two dimensional optical filtering techniques for the implementation of imaging sensors. The smart sensors projects comprise investigations of electronic and optical processing methods which can be integrated with imaging sensors to perform low level image enhancement and feature extraction within the sensor. The smart sensors research work is being performed by the Hughes Aircraft Company Research Laboratories in Malibu, California under subcontract to USC.

Section 2 of this report details the research effort on the image

understanding projects. The image processing research activities are described in Section 3, and Section 4 covers the work on the smart sensors projects. Section 5 is a list of publications by project members during the reporting period.

2. Image Understanding Projects

The effort in the image understanding projects is directed toward the eventual development of processing systems for generalized image analysis. Applications of such systems include photointerpretation, vehicle guidance, visual tracking, map matching, cartography, and image communication.

The research study is organized according to the structure of a conceptual image understanding system. This system consists of a feature extraction stage which detects and measures primitive features such as edges and texture regions from an input image. Next the primitive features are grouped into meaningful symbols such as object boundaries, segments of some common attribute, or basic shape structures. Finally the image symbols are interpreted in terms of their semantic relationships to produce a concise quantitative description of the original image. Some knowledge base is assumed available for guidance of all elements of the image understanding system.

The philosophic approach to the study has been to bring together a research team skilled in digital signal processing and concepts of artificial intelligence. Although the research effort at USC involves all stages of the overall image understanding system, initial emphasis has been deliberately placed on the feature extraction and symbolic description system elements. Rapid progress is needed in these areas to assist the research in the higher stages of the image understanding system.

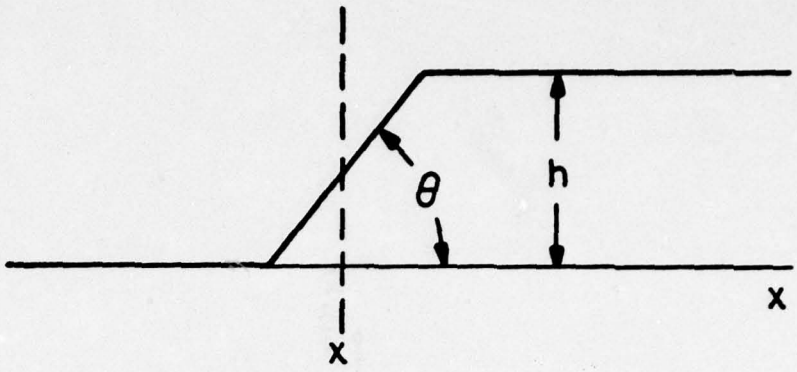
2.1 Luminance Edge Detection Techniques

William K. Pratt

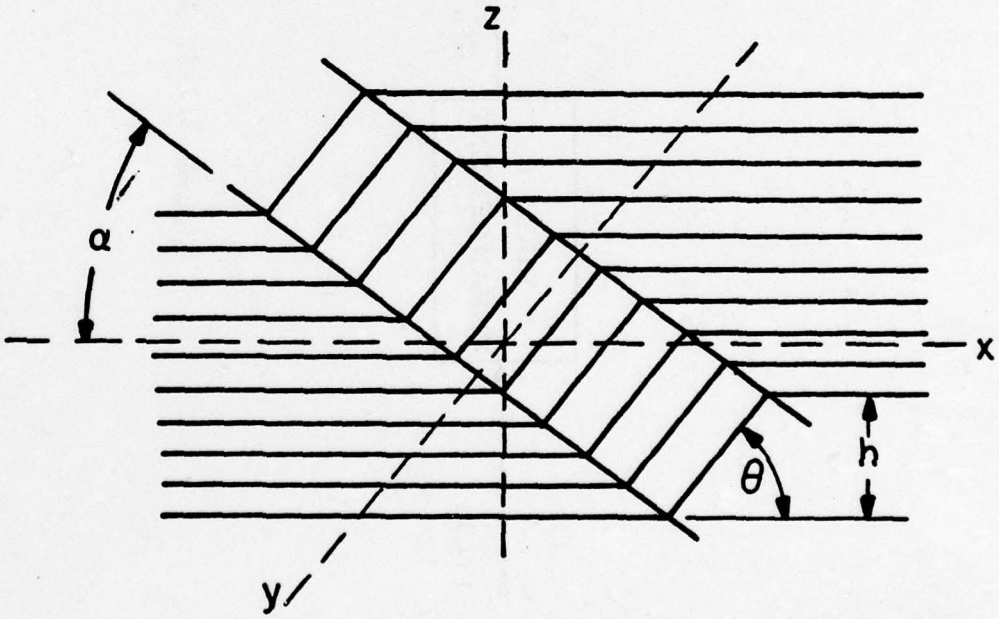
Changes or discontinuities in an image attribute such as luminance, tristimulus value, or texture are fundamentally important primitive features of an image since they often provide an indication of the spatial extent of objects within the image. Local discontinuities in image luminance or amplitude level are called luminance edges, while global luminance discontinuities are defined to be boundaries. This report presents a survey of luminance edge detection techniques.

Figure 1 contains sketches of one and two dimensional luminance edges each represented as a ramp increase in image amplitude level from a low to a high level. In the one dimensional case the edge is characterized by its height, the slope angle, and x-coordinate of the slope midpoint. An edge exists if both the slope angle and height are larger than specified critical values. For the two dimensional example the orientation with respect to the x-axis is also of importance. An ideal edge detector processing the image regions of figure 1 should produce an edge indication localized to a single pixel located at the midpoint of the slope.

A common aproach to monochrome edge detection is illustrated in figure 2 in which an original monochrome image F(j,k) undergoes a grey scale edge enhancement by linear or nonlinear processing to produce an image field G(j,k) with accentuated spatial luminance changes. Next,



(a) ONE DIMENSIONAL EDGE



(b) TWO DIMENSIONAL EDGE

Figure 2.1-1. One and two dimensional edges.

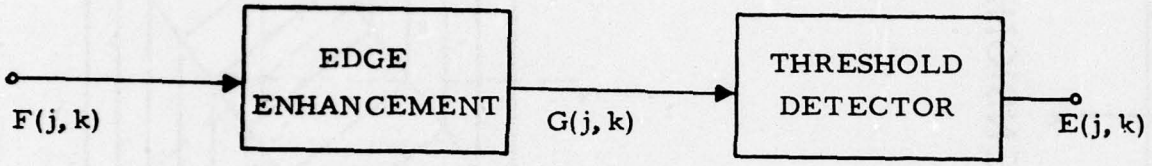


Figure 2.1-2. Edge Detection System.

a threshold operation is performed to determine the pixel location of significant edges. A negative going edge exists if

$$G(j,k) < T_L(j,k) \tag{2.1-1a}$$

and a positive going edge exists if

$$G(j,k) \geq T_U(j,k) \tag{2.1-1b}$$

where $T_L(j,k)$ and $T_U(j,k)$ are lower and upper threshold values, respectively. These threshold values may be made spatially varying to compensate for gross spatial luminance changes. Threshold selection is one of the key issues in edge detection. A threshold level set too high will not permit detection of low amplitude structural image elements. Conversely, a threshold level set too low will cause noise to be falsely detected as an image edge. An edge location map $E(j,k)$ is often generated to indicate the positions of edges within an image. For example, all positive edge locations could be indicated by white pixel values against a black background. Alternatively, positive going edges could be noted by white pixels, negative going edges by black pixels, and non-edge regions by mid-grey pixels.

A second major approach to luminance edge detection involves fitting of a local region of pixel values to some ideal representation of a one or two dimensional edge, as defined in figure 1. If the fit is sufficiently close, an edge is said to exist, and its assigned parameters are those of the ideal edge.

Linear Edge Enhancement Methods: A variety of edge enhancement techniques can be utilized to accentuate edges prior to threshold detection. One of the simplest techniques is discrete differencing analogous to continuous spatial differentiation. Horizontal edge sharpening can be obtained by the running difference operation which produces an output image according to the relation

$$G(j,k) = F(j,k) - F(j,k+1) \tag{2.1-2a}$$

Similarly, vertical sharpening results from the operation

$$G(j,k) = F(j,k) - F(j+1,k) \tag{2.1-2b}$$

Diagonal sharpening can be obtained by subtraction of diagonal pairs of pixels.

Horizontal edge accentuation can also be accomplished by forming the differences between the slopes of the image amplitude along a line according to the relation

$$G(j,k) = [F(j,k)-F(j,k-1)]-[F(j,k+1)-F(j,k)] \tag{2.1-3a}$$

or equivalently

$$G(j,k) = 2F(j,k)-F(j,k-1)-F(j,k+1) \tag{2.1-3b}$$

Similar expressions exist for vertical and diagonal slope differences.

Two dimensional discrete differentiation can be performed by convolving the original image array with the compass gradient masks listed below <1,p.111>:

North

$$\underline{H} = \begin{matrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{matrix} \qquad (2.1\text{-}4a)$$

Northeast

$$\underline{H} = \begin{matrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{matrix} \qquad (2.1\text{-}4b)$$

East

$$\underline{H} = \begin{matrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{matrix} \qquad (2.1\text{-}4c)$$

Southeast

$$\underline{H} = \begin{matrix} -1 & -1 & 1 \\ -1 & -2 & 1 \\ 1 & 1 & 1 \end{matrix} \qquad (2.1\text{-}4d)$$

South

$$\underline{H} = \begin{matrix} -1 & -1 & -1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{matrix} \qquad (2.1\text{-}4e)$$

Southwest

$$\underline{H} = \begin{matrix} 1 & -1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & 1 \end{matrix} \qquad (2.1\text{-}4f)$$

West

$$\underline{H} = \begin{matrix} 1 & 1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & -1 \end{matrix} \qquad (2.1\text{-}4g)$$

Northwest

$$\underline{H} = \begin{matrix} 1 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -1 & -1 \end{matrix} \qquad (2.1\text{-}4h)$$

The compass names indicate the slope direction of maximum response, e.g. the East gradient mask produces a maximum output for horizontal luminance changes from left to right. It should be noted that the gradient masks have zero weighting (the sum of the array elements is zero) so that there is no output response over constant luminance regions of the image.

Edge sharpening, without regard to edge direction, can be obtained by convolution of an image with a Laplacian mask. Several types of Laplacian masks are listed below:

Mask 1

$$\underline{H} = \begin{matrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{matrix} \qquad (2.1\text{-}5a)$$

Mask 2

$$\underline{H} = \begin{matrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{matrix} \qquad (2.1\text{-}5b)$$

Mask 3

$$\underline{H} = \begin{matrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{matrix} \qquad (2.1\text{-}5c)$$

Edge sharpening can be made proportional to the statistical correlation of pixel values by the statistical mask <1,p.125>

$$\underline{H} = \begin{bmatrix} \rho_C\rho_R & -\rho_C(1+\rho_R^2) & \rho_C\rho_R \\ -\rho_R(1+\rho_C^2) & (1+\rho_C^2)(1+\rho_R^2) & -\rho_R(1+\rho_C^2) \\ \rho_C\rho_R & -\rho_C(1+\rho_R^2) & \rho_C\rho_R \end{bmatrix} \tag{2.1-6}$$

in which ρ_R and ρ_C represent the assumed Markovian correlation factor between adjacent row and column pixels. If $\rho_R = \rho_C = 0$, there is no adjacent element correlation and the statistical mask has no effect; in the extreme, if $\rho_R = \rho_C = 1$, the statistical mask reduces to the Laplacian mask of eq. (5c).

Argyle <2> and Macleod <3,4> have proposed Gaussian shaped weighting functions as a means of edge enhancement. The Argyle function is a split Gaussian function defined in one dimension as

$$h(x) = \exp\left\{-\left(\frac{x}{p}\right)^2\right\} \qquad x \geq 0$$

$$h(x) = -\exp\left\{-\left(\frac{x}{p}\right)^2\right\} \qquad x < 0 \tag{2.1-7}$$

where p is a spread constant. The Macleod function given by

$$H(x,y) = \exp\left\{-\left(\frac{x}{t}\right)^2\right\}\left[\exp\left\{-\left(\frac{x-p}{p}\right)^2\right\} - \exp\left\{-\left(\frac{x+p}{p}\right)^2\right\}\right] \tag{2.1-8}$$

where p and t are spread constants. This function suppresses the effect of pixel values in the edge transition region and edges in rows above and below the edge to be detected.

A common limitation of the linear edge sharpening methods previously discussed is the amplification of high spatial frequency

noise and artifacts as a result of the inherent differencing operations involved. Noise smoothing can be encorporated into the linear edge sharpening methods by performing the linear masking on regions of pixels rather than on individual pixels <5>. This can be accomplished by forming a linear mask

$$H(j,k) = H_S(j,k) \circledast H(j,k) \qquad (2.1\text{-}9)$$

by convolving one of the edge enhancement masks previously defined with a low pass filter averaging mask $H_S(j,k)$. Such spatial averaging, of course, leads to a smoothing of edges as well as noise.

Nonlinear Edge Enhancement Methods: Nonlinear edge detection systems perform nonlinear combinations of pixel values as a means of edge enhancement prior to thresholding. Most techniques are limited to processing over small 2 x 2 or 3 x 3 pixel windows.

Roberts <6> has introduced the nonlinear cross-operation

$$G_R(j,k) = \left([F(j,k) - F(j+1,k+1)]^2 + [F(j,k+1) - F(j+1,k)]^2 \right)^{1/2} \qquad (2.1\text{-}10)$$

as a two-dimensional differencing method for edge sharpening and edge isolation. Another spatial differencing operation, which is of computationally simpler form, is given by

$$G_A(j,k) = |F(j,k) - F(j+1,k+1)| + |F(j,k+1) - F(j+1,k)| \qquad (2.1\text{-}11)$$

It can be easily shown that

$$G_R(j,k) \leq G_A(j,k) \leq \sqrt{2}\, G_R(j,k) \qquad (2.1\text{-}12)$$

Crude directional information can be extracted by noting which of the four pixels is largest at a detected edge point.

Sobel <7,p.271> has suggested a 3 x 3 nonlinear edge enhancement operator described by the pixel numbering convention of figure 3. The edge enhancement plane is defined as

$$G(j,k) = \sqrt{X^2 + Y^2} \tag{2.1-13a}$$

where

$$X = (A_2 + 2A_3 + A_4) - (A_0 + 2A_7 + A_6) \tag{2.1-13b}$$

$$Y = (A_0 + 2A_1 + A_2) - (A_6 + 2A_5 + A_4) \tag{2.1-13c}$$

Another 3 x 3 nonlinear edge enhancement algorithm has been introduced by Kirsch <8>. Referring to the notation of figure 3, the enhancement is given as

$$G(j,k) = \max\left\{1, \max_{i=1}^{7}\left[|5S_i - 3T_i|\right]\right\} \tag{2.1-14a}$$

where

$$S_i = A_i + A_{i+1} + A_{i+2} \tag{2.1-14b}$$

$$T_i = A_{i+3} + A_{i+4} + A_{i+5} + A_{i+6} + A_{i+7} \tag{2.1-14c}$$

The subscripts of S and T are evaluated modulo 8. Basically, the Kirsch operator provides the maximal compass gradient magnitude about

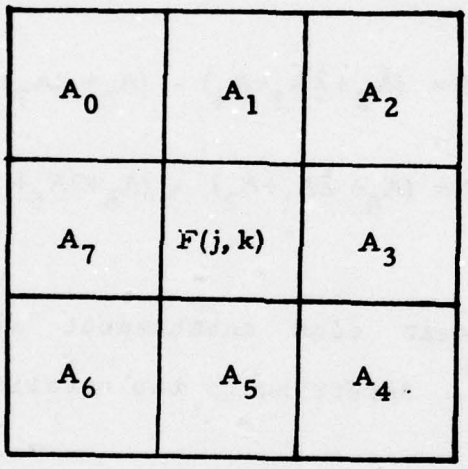


Figure 2.1-3. Numbering for 3 X 3 edge detection operators.

an image point ignoring the pixel value F(j,k).

Wallis <9> has proposed a nonlinear edge detection scheme based upon homomorphic image processing. According to this scheme an edge exists if the magnitude of the logarithm of the image luminance at a pixel exceeds the magnitude of the average logarithmic luminance of its four nearest neighbors by a fixed threshold value. With reference to figure 3, the edge enhancement plane is defined as

$$G(j,k) = \log[F(j,k)] - \tfrac{1}{4}\log(A_1) - \tfrac{1}{4}\log(A_3) - \tfrac{1}{4}\log(A_5) - \tfrac{1}{4}\log(A_7) \qquad (2.1\text{-}15a)$$

or

$$G(j,k) = \tfrac{1}{4}\log\left[\frac{F(j,k)}{A_1 A_3 A_5 A_7}\right] \qquad (2.1\text{-}15b)$$

Comparison of G(j,k) against upper and lower threshold values is exactly equivalent to comparison of the fraction in the brackets of eq. (15b) against a modified threshold. Therefore, logarithms need not be computed. The principal advantage of the logarithmic edge detector besides its computational simplicity is that the technique is insensitive to multiplicative changes in luminance level.

The logarithmic edge enhancement scheme defined by eq. (15) can be considered as a linear enhancement with the Laplacian mask of eq. (5a) performed on the logarithms of the pixel values. In this context other edge enhancement methods can be easily formulated as a concatenation of point nonlinear operations followed by linear edge enhancement and thresholding.

Rosenfeld <10> has developed a nonlinear product averaging mask for edge sharpening and edge isolation. With this operator the running one-dimensional average

$$D_M(j,k)=\frac{1}{M}[F(j+M-1,k)+F(j+M-2,k)+\ldots+F(j,k) -F(j-1,k)-F(j-2,k)-\ldots F(j-M,k)] \quad (2.1\text{-}16)$$

is formed at each pixel point where $M=2^m$ and m is an integer. This operation is performed for M=1,2,4,8,16,...,etc. up to some desired upper limit. Then, the product

$$P_M(j,k)=D_1(j,k)D_2(j,k)\ldots D_M(j,k) \quad (2.1\text{-}17)$$

is formed at each pixel. Conceptually, the higher order averaging masks provide a broad indication of edges and some degree of noise suppression, while the lower order averaging masks yield localized derivatives, which are much more noise sensitive. Together, it is hypothesized, the product of the variable length averaging masks should give a positive indication in the vicinity of true edges. Rosenfeld's justification of the operation is as follows: "...the result $P_M(j,k)$ tends to yield sharply localized detections of major edges while suppressing noise. Intuitively, this is because the product is large only when all factors are large, and as soon as one moves away from a position "just at" and edge point, the factors with low m's become small; while if one is not at or near a major edge, the factors with high m's are small."

Rosenfeld <11> has also proposed a nonlinear processing procedure for isolating large sharp edges in the neighborhood of smaller edges.

This procedure, which will be called dominant neighbor suppression, is performed by scanning the edge enhanced plane G(j,k) with a small pixel window. The value of G(j,k) in the center of the window is suppressed (set to zero) unless its magnitude is the greatest of all samples within the window. Conventional amplitude thresholding then follows. A variation of the process is to permit suppression of G(j,k) only if a neighbor in the window dominates by a significant amount. The dominant neighbor suppression thresholding algorithm has proven quite effective for edge detection when coupled with an edge enhancement method that provides some noise smoothing.

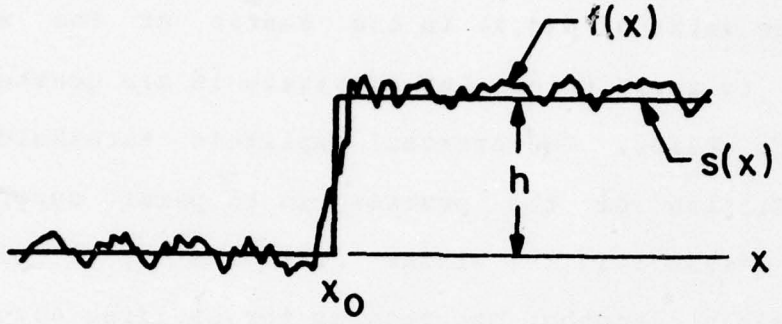
Edge Fitting Methods: Ideal edges may be viewed as one or two dimensional ramp signals of the form sketched in figure 1. Actual image data can then be matched against, or fit to, the ideal edge models. If the fit is sufficiently accurate at a given image location, an edge is assumed to exist with the same parameters as the ideal edge model.

In the one dimensional edge fitting case described in figure 4 the image signal f(x) is fit to a step function

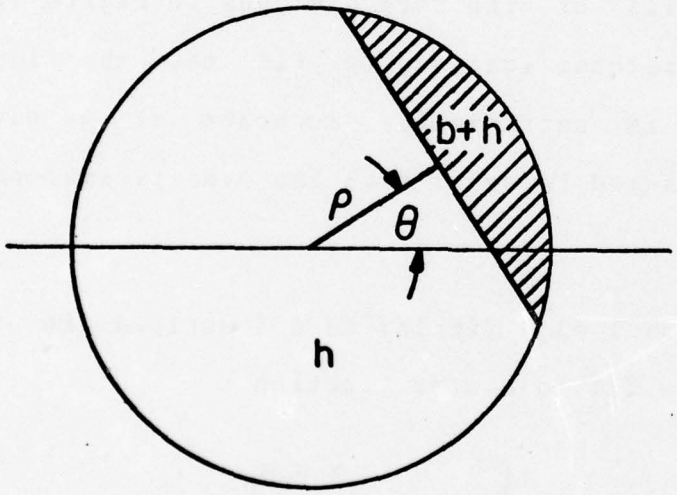
$$s(x) = \begin{cases} b & x < x_0 \\ b+h & x \geq 0 \end{cases} \qquad (2.1\text{-}18)$$

An edge is assumed present if the mean square error

$$\varepsilon = \int_{x_0-L}^{x_0+L} [f(x)-s(x)]^2 dx \qquad (2.1\text{-}19)$$



(a) ONE DIMENSIONAL



(b) TWO DIMENSIONAL

Figure 2.1-4. One and two dimensional edge fitting.

is below some threshold value. In the two dimensional formulation the ideal step edge is defined as

$$S(x,y) = \begin{cases} b & x\cos\theta + y\sin\theta < \rho \\ b+h & x\cos\theta + y\sin\theta \geq \rho \end{cases} \qquad (2.1\text{-}20)$$

where θ and ρ represent the polar distance from the center of a circular test region to the normal point of the edge. The edge fitting error is

$$\varepsilon = \iint_{\substack{\text{over} \\ \text{circle}}} [F(x,y) - S(x,y)]^2 dxdy \qquad (2.1\text{-}21)$$

Hueckel <12> has developed a procedure for two dimensional edge fitting in which the image points within the circle of figure 4 are expanded in a set of two dimensional basis functions by a Fourier series in polar coordinates. Let $H_i(x,y)$ represent the basis functions. Then the weighting coefficients for the expansions of the image and the ideal step edge become

$$a_i = \iint H_i(x,y)F(x,y)dxdy \qquad (2.1\text{-}22a)$$

$$b_i = \iint H_i(x,y)S(x,y)dxdy \qquad (2.1\text{-}22b)$$

It should be noted that $S(x,y)$ is defined parametrically in terms of the set (b,h,ρ,θ). In Hueckel's algorithm the expansion is truncated to eight terms for computational economy and to provide some inherent noise smoothing. Minimization of the mean square difference of

eq.(21) is equivalent to minimization of $(a_i - b_i)^2$ for all coefficients. Hueckel has performed this minimization, invoking some simplifying approximations, and has formulated a set of nonlinear equations expressing the edge parameter set (b,h,ρ,θ) in terms of the expansions coefficients a_i. A test is performed to compare the actual image data F(x,y) with the edge fit, and if the resulting fit is poor because of noise, no edge is judged present. If the fit is sufficiently close, a further test is made to determine if the edge contrast factor h is greater than a threshold factor.

The complexity of the Hueckel algorithm renders it difficult to analyze theoretically. However, experimental evidence indicates that the Hueckel operator is quite sensitive to the detection of low contrast edges in reasonably noisy images.

References

1. J.M.S. Prewitt, "Object Enhancement and Extraction," in Picture Processing and Psychopictorics, B.S. Lipkin and A. Rosenfeld, eds., Academic Press, New York, 1970.

2. E. Argyle, "Techniques for Edge Detection," Proceedings IEEE, Vol. 59, No. 2, February 1971, pp. 285-287.

3. I.D.G. Macleod, "On Finding Structure in Pictures," in Picture Language Machines, S. Kaneff, ed., Academic Press, New York, 1970, p. 231.

4. I.D.G. Macleod, "Comments on 'Techniques for Edge Detection,'"

Proceedings IEEE, Vol. 60, No. 3, March 1972, p. 344.

5. A. Rosenfeld, M. Thurston and Y. Lee, "Edge and Curve Detection: Further Experiments," IEEE Transactions on Computers, Vol. C-21, No. 7, July 1972, pp. 677-715.

6. L.G. Roberts, "Machine Perception of Three-Dimensional Solids," in Optical and Electro-Optical Information Processing, J.T. Tippett, et.al., eds., Massachusetts Institute of Technology Press, Cambridge, Massachusetts, 1965, pp. 159-197.

7. R.O. Duda and P.E. Hart, Pattern Classification and Scene Analysis, John Wiley and Sons, New York, 1973.

8. R. Kirsch, "Computer Determination of the Constituent Structure of Biological Images, Computers and Biomedical Research, Vol. 4, 1971.

9. R. Wallis, private communication.

10. A. Rosenfeld, "A Nonlinear Edge Detection Technique," Proceedings IEEE Letters, Vol. 58, No. 5, May 1970, pp. 814-816.

11. A. Rosenfeld and M. Thurston, "Edge and Curve Detection for Visual Scene Analysis," IEEE Transactions on Computers, Vol. C-20, No. 5, May 1971, pp. 562-569.

12. M. Hueckel, "An Operator Which Locates Edges in Digital Pictures," Journal ACM, Vol. 18, No. 1, January 1971, pp. 113-125.

2.2 A New Class of Edge and Feature Detection Operators

Werner Frei and Chung-Ching Chen

Human or machine scene analysis depends greatly upon the ability to partition an image into regions that correspond to individual objects. One popular approach is the detection of edges, which are crudely defined as the boundaries between image regions having different properties. Such properties are brightness, color, texture or whatever might enable the detection of object boundaries within the image. Clearly then, edge detection must be based upon the inspection of more than one picture element within sub-areas of an image in order to decide whether an edge segment is present within each area. This operation can be expressed as

$$\text{edge} = T\{f(A)\} \qquad (2.2\text{-}1)$$

where $f(\cdot)$ represents some function of the picture elements within area A, and $T(\cdot)$ is a threshold or decision function. The resulting edge elements can be characterized by variables such as amplitude, orientation, position, and a confidence factor.

Upon examination of the whole picture, object boundaries are constructed by connecting the edge elements detected into lines, an operation that can be based upon simple syntactic rules, for example, to connect neighbor edge elements that have similar orientation and delete isolated elements. Some of the difficulties of edge detection are created by noise, but much more so by the facts that visually distinct edges sometimes cannot be detected within an area of size A,

or conversely, that what appears to be an edge within A sometimes belongs to a homogeneously textured area of the picture. Increasing the size of A would apparently solve the problem, except for computational limitations. Also note that if the size of A is increased greatly, the description of an edge within A becomes very complex. It is generally recognized that boundary detection is therefore best done by the combination of a relatively simple edge operator followed by algorithms such as thinning and linking that delete irrelevant edge points and connect the remaining ones into lines. A number of such edge detection operators have been published, for example, Roberts gradient <1>, Kirsch <2>, Sobel <3>, Prewitt <4>, smoothed gradient <5>, Robinson <6> and Hueckel <7,8>. All of the above algorithms can be expressed in general terms as

$$\text{"edge"} = T\{f(\theta \cdot A)\} \qquad (2.2\text{-}2)$$

where θ is a linear transformation of A and f is a non-linear distance function. Edge detection is then accomplished by moving the area A over the whole picture sequentially, and recording the edge parameters found. Within that class of algorithms, the computationally fast operators consist of a set of templates W of size 2 x 2 or 3 x 3 pixels that are convolved simultaneously with the image to be analyzed. The maximum, sum of squares, or absolute sum of the convolutions is then thresholded to obtain edge points. Edge direction can also be determined in a straightforward manner, either by identifying the orientation of the template which yields the largest convolution value at each point, or by an arctangent

computation <1-6>.

Comparison of the above operators reveals similarities that can be attributed to the heuristic design approaches used. Yet there does not seem to exist a generally recognized criterion for optimization; the evaluation of such an algorithm is performed chiefly by visual inspection of processed edge pictures. The exception is the Hueckel operator, in which an ideal edge is defined a priori, and the algorithm optimized to detect such ideal edges.

The present contribution follows a similar approach in the sense that ideal edge elements are defined first. The analysis reveals that no more than two templates are required for the optimal detection of such edges, and the parameters of the operator can be adjusted to match preferred types of non-ideal edges (sharp versus smooth). Given the axiomatic criterion of ideal edge elements, most existing algorithms can be readily evaluated, and it is easy to verify that many are computationally redundant.

The edge operator is then generalized for other image features such as ideal line elements and discrete Laplacian (ideal points) which, taken together, constitute a rudimentary but fast orthogonal feature transform. Examination of the characteristics of the transform domain suggests the elaboration of a measure of confidence, or conversely, the design of adaptive threshold strategies for edge and line detection. Furthermore, experimental results suggest that the above features provide appropriate elements for contextual thinning and linking algorithms.

Definitions: Consider an image region A of size n x n sampling intervals shown in figure 1a. An ideal edge element in the continuous image domain is defined as a straight boundary passing through the center of A that separates two regions of different average brightness $\bar{b}_1$ and $\bar{b}_2$. Adopting the convention $\bar{b}_1 > \bar{b}_2$, the direction Φ_E of the edge element is uniquely determined with respect to any arbitrary direction (figure 1b). Note that the ideal edge element is characterized by its magnitude $|\bar{b}_1 - \bar{b}_2|$ and orientation Φ_E, where $0 \leq \Phi_E \leq 2\pi$. Possible discrete representations of the sampled ideal edge element will now be examined.

Consider the set of n^2 brightness samples b_{ij} of A as an element of an n^2-dimensional vector space $\mathcal{B}$. The elements of $\mathcal{B}$ can be represented by a matrix B or a column vector $\underline{b}$. For example for n = 3

$$\underline{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \qquad (2.2\text{-}3a)$$

$$\underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_9 \end{bmatrix} \qquad (2.2\text{-}3b)$$

An inner product (dot product) on $\mathcal{B}$ is defined as

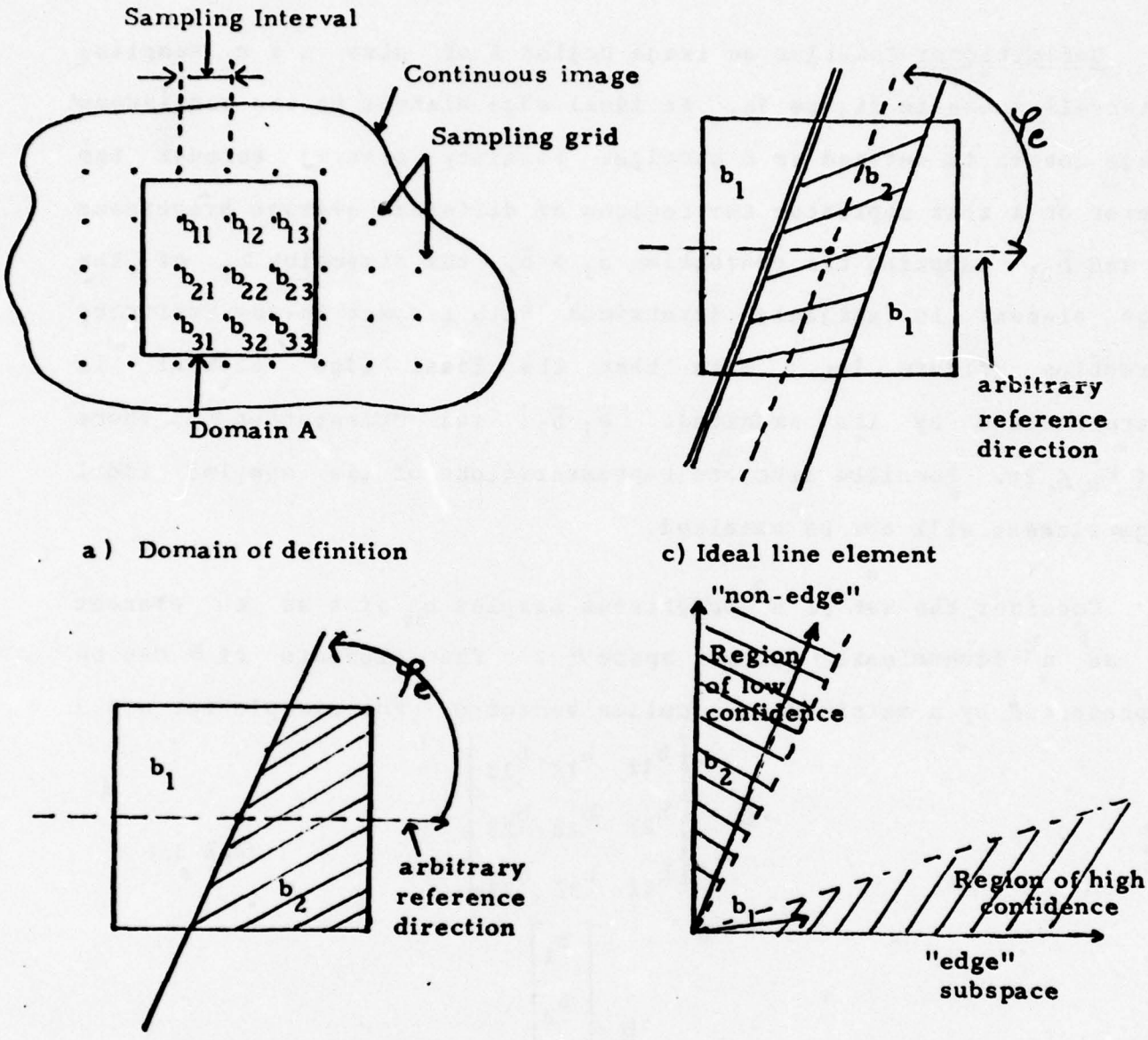


Figure 2.2-1. Definitions of ideal structural elements and concept of confidence.

$$(\underline{B}, \underline{C}) = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} c_{ij} \qquad (2.2\text{-}4a)$$

$$(\underline{b}, \underline{c}) = \sum_{i=1}^{n^2} b_i c_i \qquad (2.2\text{-}4b)$$

All of the fast edge operators mentioned previously can now be easily expressed in terms of the above notation. For example Roberts gradient is given by

$$\text{edge} = T\{(\underline{B}, \underline{W}_1)^2 + (\underline{B}, \underline{W}_2)^2\} \qquad (2.2\text{-}5)$$

with

$$\text{with} \quad \underline{W}_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \underline{W}_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad (2.2\text{-}6)$$

where $T(\cdot) = 0$ or 1 if the argument is smaller or larger, respectively, than a preset value. Note that $\underline{W}_1$ and $\underline{W}_2$ can be regarded as a pair of basis vectors for $\mathcal{B}$. The following hypothesis is now formulated:

> The discrete representation of all ideal edge elements defined above forms a proper subset $\mathcal{E}$ of $\mathcal{B}$, of dimension two. As a consequence, there must exist a basis consisting of $\underline{W}_1$ and $\underline{W}_2$ which spans $\mathcal{E}$ such that all ideal edge elements can be represented by no more than two linearly independent coefficients as

$$f_i = (\underline{B}, \underline{W}_i) \qquad (2.2\text{-}7)$$

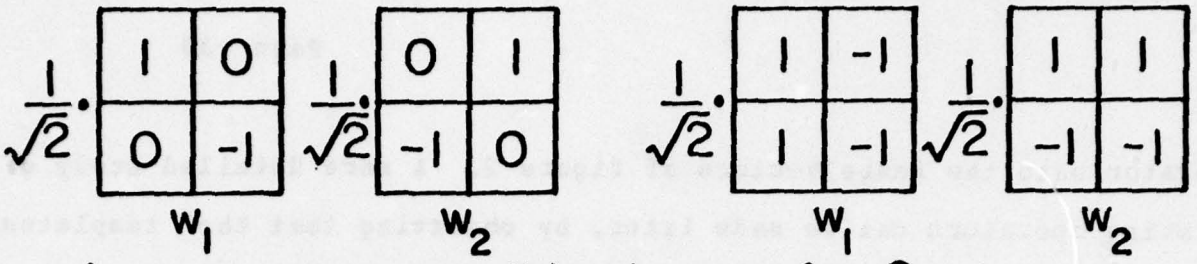
for i = 1,2.

Note that this formulation is intuitively satisfying in that ideal edge elements are characterized by two independent quantities, their magnitude and direction. If $\underline{h}_1$ and $\underline{h}_2$ are orthonormal, these quantities can be easily obtained from $\underline{f}_1$ and $\underline{f}_2$ by

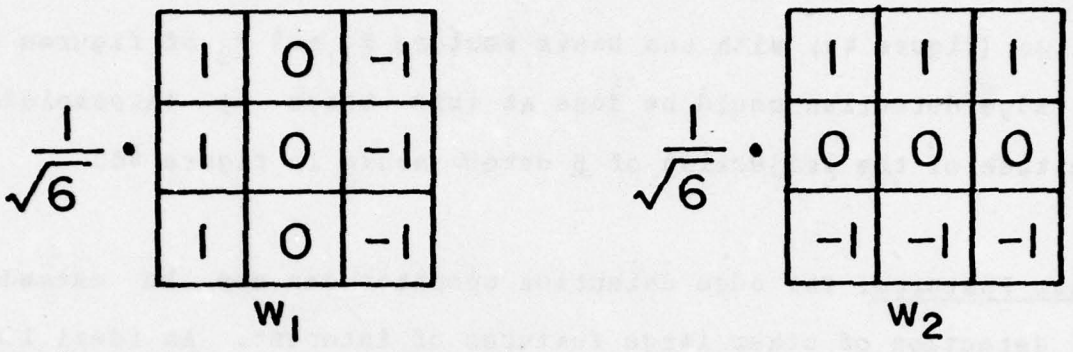
$$\text{"magnitude"} = (f_1^2 + f_2^2)^{\frac{1}{2}} \qquad (2.2\text{-}8a)$$

$$\text{"orientation"} = \operatorname{atan}\left(\frac{f_1}{f_2}\right) \qquad (2.2\text{-}8b)$$

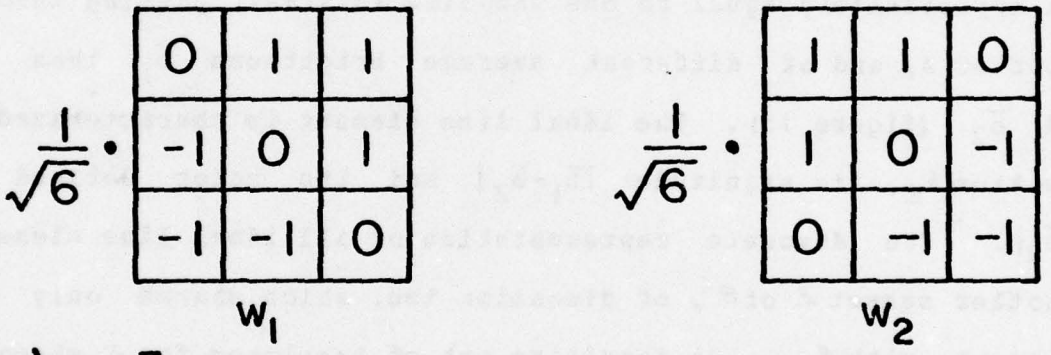
A more satisfying proof can be obtained by observing that all possible discrete representations of non-ideal edge elements can be obtained as a linear combination of $\underline{h}_1$, $\underline{h}_2$ and other non-edge basis vectors to be discussed. It has been found however that for n > 2, the subset δ is not unique, but depends upon the weighting function used to compute $\underline{h}_1$ and $\underline{h}_2$. Figure 2 shows examples of such basis vectors for n = 2 and n = 3. Note that what distinguishes the bases of figures 2b, 2c and 2d are the weights given to pixels as a function of distance from the separating boundary. This permits tuning of the edge operator to a certain extent for the detection of sharp or smooth edges. It is now possible to assess existing edge operators by projection of their templates onto the basis vectors found. One conclusion that can be drawn immediately is that any operator that makes use of more than two templates is computationally redundant because the excess templates are linear combinations of two edge basis vectors. Consequently, the excess convolutions are linearly dependant as well, and therefore redundant. Figure 3 shows projections of the templates of the Kirsch



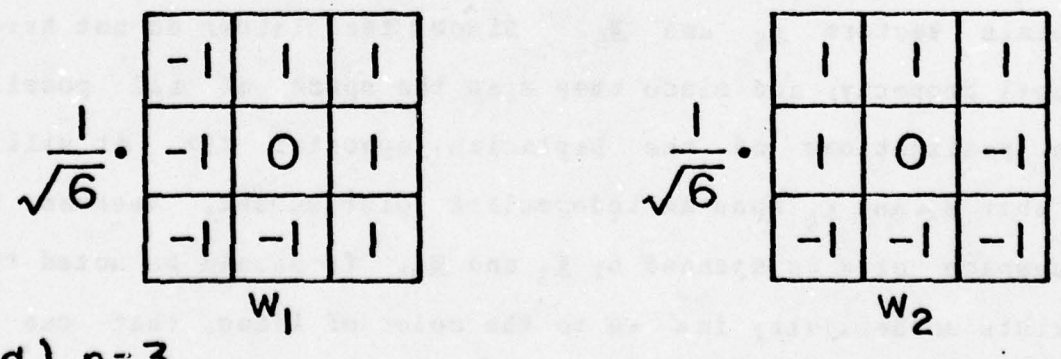
a) n = 2; two possible bases for $\mathcal{E}$



b) n = 3



c) n = 3



d) n = 3

Figure 2.2-2. Possible orthogonal basis vectors spanning $\mathcal{E}$ for n=2 and n=3. (The factors to the left of the templates render the basis vectors orthonormal. Often not necessary).

operator onto the basis vectors of figure 2. A more detailed study of existing operators can be made later, by observing that the templates used are sometimes a linear combination of $\underline{W}_1$, $\underline{W}_2$ and other non-edge basis vectors to be discussed. Figure 4 shows two convolutions of a test image (figure 4a) with the basis vectors $\underline{W}_1$ and $\underline{W}_2$ of figures 4b and 4c. Edge detection could be done at this stage by thresholding the magnitude of the projection of $\underline{B}$ onto $\mathcal{B}$ shown in figure 4d.

Other Features: The edge detection operator can now be extended to the detection of other image features of interest. An ideal line element is defined in the continuous image domain as a straight strip of width approximately equal to one sampling interval, passing through the center of A, and of different average brightness $\bar{b}_1$ than its surround $\bar{b}_2$ (figure 1c). The ideal line element is characterized by its direction θ_E, its magnitude $|\bar{b}_1-\bar{b}_2|$ and its color defined as $\text{sgn}(\bar{b}_1-\bar{b}_2)$. The discrete representation of all ideal line elements forms another subset $\mathcal{L}$ of $\mathcal{B}$, of dimension two, which shares only the zero element with $\mathcal{S}$. An intuitive set of templates for $\mathcal{L}$ shown in figure 5a can be reduced to two line basis vectors $\underline{W}_3$ and $\underline{W}_4$ and two point basis vectors $\underline{W}_5$ and $\underline{W}_6$. Since the latter do not have a directional property, and since they span the space of all possible discrete realizations of the Laplacian operator <9>, it will be assumed that $\underline{W}_5$ and $\underline{W}_6$ span an independant point subset, whereas the line subspace of $\mathcal{L}$ is spanned by $\underline{W}_3$ and $\underline{W}_4$. It should be noted that there exists an ambiguity in $\mathcal{L}$ as to the color of lines, that can be resolved by examination of the components in the point subspace.

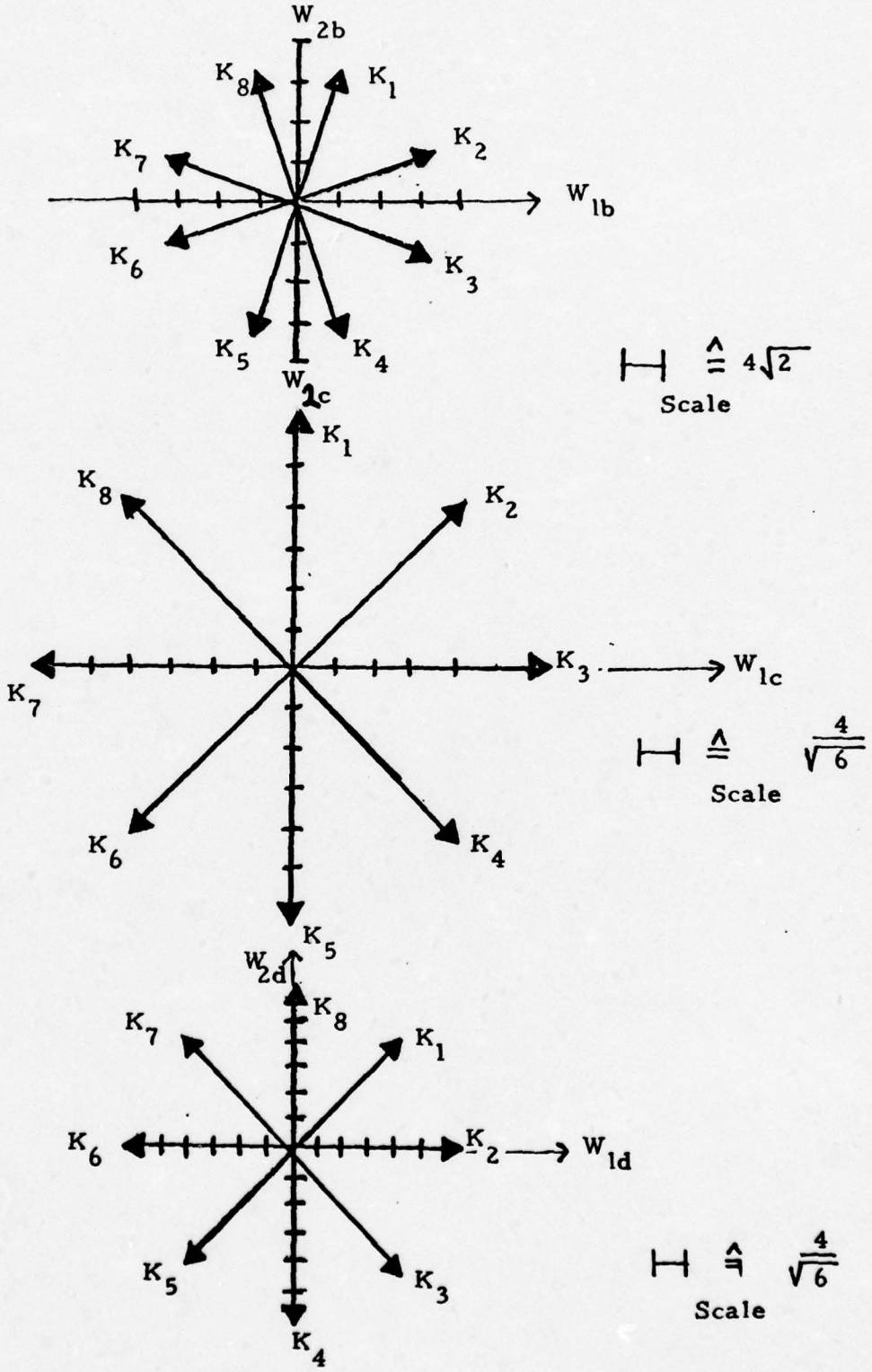
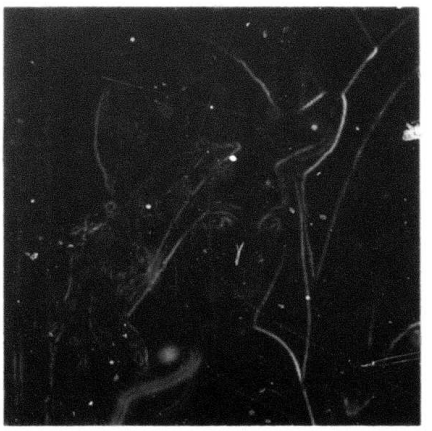


Figure 2.2-3. Projections of the Kirsch templates onto the edge subspaces spanned by the bases of figure 2b, c, d respectively.

(a) Original

(b) Convolution with W_1 (scaled)

(c) Convolution with W_2 (scaled)



(d) Magnitude of projection of original image onto the edge subspace $\mathcal{E}$

Figure 2.2-4. Original image, projections onto the basis vectors of the edge subspace, and magnitude of projection onto the same subspace.

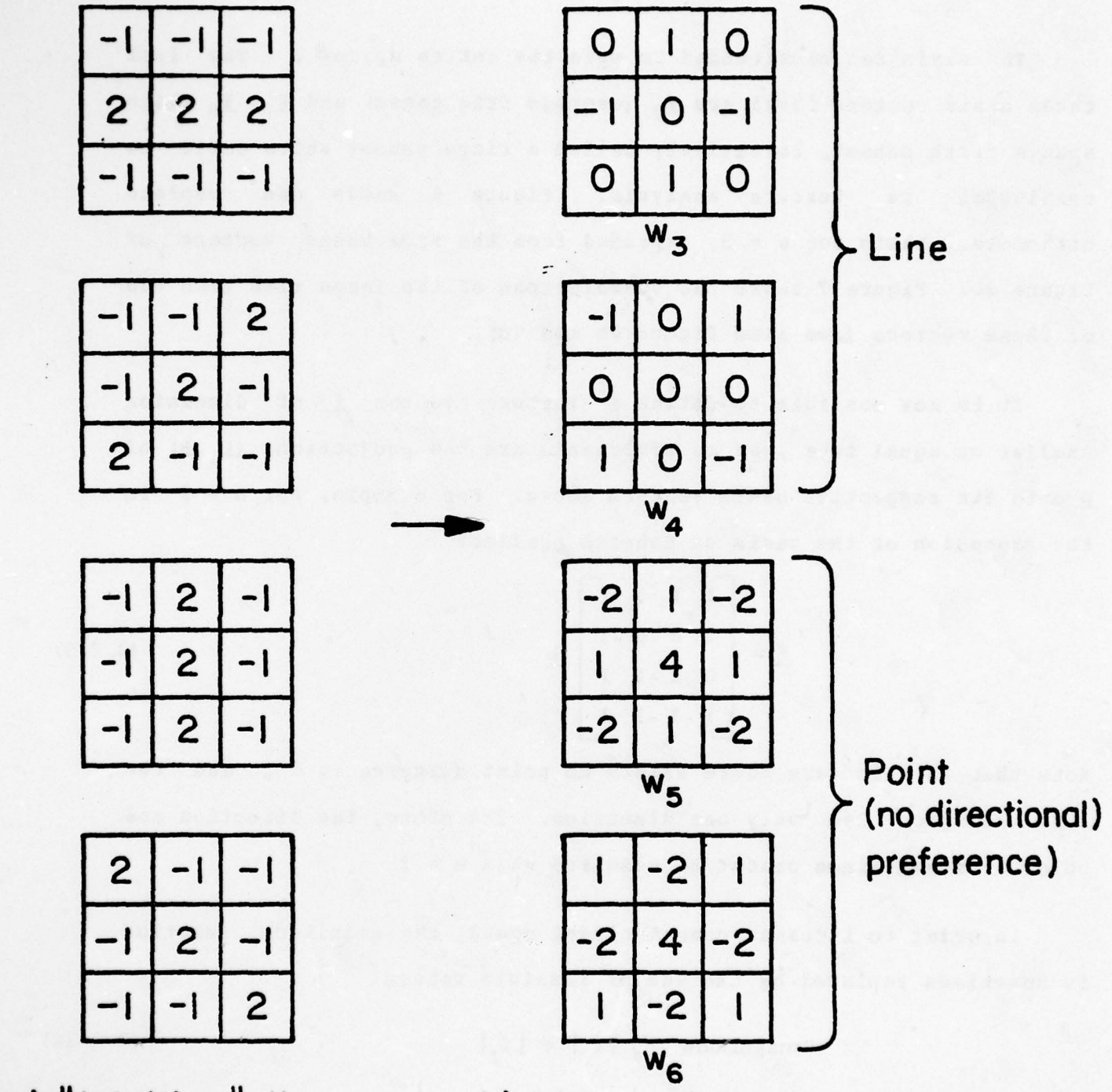


Figure 2.2-5. "Intuitive" line detection templates and decomposition into line and point basis vectors.

The basis can be extended to span the entire space $\mathcal{B}$. The last three basis vectors found are $\underline{W}_0$ (average brightness) and $\underline{W}_7$, $\underline{W}_8$ which span a fifth subset, tentatively called a ridge subset which could be meaningful in texture analysis. Figure 6 shows one complete orthonormal basis for n = 3, expanded from the edge basis vectors of figure 2b. Figure 7 shows the convolutions of the image with each one of these vectors (see also figure 4b and 4c).

It is now possible to define a feature vector $\underline{f}$ of dimension smaller or equal to n^2, whose components are the projections $(\underline{W}_i, \underline{B})$ of $\underline{B}$ onto the respective basis vectors above. For example, for n = 2 in the expansion of the basis of Roberts gradient

$$\underline{f} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \underline{b} \qquad (2.2\text{-}9)$$

Note that in this case there exists no point subspace in $\mathcal{B}$, and the line subspace has only one dimension. Therefore, the direction and color of ideal lines cannot be resolved when n = 2.

In order to increase computational speed, the magnitude function is sometimes replaced by the sum of absolute values

$$\text{"magnitude"} = |f_1| + |f_2| \qquad (2.2\text{-}10)$$

which can be in error by up to 41%. It has been suggested that such errors do not seriously affect the performance of edge detectors, but it is easy to see that they give preference to certain orientations.

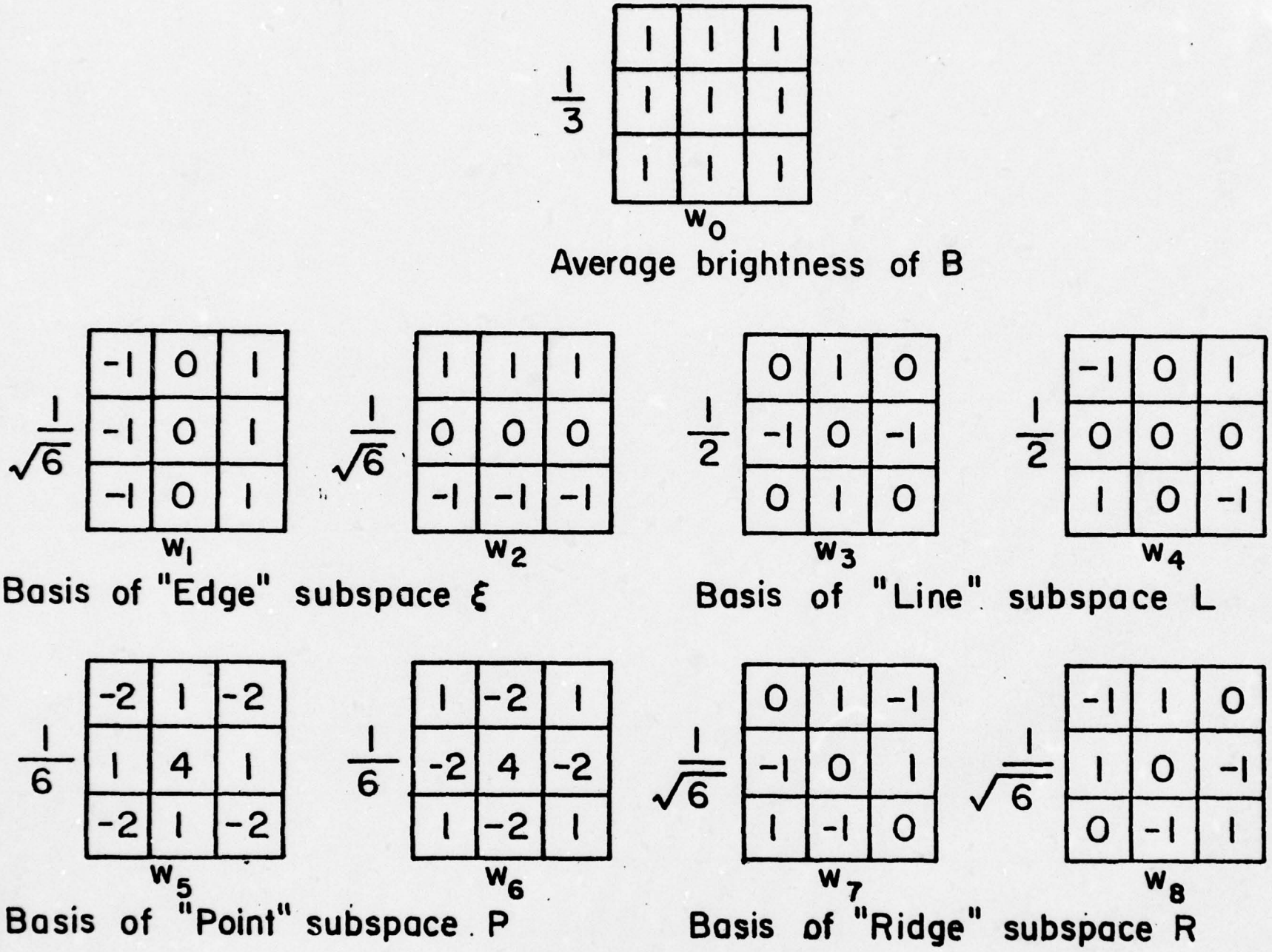
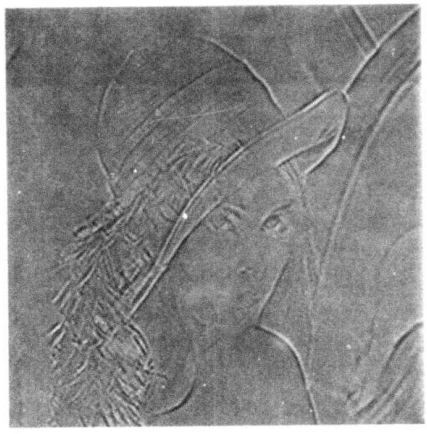


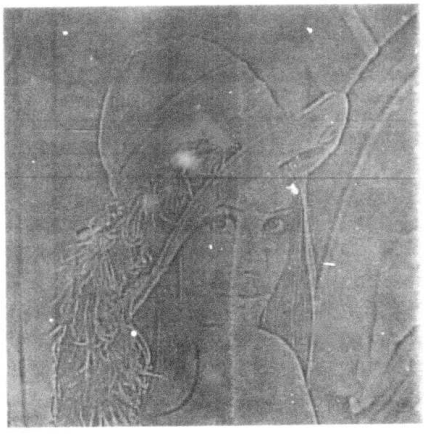
Figure 2.2-6. One possible orthogonal basis for B (orthonormal if the factors to the left of the basis vectors are included).

(a) B ⊛ W_3

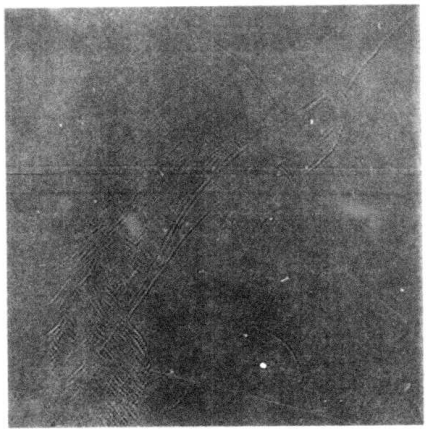
(b) B ⊛ W_4



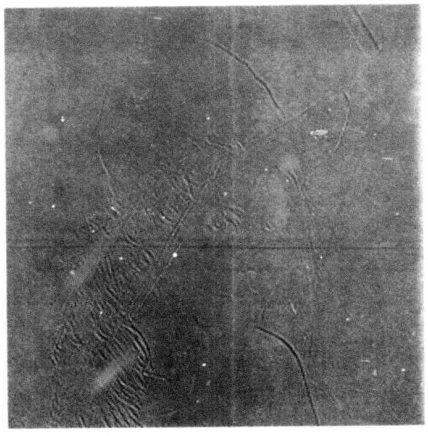
(c) B ⊛ W_5



(d) B ⊛ W_6



(e) B ⊛ W_7



(f) B ⊛ W_8

Figure 2.2-7. Convolutions of original image with basis vectors, W_3, W_4, W_5, W_6, W_7, W_8 (W_1 and W_2. See figure 4b and 4c)

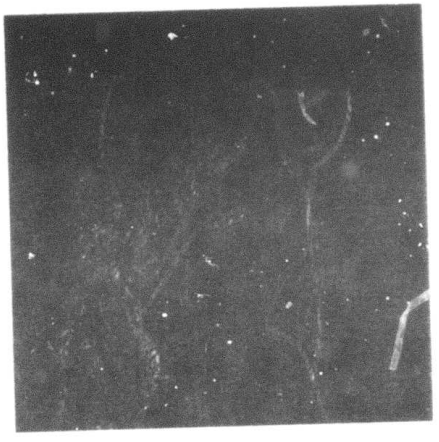
Extensions: Several properties of the feature transform are now being actively explored. Of particular interest are the magnitudes of the projections of $\underline{B}$ onto the subspaces $\mathcal{S}$, $\mathcal{L}$, $\mathcal{O}$, $\mathcal{R}$. Figures 4d and 8a, 8b, 8c show these magnitudes computed as

$$\text{"magnitude"} = (f_i^2 + f_j^2)^{\frac{1}{2}} \qquad (2.2\text{-}11)$$

for (i, j) = (1,2), (3,4), (5,6), (7,8). Since these magnitudes are indicative of the respective contributions of ideal edge, point and ridge elements present within an actual image segment, it appears that a measure of confidence can be derived as a function that compares the magnitudes of the projections of A onto the edge subspace $\mathcal{S}$ and the non-edge subspaces $\mathcal{L}$, $\mathcal{O}$, $\mathcal{R}$, respectively. Conversely, it is conceivable that the threshold decision can be based upon the relative magnitudes of the components, instead of using a fixed threshold value. This is illustrated conceptually in figure 1d, where the vector $\underline{b}_1$ would be classified as an edge but not vector $\underline{b}_2$, although $\underline{b}_2$ has a larger projection onto the edge subspace, and would be detected as an edge by conventional enhancement/threshold edge detection algorithms.

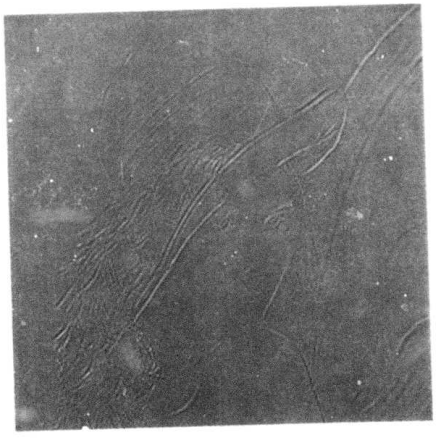
Next it is observed that the magnitude images of figures 4a, 8a, 8b, and 8c reveal a complementary nature of the edge and, to a lesser extent, ridge versus line, and point subspaces. An intuitive argument for this observation is made by considering a non-ideal line of width larger than one sampling interval. It is easy to see that edge points

(a) $\left[(B \circledcirc W_3)^2 + (B \circledcirc W_4)^2\right]^{-\frac{1}{2}}$
(Line subspace)



(b) $\left[(B \circledcirc W_5)^2 + (B \circledcirc W_6^2)\right]^{-\frac{1}{2}}$
(Point subspace)

(c) $\left[B \circledcirc W_7^{?} + (B \circledcirc W_8)^2\right]^{-\frac{1}{2}}$
(Ridge subspace)



(d) Convolution of original image with W_0(low-pass)

Figure 2.2-8. Magnitudes of projections of original image (fig. 4a) onto the three subspaces L, P, R resp. (see also figure 4d)

will be detected on each side of the line. Conversely, lines and points are detected on each side of non-ideal edges, a fact that is less apparent intuitively, but can be verified readily in the pictures of figures 4d, 8a, 8b, and 8c, or by manual computation. This observation suggests that contextual information supplied by the remaining elements of the feature vector $\underline{f}$ can be used to eliminate edges due to noise, as well as to thin and link edges or lines, since any edge point is surrounded by line points of similar orientation, as well as point elements, and vice-versa.

Conclusions: It is believed that the present feature transform is a powerful tool for image analysis, and enables an objective analytical evaluation of edge detectors. In addition to the extensions mentioned above, it is also pointed out that the present approach can be extended immediately for $n > 3$. However, it appears that arbitrary large values of n are not very desirable in terms of computational load, because the contextual techniques outlined could give similar results at a lower cost.

References

1. L.G. Roberts, "Machine Perception of Three-Dimensional Solids," in Optical and Electro-Optical Information Processing, J.T. Tippett, et.al., eds., MIT Press, 1965.

2. R. Kirsch, "Computer Determination of the Constituent Structure of Biological Images," Computers and Biomedical Research, Vol. 4, 1971, pp. 315-328.

3. R. O. Duda and P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1971.

4. J.M.S. Prewitt, "Object Enhancement and Extraction," in Picture Processing and Psychopictorics, B.S. Lipkin and A. Rosenfeld, eds., Academic Press, New York, 1970, pp. 75-149.

5. T.S. Huang, Purdue University, private communication, 1975.

6. G.S. Robinson, "Detection and Coding of Edges using Directional Masks," see present report.

7. M.S. Hueckel, "An Operator which Locates Edges in Digitized Pictures," J. Association for Computing Machinery, Vol. 18, No. 1, January 1971, pp. 113-125.

8. M.S. Hueckel, "A Local Visual Operator which Recognizes Edges and Lines," J. Association for Computing Machinery, Vol. 20, October 1973, pp. 634-647.

9. Z. Meiri (Bell Laboratories), private communication, January 1976.

2.3 Detection and Coding of Edges using Directional Masks

Guner S. Robinson

A new image coding system is described, which combines the detection and coding of visually significant edges in natural images.

Edges are defined as amplitude discontinuities between different regions of an image. The edge detection system makes use of 3 x 3 masks, which are well suited for digital implementation. Edge angles are quantized to eight equally spaced directions, suitable for chain coding of contours. Use of an edge direction map improves the simple thresholding of gradient modulus images. The concept of local connectivity of the edge direction map is useful in improving the performance of edge operators such as Kirsch and Sobel operators. The concepts of an "edge activity index" and a "locally adaptive threshold" improve performance even further.

Directional Mask Operators: Two dimensional mask operators are commonly used for filtering and enhancement <1,2>. Two dimensional discrete differentiation can be performed by convolving the original image with the compass gradient masks shown in figure 1. The compass names indicate the slope direction of maximum response, e.g. the North gradient mask produces a maximum output for vertical luminance changes, i.e. for horizontal edges. The direction of this horizontal edge could be from left to right or from right to left. In the following, edge directions corresponding to the eight compass gradients are determined such that the bright side of the edge is always to the left as one moves in the direction of the edge. The directions of the edges corresponding to the eight compass gradient masks are also shown in figure 1. The numbers 0,1,....,7 are used for the eight principal directions in a 3 x 3 grid, as shown in figure 2. This notation is used in order to be consistent with the Freeman chain

Direction of Edge	Direction of Gradient	Prewitt Masks	Kirsch Masks	Three-level Simple Masks	Five-level Simple Masks
0	North	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$
1	North-West	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -2 \end{bmatrix}$
2	West	$\begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & -1 \end{bmatrix}$	$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$
3	South-West	$\begin{bmatrix} 1 & -1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix}$
4	South	$\begin{bmatrix} -1 & -1 & -1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$
5	South-East	$\begin{bmatrix} -1 & -1 & 1 \\ -1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$
6	East	$\begin{bmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$
7	North-East	$\begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix}$	$\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix}$

Figure 2.3-1. Examples of Compass Gradient Masks.

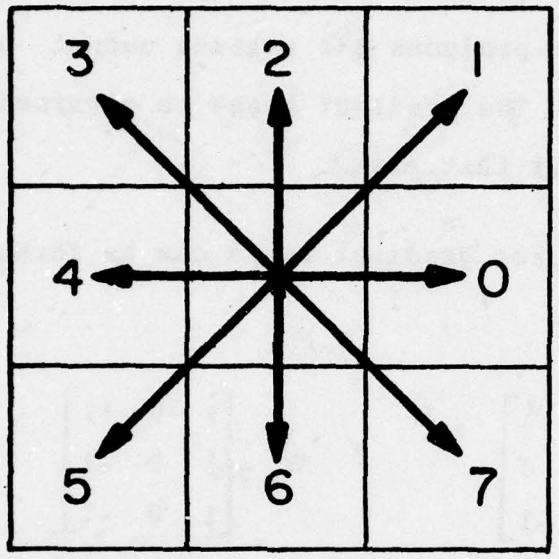


Figure 2.3-2. The eight principal directions on a 3 X 3 grid.

coding scheme <3, p.249> for subsequent chain coding of extracted boundaries.

The Kirsch operator <18> can be thought as taking the output of one of the eight compass gradient masks shown in the second column of figure 1. The mask which produces the maximum output determines the direction of the edge. The gradient image is obtained by taking the magnitude of the output of that mask.

A simple set of compass gradient masks can be formed by rotating the differentiation masks

$$W_x = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & -1 & -1 \end{bmatrix} \qquad W_y = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

given by Prewitt <1,p.108) , as shown in the third column of figure 1. These directional masks are called three-level simple masks in order to distinguish them from the five-level directional masks which will be given later. The two orthogonal masks, W_x and W_y, which measure the gradients in the North and West directions, approximate the partial derivatives in the x-direction and y-direction, respectively. Application of W_x and W_y to an image results in spatial differentiation in two orthogonal directions. The gradient magnitude and direction can then be obtained by taking the magnitude and direction cosines at each point. An analog gradient picture results when the gradient magnitudes are displayed as gray values.

The Sobel operator (5,p.271) can be thought as a combination of two gradiant masks, one in the North and the other in the East direction. The outputs of these two masks approximate the partial derivative in their respective directions. A gradient modulus image is obtained by taking the magnitude of the two orthogonal mask outputs. The direction of the edge can also be determined from the outputs of these two masks.

In the following a set of five-level simple directional masks are considered for obtaining the analog gradient image and the edge direction in a simple manner. These masks, called five-level simple masks, assume five integer weights between -2 and +2. The five-level simple compass gradient masks are shown in the fourth column of figure 1. The advantages of the particular choice of weights are that:

(a) The two orthogonal masks

$$M_x = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \qquad M_y = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

approximate the partial derivatives in the x-direction and y-direction, respectively.

(b) The zero weights in the center of the masks result in ignoring the transient line where an edge might occur.

(c) Computing the output of the first four masks is enough to obtain both the analog gradient image and the edge direction map.

For example, if the result of the North mask is positive, then the edge direction is "0" and if it is negative, then the direction is "4".

(d) A defocusing operation can be performed on an image by using a low-pass mask such as

$$M_0 = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

which has integer weights similar to M_x and M_y. This type of defocusing mask could be used for obtaining local threshold values in edge detection by simple directional masks. Mask M_0 has a better sidelobe structure than the low-pass counterpart W_0 of W_x and W_y

$$W_0 = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

which makes a simple averaging operation of nine pixels in a 3 x 3 grid of an image. It has been verified that M_0, M_x and M_y are members of a seven-level orthogonal set of masks similar to the set given by Frei and Chen <6>.

(e) Five-level simple directional masks yield a higher gradient amplitude in the diagonal direction than the horizontal or vertical direction. It is known that the visual acuity in the

horizontal and vertical directions is superior to the visual acuity in the diagonal directions for a large range of spatial frequencies, and it is conjectured that the basis of this phenomenon lies in the retina and higher order visual pathways <7,8>. The directional masks of figure 1 can be compared in the x-y domain defined by the partial derivative operators W_x and W_y. For example, the direction and normalized amplitude of the edge corresponding to the Northwest diagonal mask is described below:

Type of Mask	Angle (Degrees)	Amplitude
Prewitt	45	0.943
Kirsch	45	0.943
Three-level simple	45	0.943
Five-level simple	45	1.067

This table shows that all four types of directional masks are nonisotropic. However only the five-level simple masks compensate for the lower visual acuity in the diagonal directions by weighting the diagonal elements higher in the corresponding masks.

(f) The structure and integer weights of the simple masks make them especially suitable for fast computation of gradient magnitudes and directions digitally.

Edge Detection System: The block diagram of the proposed edge detection system is shown in figure 3. The application of the first four simple masks to a 3 x 3 grid surrounding a picture element gives the gradient magnitude and direction. The gradient picture is obtained by taking the maximum gradient value at each point. The mask

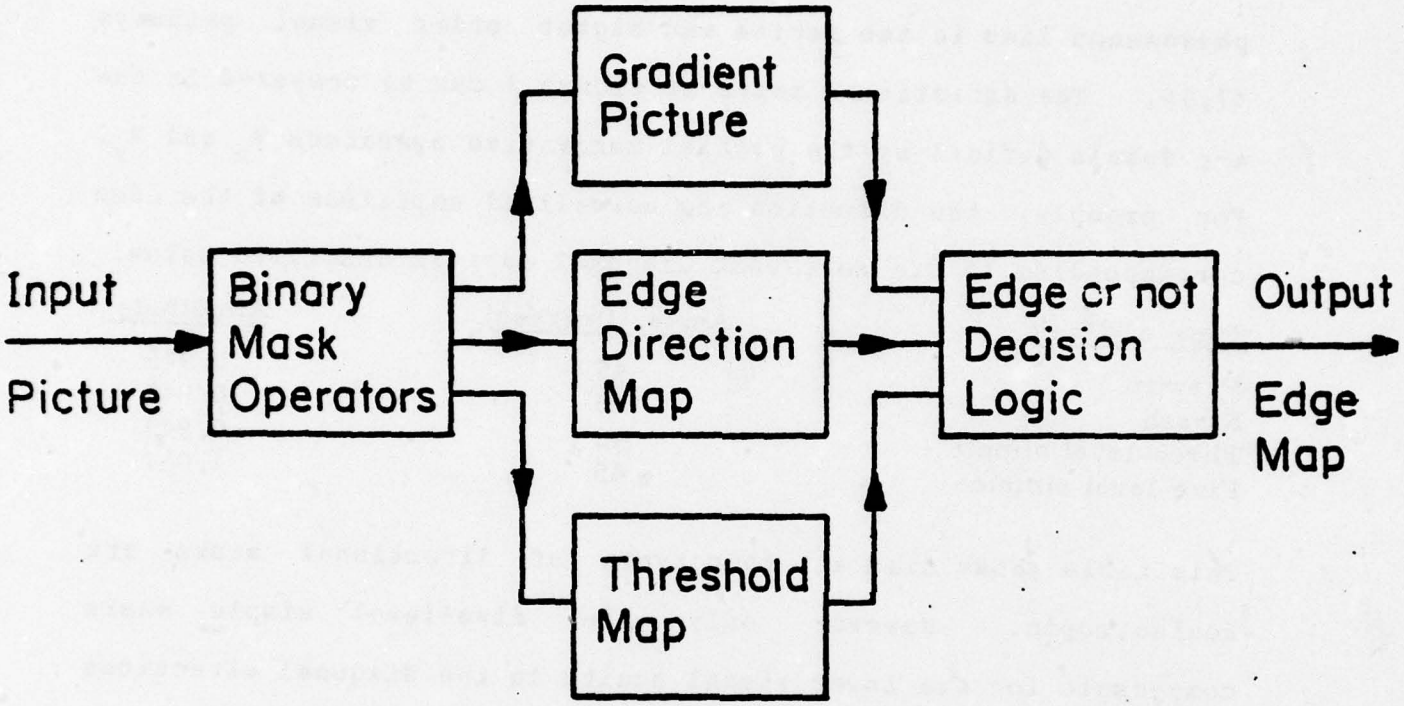


Figure 2.3-3. Block diagram of the proposed edge detection system.

which yields the maximum gradient value determines the direction of the edge. The edge map is a two dimensional array of numbers which range between 0 and 7. The threshold map can be obtained in various ways. In general it is a two dimensional array which is used in determining whether the gradient value is large enough to accept or reject the presence of an edge point. The presence of an edge is determined by examining the gradient values, edge direction map and the threshold map. If the edges in a 3 x 3 grid surrounding a point satisfy the local connectivity conditions depicted in figure 4, and if they are above the threshold set by the threshold map, then it is determined that there is an edge point. A binary edge map is thus generated at the output.

Figure 5 shows some examples of analog gradient images. Figure 6 shows the importance of the three basic blocks of the edge detection system for extracting the edges in a picture of a toy tank. The comparison of the results with the results of the Kirsch and Sobel operators shows that the use of the local connectivity and the locally adaptive threshold is necessary to improve the extraction of visually significiant edges. Figure 7 shows the binary edge maps of two other images using a fixed threshold equal to the average intensity of the corresponding analog gradient images.

Edge Activity Index and Locally Adaptive Threshold: Examination of the analog gradient images shown in figures 5, 6 and 7 shows that, a suitable fixed threshold value would produce the edges directly. However, it should be observed that some edges are very faint, while

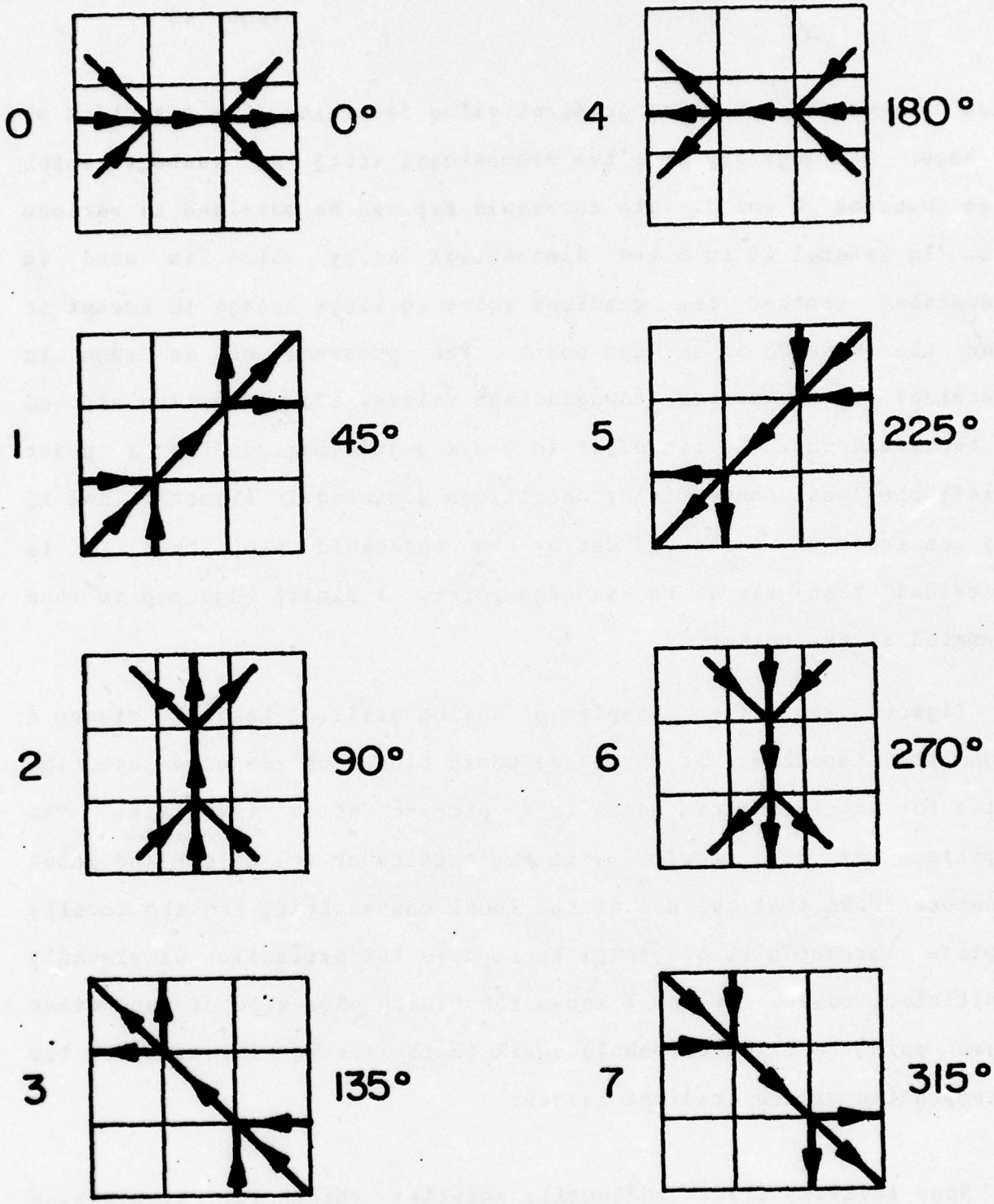
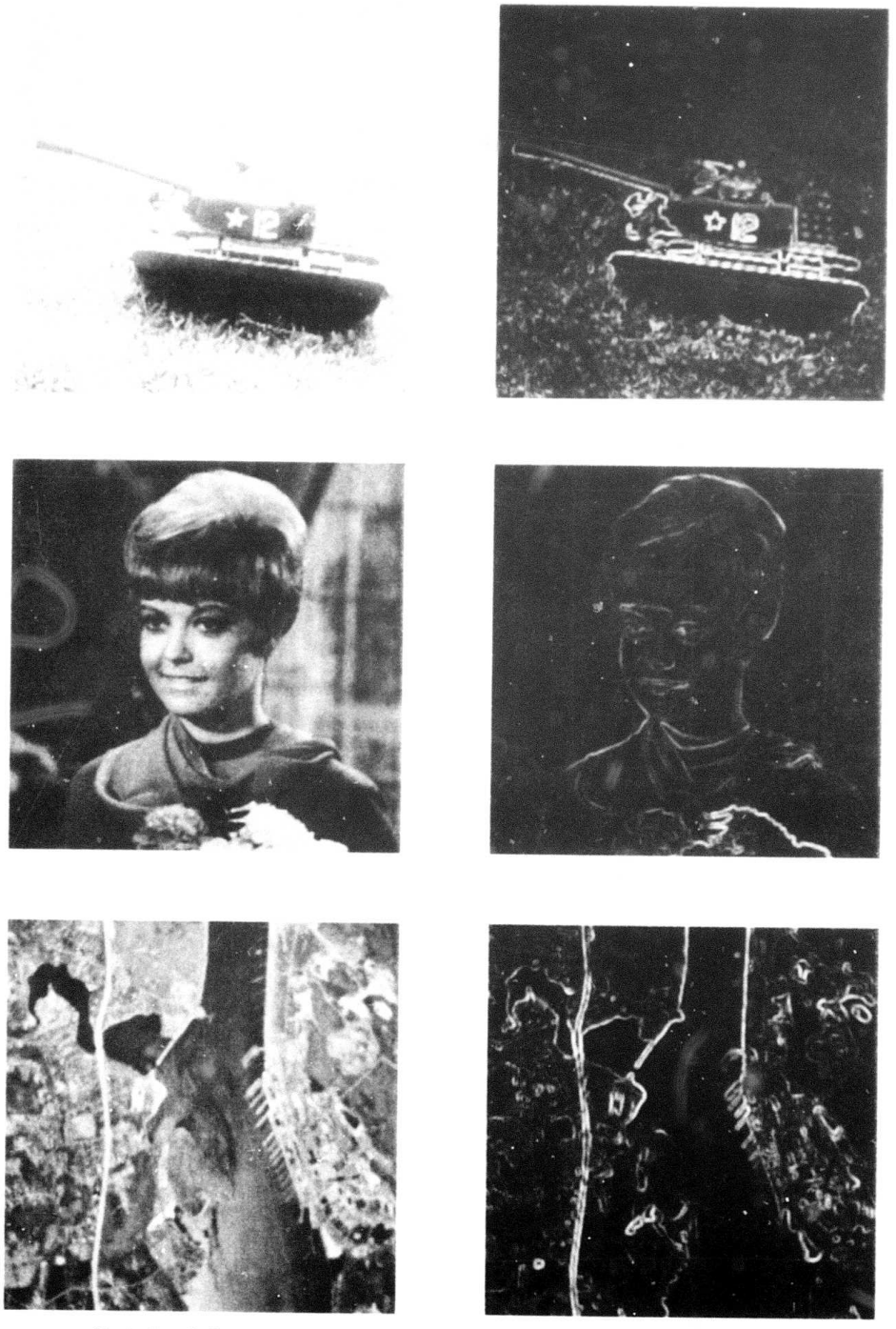


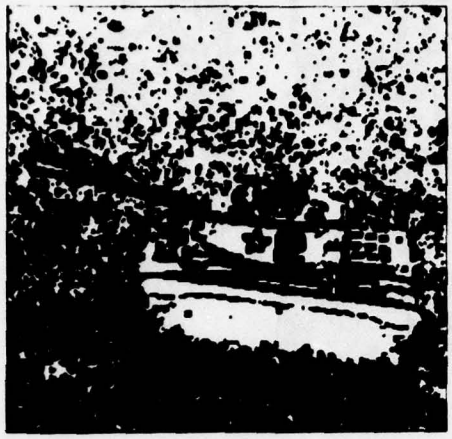
Figure 2.3-4. Local connectivity of edges on a 3 X 3 grid.



Original Images

Analog Gradient Images

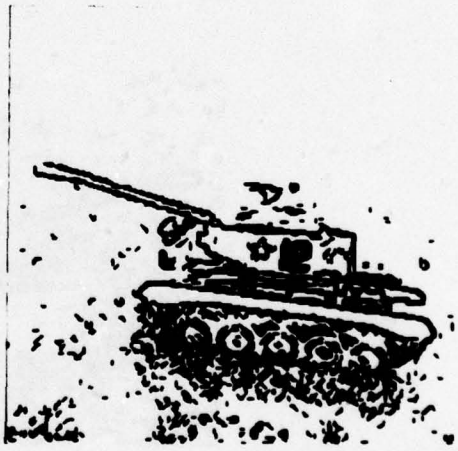
Figure 2.3-5. Some examples of Analog Gradient Images.



(a) Fixed threshold
No connectivity test

(b) Adaptive threshold
No connectivity test

(c) Fixed threshold
and connectivity test



(d) Local threshold
and connectivity test

Figure 2.3-6. Performance of the Edge Detection System using five-level simple directional masks.

(a)

(b)

(c)

Figure 2.3-7. Performance of the Edge Detection system using directional Kirsch masks and a fixed threshold.

some are stronger. If the threshold is too low, too many edge points are obtained and if the threshold is too high, then some significiant edges are lost. This suggest that improvement of the analog gradient image and the use of a local threshold is necessary.

The improvement of the analog gradient image can be obtained by the use of the edge activity index EAI defined as the ratio of the maximum gradient magnitude at an image point to the average magnitude of gradients in eight compass directions. If the eight directional gradient values at a pixel (i,j) are $y_0, y_1, \ldots, y_7$, then

$$EAI = \frac{\text{Max}\{|y_k|,\ k = 0,1,\ldots,7\}}{\sqrt{\sum_{k=0}^{7} y_k^2}}$$

could serve as an objective measure of busyness on a 3 x 3 grid of an image. If the EAI is greater than a threshold, i.e. if the edge activity is considerably superior in the direction of the maximum gradient, then the maximum gradient value is taken, otherwise the gradient value is set to zero. This operation results in a sharper histogram for the analog gradient image. It has been experimentally verified that the gradient image has a Rayleigh-like distribution. The histogram is very peaked for a large class of natural grey tone images. The shape of the histogram does not change for various types of directional masks. A simple value for a fixed threshold value in edge detection is the mean intensity of the gradient image. In the examples edge maps generated with fixed threshold are obtained using the mean intensity as the threshold on the corresponding gradient

image, in addition to the local connectivity tests on the corresponding edge direction map. Another choice of the fixed threshold can be obtained by integrating the histogram starting from the high intensities, since the edges have high intensity values in the gradient image. More and more edge points can be added in order to attain a fixed number of edge points in the resultant binary edge map.

A locally adaptive threshold has been obtained by comparing the analog gradient image with a blurred version of the original image, which is obtained by a low-pass operation on the image. The particular low pass operation can be performed by the mask M_0. Thus the locally adaptive threshold has been defined as

$$\text{LAT} = \frac{\text{Max}\{|y_k|,\ k = 0,1,\ldots,7\}}{\text{Output of the lowpass filter } M_0 \text{ at pixel } (i,j)}$$

Figure 6 shows the result of the use of a locally adaptive threshold on the toy tank.

Coding: Preliminary results show that the transmission of edges extracted using compass gradient masks may be enough for some low level applications. For more sophisticated applications, a simple grey tone image can be generated by transmitting a reduced size version of the original image. The size reduction is obtained by successive size halving of the input image by averaging four neighboring picture elements on two lines. The reduced size image

(32 x 32 or 64 x 64 for example) can be coded by standard two-dimensional coding techniques. At the receiver a full size blurred image is formed by successive linear interpolation. The superposition of the edges with this gray tone image is expected to give more visual information than just an edge image, while providing a good data rate reduction. Investigation of efficient coding methods as well as receiver interpolation of edge intensities is presently under way.

References

1. J.M.S. Prewitt, "Object Enhancement and Extraction," in Picture Processing and Psychopictorics, B.S. Lipkin and A. Rosenfeld, eds., Academic Press, New York, 1970, pp. 75-149.

2. W.K. Pratt, Digital Image Processing, Wiley, New York, 1977.

3. H. Freeman, "Boundary Encoding and Processing," in Picture Processing and Psychopictorics, B.S. Lipkin and A. Rosenfeld, eds., Academic Press, 1970, pp. 241-266.

4. R. Kirsch, "Computer Determination of the Constituent Structure of Biological Images, Computers and Biomedical Research, Vol. 4, 1971, pp. 315-328.

5. R.O. Duda and P.E. Hart, Pattern Classification and Scene Analysis, John Wiley and Sons, New York, 1973.

6. W. Frei and C-C. Chen, "A New Class of Edge and Feature

Detection Operators," in this report.

7. H. Leibowitz, "Some Observations and Theory of the Variation of Visual Acuity with the Orientation of the Test Objects," Journal Optical Society of America, Vol. 43, October 1953, pp. 902-905.

8. D.E. Mitchell, R.D. Freeman and G. Westheimer, "Effect of Orientation on the Modulation Sensitivity for Interference Fringes on the Retina," Journal of the Optical Society of America, Vol. 57, February 1967, pp. 246-249.

2.4 Color Edge Detection

Guner S. Robinson

Color images may be described quantitatively at each pixel by a set of three tristimulus values which are proportional to the amount of red, green, and blue primary lights required to match the pixel color. The red, green and blue components may be transformed by some linear or nonlinear invertible function to quantities which are more suitable for efficient transmission, color perception, redundancy reduction, edge extraction, etc. <1>. The N.T.S.C. transmission color coordinates Y, I and Q are obtained from the tristimulus color components by a linear transformation. The Y component, which is a weighted sum of the tristimulus values, is a measure of the luminance of the color. The I and Q components jointly describe the hue and saturation of the image.

Several definitions of color edges are possible, depending on the choice of the color coordinate system. In monochrome images an edge is defined as a discontinuity in the brightness function. A generalization of this definition would be possible by defining a discontinuity in the three-dimensional vector space defined by R, G, B or other color coordinates obtained from R, G, B by a linear or nonlinear transformation. For example an edge in a color image could be defined as a discontinuity in the luminance component, ignoring the discontinuities in I and Q in regions of constant luminance. Then the color edge detection problem reduces to the case of monochrome edge detection. It is known that the red, green and blue components of an image are highly correlated. The transformation which gives the N.T.S.C. color transmission coordinates partially removes this correlation., i.e. the Y, I, Q components are still partially correlated. A second approach to the color edge detection problem involves use of the Karhunen-Loeve transformation to decorrelate the color components. Edge detection is then performed on the individual principal components of the particular image. Still another approach is to perform the edge detection operation on each component of the color space and combine the edges suitably, depending on the definition of the color coordinates. This approach then reduces the problem of color edge detection to repetition of the monochrome edge detection scheme for all three components. The U V W and L a b color coordinate systems could also be used for color edge detection since they posses a metric color difference sensitivity to a good approximation, while the R, G, B space does not <1>.

In order to substantiate the above ideas and conjectures, various color components have been computed using the red, green and blue components of a 256 x 256 color image shown in figure 1. The sets of color components that were considered are: principal components $P_1P_2P_3$, Y I Q, L a b, and $G_1G_2G_3$. The components G_1, G_2 and G_3 are obtained using the model of the human visual system developed by Frei <2>. Figures 2 to 5 show the various components, which are scaled to have the same range of values, 0-255, for display. The percentage energy each component represents is given in Table 1. The normalized covariance matrix of the red, green and blue components of the girl picture is

$$C_{R,G,B} = \begin{bmatrix} 1.000 & 0.221 & 0.172 \\ 0.221 & 1.000 & 0.228 \\ 0.172 & 0.228 & 1.000 \end{bmatrix}$$

The principal components have been obtained using the above covariance matrix.

Experimental Results for Color Edge Detection: Experimental results in the edge detection are confined to the directional mask edge extraction algorithm <3>. The directional mask edge extraction algorithm was applied to each component separately. The binary edge maps were generated by using local connectivity and a fixed threshold. The threshold in each case was the mean intensity of the analog gradient picture. The color edge was obtained simply by comparing the

R

G

B

Figure 2.4-1. Red, Green and Blue Components of a Color Image.

Color Component	Range of Values (min-max)		Energy Content (%)
R	56.0	255.0	37.6
G	3.0	245.0	43.8
B	43.0	219.0	18.6
P_1	59.2	400.0	90.6
P_2	-91.4	60.4	8.4
P_3	24.7	101.2	0.8
Y	25.9	241.6	89.2
I	14.9	88.0	10.3
Q	25.8	66.0	0.5
L	38.2	98.0	82.1
a	- 1.4	61.0	16.8
b	-30.9	34.3	1.1
G_1	70.0	118.0	96.9
G_2	-17.5	1.76	2.9
G_3	- 2.89	6.05	0.2

Table 1. The range of values and the percentage energy content of various color components.

P_1

P_2

P_3

Figure 2.4-2. Principal Components of a Color Image.

Y

I

Q

Figure 2.4-3. N.T.S.C. Color Transmission Components of a Color Image.

L

a

b

Figure 2.4-4. L, a, b, Components of a Color Image.

G_1

G_2

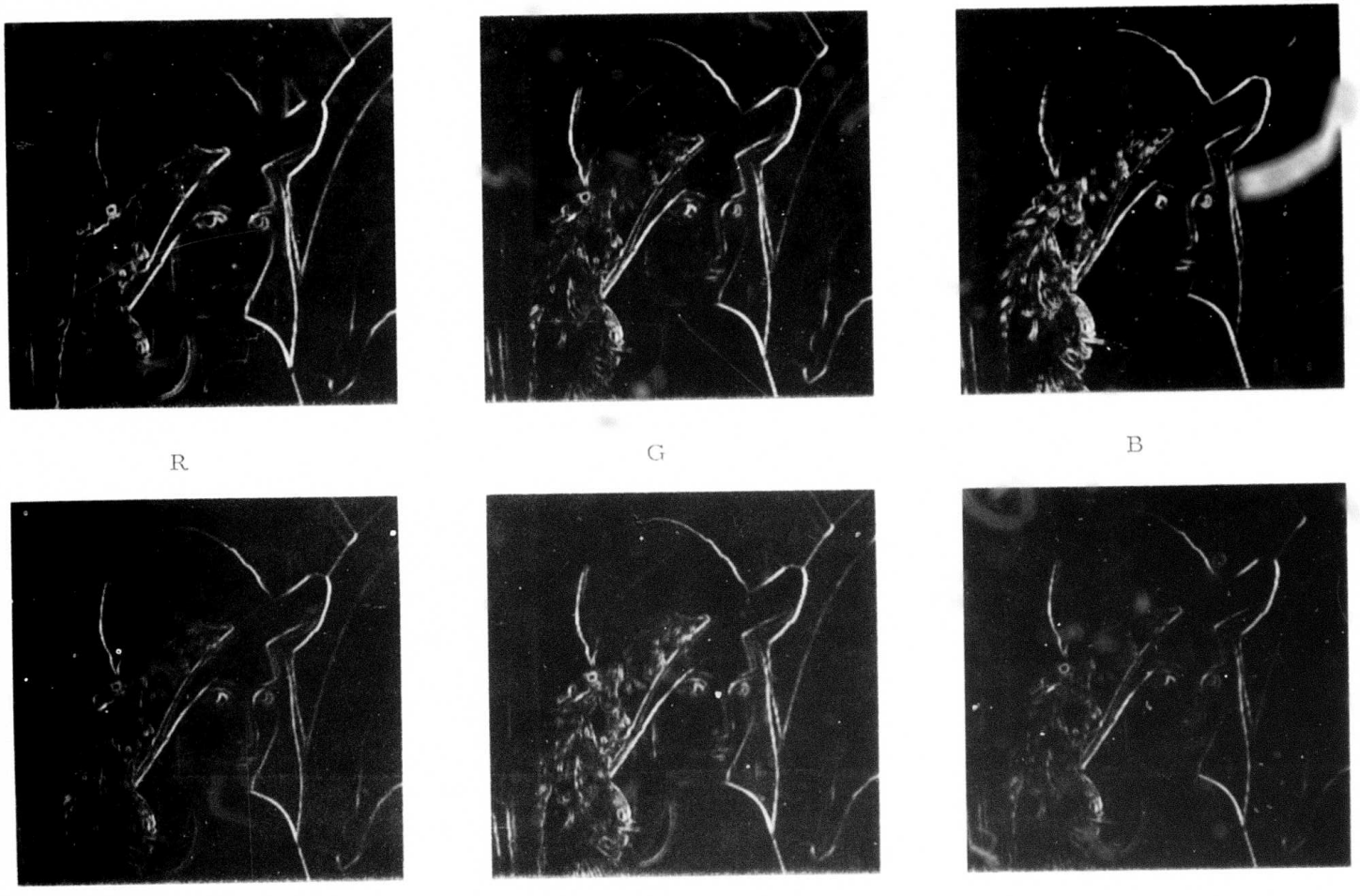
G_3

Figure 2.4-5. Components of a Color Image derived using a Model of the Human Visual System.

twenty four gradient values in three components and eight directions. The gradient that gives the maximum value at a pixel thus determines the direction of the edge, whether it is due to a discontinuity in the first, second or the third color component. Figure 6 shows some examples of gradient images of color components and figures 7 and 8 illustrate examples of the color edge extraction algorithm.

Comparison of Color Coordinates: Visual examination of the color components shown in figures 2 to 5 shows that the conjectures that were made in the selection of five sets of color coordinates are basically correct. The percentage energy each component represents is, in general, indicative of the edge activity in that component. The principal components do not have a physical meaning in terms of the perception of colors. However, the first principal component is a weighted sum of the tristimulus values, and therefore, is a measure of the brightness of the image similar to the Y and L components. The first principal component has the most energy content, which is most likely due to the sharper amplitude discontinuities. In the case of the L a b components, "L" reflects the brightness of the image, the values of "a" and "b" reflect the redness-greenness and yellowness-blueness of the image, respectively.

Objective measures in addition to the subjective evaluation are necessary for comparing color components for suitability for edge extraction. The edge activity index EAI could be used as an objective measure of edge busyness in an image <3>. If the edge direction angles at a point (i,j) are quantized to eight compass directions, the



R G B

R, G, B combined L, a, b combined P_1

Figure 2.4-6. Examples of Color Image Gradients

Figure 2.4-7. Edges detected in the RGB color space and comparison with edge maps obtained in the R, G and B components separately, using directional Kirsch masks.

$P_1P_2P_3$

YIQ

Lab

$G_1G_2G_3$

Figure 2.4-8. Edges detected in various color coordinate systems using directional Kirsch masks.

EAI provides a measure of the superiority of the maximum gradient direction to the gradients in the other seven compass directions. If the values of EAI are computed on the edge positions throughout the image, then the mean and the variance of the EAI is expected to be indicative of the edge activity in that image. The study of the edge activity index as an objective measure in comparing the above five sets of color components is presently under way.

References

1. W.K. Pratt, Digital Image Processing, Wiley, New York, 1977.

2. W. Frei, "A Quantitative Model of Color Vision," USCIPI Report 540, University of Southern California, pp. 69-83.

3. G.S. Robinson, "Detection and Coding of Edges using Directional Masks," in this report.

2.5 Hueckel Color Edge Detector

Ram Nevatia

A common operation performed on images is that of edge detection. Edges are defined as a discontinuity in some image attribute, usually luminance. Boundaries between two objects are often accompanied by a discontinuity in luminance. In color images different objects may have different colors and it seems useful to look for color discontinuities or color edges. This paper provides a definition of a color edge as a generalization of the Hueckel edge operator <1-2>, and

a technique for its computation. Some experimental results are presented.

Representation of a Color Image: A digital image is described by a function, say I(x,y), defined at chosen grid points of the image. For a grey level (achromatic) image, the function I is scalar valued; its value being the luminance of the image at a certain point. For color images, three values must be specified at each point, i.e. the function I is vector valued and has three components.

A common choice for the three components of I is the Red, Green and Blue (R,G,B) components of a scene viewed through red, green and blue filters. The R, G and B components can be transformed to other quantities, more closely associated with human visual sensations of color, such as brightness, hue and saturation. The choice of components for representing color is discussed later.

Definition of a Color Edge: An edge in an achromatic image is defined by a discontinuity in the luminance function, I. The precise definition of a discontinuity varies with the different edge operators. To determine an edge in a color image, a definition for discontinuity in a three-dimensional color space is required. The simplest scheme is perhaps to compute edges in the three color components separately and determine an edge in the color image if certain relations between edges in individual components are satisfied. This paper describes a technique in which a constraint is imposed before the computation of individual edges. This approach is

based on an edge operator developed by Hueckel <2> for a single, grey level image which is first described briefly.

The Hueckel operator determines the presence of an edge in a circular neighborhood by fitting an optimal step to the input signal. An ideal edge step is defined to have a step or line profile in amplitude, this profile being constant along a straight line in the signal window. The edge is characterized by the position and orientation α of this line, and six other parameters which determine the amplitude and the profile of the edge (for details, see <2>). Hueckel's method determines the parameters of the edge step that matches best with the image in a given neighborhood.

The matching is simplified by expanding both the input signal and an ideal step into the first eight elements of a series of orthogonal functions. Let the coefficients of expansion be a_i and s_i for $0 \leq i \leq 7$. Note that the s_i are parameterized by the edge parameters. Optimal edge parameters are found by minimizing

$$N^2 = \sum_{i=0}^{7} (a_i - s_i)^2 \qquad (2.5\text{-}1)$$

It is shown in <2> that the optimal value for the edge direction α can be determined independent of the other edge parameters as a function of coefficients a_i. The remaining edge parameters are then computed separately. An edge is said to be present if the amplitude of the step with the best fit is sufficiently large and the fit is sufficiently good (N is small compared to the step amplitude).

To extend this concept of an edge to a color edge, an optimal step in the vector color space is defined first. Let a step in the color space have three components S_R, S_G and S_B in the three color components, and let the corresponding degrees of fit be N_R, N_G and N_B, where

$$N_R^2 = \sum_{i=0}^{7} (a_i - s_{iR})^2 \qquad (2.5\text{-}2)$$

Similar equations exist for N_G and N_B. Now, an optimal step in color space is defined by determining S_R, S_G and S_B such that

$$N^2 = N_R^2 + N_G^2 + N_B^2 \qquad (2.5\text{-}3)$$

is minimized. If the three step components are allowed to be independent, then minimizing N in eq.(3) is equivalent to minimizing the three components separately.

To simplify the optimization the three components are allowed to be independent, except that the spatial angles are constrained to be identical (i.e. $\alpha_R = \alpha_G = \alpha_B$). Once an optimal step with this constraint is obtained, a decision process for judging the presence of an edge is needed. The decision may be based on the degree of fit, determined by N, and a weighted sum of the squares of amplitudes of the steps in the three components, analagous to the achromatic image case.

Alternatively, the presence of edges in the three components may be determined separately from components of the optimal step above. Presence of an edge in the color image may now be based on

relationships between the edges in the three components. Experimental results using two types of relations are presented later.

Computation of a Color Edge: An expression for N, defined in eq.(3) can be easily generated from expressions for individual components (following eq.(12) in <2>). From this expression, it turns out that the angle α can still be chosen independent of other edge parameters (see Appendix 1). The remaining edge parameters may now be computed exactly as for the black and white images.

Computing the optimal value of α requires the solution of an eighth order polynomial. As an approximation, α can be determined as a weighted average (weighted by the step amplitudes) of the optimal angles of steps in the three components, determined separately. This is the method used for the results presented later.

In summary, an optimal step in color space is computed as follows. First, optimal steps are computed for the three components separately. A common angle is determined by a weighted average of the three angles. Using this new angle, the other edge parameters are recomputed, each component being treated independently again. Also new values for N_R, N_G and N_B are obtained. The resulting edge parameters are taken to define an optimal step in the color space. Next, the presence of edges in the three components of this optimal step is detected separately. Relationships between these edges are used to infer the presence of color edges as described below.

Experimental Results: Figures 1a, 1b and 1c show the R, G and B components of a color picture of a girl, an SMPTE (Society of Motion Pictures and Television Engineers) test picture. Figures 2a and 2b show the color edges obtained by computing an optimal step in color space as described earlier. Figure 2a shows the edges that occur in at least two components of the color step (the displayed edges are those occuring in the strongest component). Figure 2b shows edges occurring in all three components. These results may be compared with the edges obtained from an achromatic image derived from the color girl image and shown in figure 2c.

The edges shown in figure 2a contain fewer undesired edges. However, some desired edges may be lost because of the strict requirement of the presence of edges in all three components. Figure 2b contains a larger number of edges, filling in some of the gaps in figure 2a.

The number of components occurring in a color edge may be considered as a measure of the confidence or prominence of the edge. In subsequent processing only edges present in all three components may be processed initially. Edges with fewer components may then be used to fill in as analysis proceeds and more context becomes available.

Another approach is to accept a color edge if an edge is present in any of its color components. The color edge is further characterized by the components in which edges are present. This description is used in further processing of the edge data.

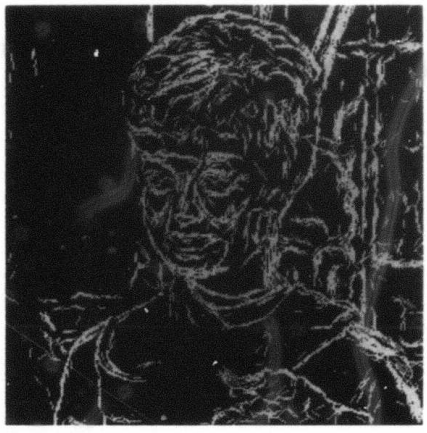
(a) "R" component

(b) "G" component

(c) "B" component

Figure 2.5-1. Three components of a color picture

(a) Color edges in two components



(b) Color edges in 3 components

(c) Edges in an achromatic image

(d) Color edges linked into segments

Figure 2.5-2. Edges detected in pictures of figure 1.

A common step in the processing of an image is to link the local edges into longer segments. One technique of linking edges into nearly straight line segments is described in <3>. It is also suggested there that the elongated segments are often useful in isolating the edges belonging to object boundaries. Only neighboring edges having orientations within specified limits of each other are linked. Further, only edge segments that contain at least a certain minimum number of edge elements are preserved.

The description of a color edge as defined above allows further constraints to be imposed on linking of two edges. The two edges to be linked are required to have compatible color characteristics. Two edges are defined to be color compatible, if the color components in which one edge occurs are the same as, or a subset of, the components in which the other edge occurs. Figure 2d shows the result of linking edges obtained from the girl picture with this additional constraint and using the method described in <3>. Only segments consisting of more than five edge elements are shown. Again, linking with tight constraints may be used to obtain boundary segments initially. These constraints may then be relaxed based on the additional context available now.

Choice of Color Coordinates: The edge detection scheme described above is applicable to any image where each pixel is described by three numbers. The conventional R, G and B components of an image were used for the results presented in the previous section. These components do not correspond directly to the human perceived color

attributes of an image. Human perception of color is normally described by the attributes of brightness, hue and saturation. The R, G and B components can be transformed to give these attributes. Edges obtained in the transformed space have more intuitive meaning and may be easier to use in further processing. e.g. it may be useful to extract edges differing in hue only.

Edge detection experiments, similar to those described for the R,G and B components, were performed in a transformed color space using a transformation based on a model of human visual system developed by Frei <4>. The transformed space contains a luminance component and two chromaticity components. The chromaticity components are claimed to be independent of the luminance for a wide range. The three components were scaled to have the same range of values. The results obtained are comparable to those of figures 2a, 2b and 2d. However, here the chromaticity edges are separated from the luminance edges and the linking of edges can be more selective.

Conclusions: An optimal step in color space that allows the steps in the three components to be combined in useful ways has been defined. Some alternatives for combining the components were discussed and experimental results presented. An important issue requiring further experimental investigation is the choice of color space used to represent an image.

Acknowledgements: The definition of optimal color step presented here was suggested to the author by Dr. T.O. Binford. W.D. Miller

helped with the early analytical and programming work. Dr. W. Frei provided much useful information on the choice of color coordinates.

Appendix 1

In reference <2>, Hueckel presents a formula, eq.(12), for computing N (for an achromatic image) which may be abbreviated as follows

$$N = f_1(a) + f_2(\alpha) + f_3(v) \qquad (2.5\text{-}4)$$

where f_1 is a function of the coefficients a_i, f_2 is a function of α only and f_3 is a function of vector $\underline{v}$ containing all the edge parameters. It is shown in <2> that f_3 is always equal to zero at the optimal solution point, and that to minimize N, α may be chosen independent of other parameters in $\underline{v}$.

For the color case it is necessary to minimize

$$N^2 = \sum_{j=R,G,B} (f_{1j}(a) + f_{2j}(\alpha))^2 \qquad (2.5\text{-}5)$$

Differentiation with respect to α yields an eighth order polynomial in $\cos(\alpha)$. After α is computed, other parameters of $\underline{v}_R$, $\underline{v}_G$ and $\underline{v}_B$ can be computed separately as for the case of black and white images.

References

1. M. Hueckel, "An Operator Which Locates Edges in Digitized Pictures," Journal ACM, Vol. 18, No. 1, January 1971, pp. 113-125.

2. M. Hueckel, "A Local Visual Operator Which Recognizes Edges and Lines," Journal ACM, Vol. 20, No. 4, October 1973, pp. 634-647.

3. R. Nevatia, "Object Boundary Determination in a Textured Environment," Proceedings of the ACM'75 Conference, Minneapolis, Minnesota, October 1975, pp. 32-36.

4. W. Frei, "A Quantitative Model of Color Vision," University of Southern California, USCIPI Report 540, pp. 69-83.

2.6 Recursive Estimator for the Determination of Boundaries

Nasser E. Nahi and Simon Lopez-Mora

The purpose of this research is to develop recursive algorithms for the estimation of boundaries of objects in noisy pictures. The available information is assumed to be limited to the statistics of the object boundary foreground and background.

The scanned picture is represented by

$$s(k) = \lambda(k)\, s_o(k) + (1-\lambda(k))s_b(k) \qquad (2.6\text{-}1)$$

for $k = 1,\ldots,N^2$ where N is the number of lines in the picture, s_o and s_b denote the intensity values of object and background, assumed to be sample functions of two statistically independent, cyclo-stationary random sequences whose first two moments are available, and λ is a binary valued function taking values of 1 or 0 corresponding to points of the image belonging to the object or the background respectively. The domains of s_o, s_b and λ correspond to the entire picture.

A set of observations

$$y(k) = s(k) + v(k) \tag{2.6-2}$$

is assumed, with $s(k)$ as defined in eq. (1) and $v(k)$ representing a zero mean Gaussian white noise sequence of variance σ^2.

The dynamic system chosen to represent $\lambda(k)$ is given by

$$\begin{pmatrix} v \\ w \\ v \\ s \end{pmatrix}(i+1) = A(i) \begin{pmatrix} v \\ w \\ v \\ s \end{pmatrix}(i) + B(i) \begin{matrix} u_1 \\ u_2 \end{matrix} (i) \tag{2.6-3a}$$

$$\lambda(k) = .5\{1 + \text{sign}\left(\frac{2c}{N}\left(\tilde{w}(i) + \bar{w}\right) - c - TR1(k - \tilde{s}(i) - s)\right)\} \tag{2.6-3b}$$

$$1 \le (i-1)N+1 \le k \le iN \le N^2$$

where,

$$\text{sign}(z) = \begin{matrix} 1 \\ 0 \\ -1 \end{matrix} \left\{ \begin{matrix} \text{if } z > 0 \\ \text{if } z = 0 \\ \text{if } z < 0 \end{matrix} \right. \tag{2.6-4a}$$

$$TR1(z) = \begin{cases} c + \frac{4c}{N} z & \text{if } -\frac{N}{2} \le z < 0 \\ c - \frac{4c}{N} z & \text{if } 0 \le z \le \frac{N}{2} \end{cases} \tag{2.6-4b}$$

$$TR1(z \pm N) = TR1(z) \qquad \text{for all } z \tag{2.6-4c}$$

for $c > 0$. It is shown that this nonlinear model represents the statistics of λ, where $\tilde{w}(i)$ is the variation of the object width at line i about its mean $\bar{w}$, $\tilde{s}(i)$ is the deviation of the width center from its mean $\bar{s}$ (measured from the middle of the picture), $A(i)$, $B(i)$ are matrices of appropriate dimensions, and $\underline{u} = (u_1, u_2)^T$ is a zero

mean white Gaussian sequence of unit covariance.

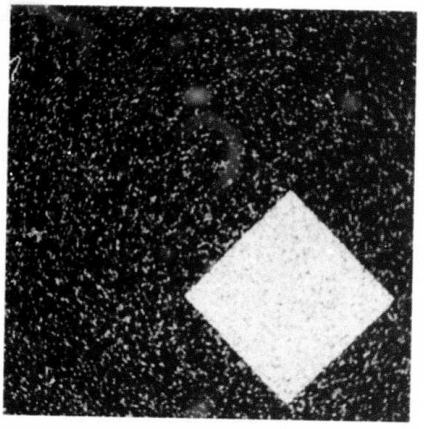
When dynamic models representing the statistics of the object and the background are adjoined to eq.(3) it is possible to formulate a nonlinear minimum mean square error estimation problem for which the method introduced by Wahi and Naraghi <1> can be implemented. To illustrate the performance of the estimator, a binary picture with an object intensity of unity and a background intensity of zero has been used. The parameters for the diamond shape in eq.(3) were calculated as

$$A(i) = \begin{pmatrix} .998 & 0 \\ 0 & .998 \end{pmatrix} \qquad (2.6\text{-}5a)$$

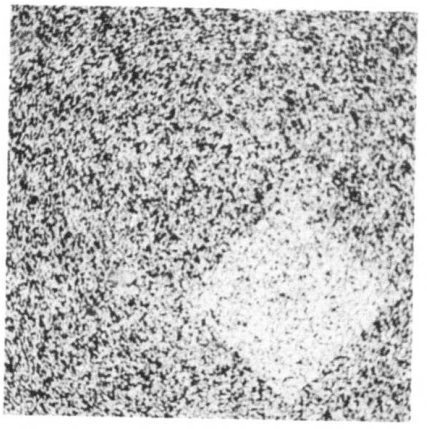
$$B(i) = \begin{pmatrix} 1.99 & 0 \\ 0 & .81 \end{pmatrix} \qquad (2.6\text{-}5b)$$

with $\bar{x} = 65$, $\bar{z} = 0$, $s = 256$ and $c = 1$. To single out the boundary estimation itself, $s_o(k) = 1$ and $s_b(k) = 0$ have been used. Note that $\bar{x} = 0$ corresponds to objects that should be centered, even though the original picture is not. Figure 1a shows the observation sequence when Gaussian noise of standard deviation 0.5 has been added to the original picture whose object intensity is unity and with background intensity zero while figure 1c shows the estimated (brighter trace) and original boundaries. Similar results for the case of Gaussian noise with variance 4 appear in figures 1b and 1d.

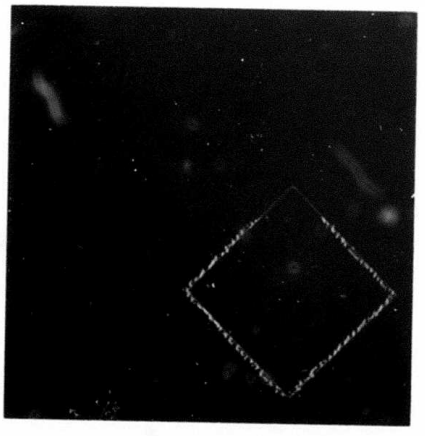
Further work is underway to generalize these results to more



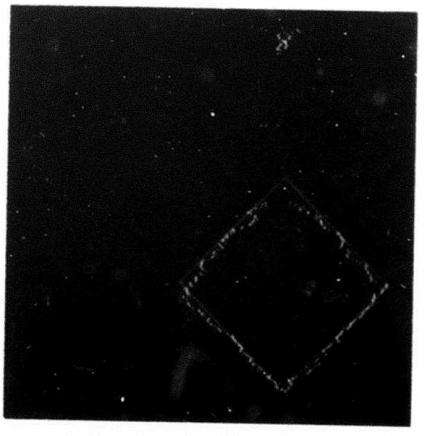
(a) Noisy observation, $\sigma = .5$



(b) Noisy observation, $\sigma = 2$



(c) Estimated and original boundaries



(d) Estimated and original boundaries

Figure 2.6-1. Recursive estimator for boundaries.

complex shaped objects possessing multiple gray levels.

References

1. M. Nahi and M. Naraghi, "A General Image Estimation Algorithm Applicable to Multiplication and Non-Gaussian Noise," 18th Midwest Symposium on Circuits and Systems, Montreal, Canada, August 11-12, 1975.

2.7 Figure of Merit for Edge Location

William K. Pratt

Relatively few studies of edge detector performance have been reported in the literature <1,2>. A performance evaluation is difficult because of the large number of proposed methods, difficulties in determining the best parameters associated with each technique, and the lack of definitive performance criteria.

In developing performance criteria for an edge detector, it is wise to distinguish between mandatory and auxiliary information to be obtained from the detector. Obviously, it is absolutely essential to determine the pixel location of an edge. Other information of interest includes the height and slope angle of the edge as well as its spatial orientation. Another useful item is a confidence factor associated with the edge decision, for example, the closeness of fit between actual image data and an idealized edge model. Unfortunately, few edge detectors provide this full gamut of information.

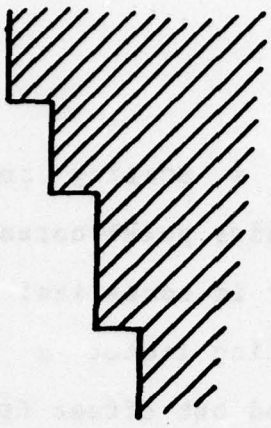
There are three major types of error associated with determination of an edge location: (a), missing valid edge points; (b), non-localized edge points; (c), improper classification of noise pulses as edge points. Figure 1 illustrates a typical edge segment in a digital image, an ideal edge representation, and edge representations subject to various types of error.

A common strategy in signal detection problems is to establish some bound on the probability of false detection resulting from noise, and then attempt to maximize the probability of true signal detection. Extending this concept to edge detection simply involves the setting of the edge detection threshold at a level such that the probability of false detection resulting from noise alone does not exceed some desired value. The probability of true edge detection can be readily evaluated by a coincidence comparison of the edge maps of an ideal and an actual edge detector. The penalty for non-localized edges is somewhat more difficult to assess. Edge detectors which provide a smeared edge location should clearly be penalized, however, credit should be given to edge detectors whose edge locations are localized but biased by a small amount.

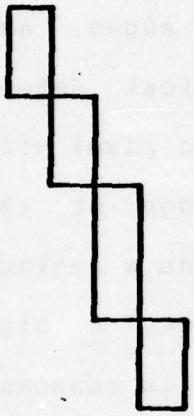
An edge location figure of merit for digital images has been developed. The figure of merit is defined as

$$F = \frac{1}{I_N} \sum_{i=1}^{I_A} \frac{1}{1+\alpha d^2} \qquad (2.7\text{-}1)$$

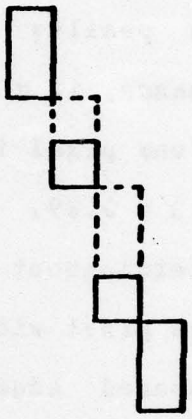
where $I_N = \max(I_I, I_A)$ and I_I and I_A represent the number of ideal and



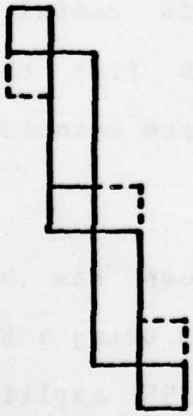
(a) IMAGE SEGMENT



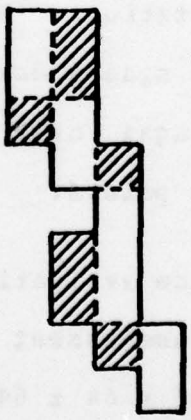
(b) IDEAL INDICATION



(c) FRAGMENTED INDICATION



(d) OFFSET INDICATION



(e) SMEARED INDICATION

Figure 2.7-1. Indications of edge location.

actual edge map points, α is a scaling constant and d is the separation distance of an actual edge point normal to a line of ideal edge points. The rating factor is normalized so that R = 1.0 for a perfectly detected edge. The scaling factor α may be adjusted to penalize edges which are localized but offset from the true position. Normalization by the maximum of the actual and ideal number of edge points insures a penalty for smeared or fragmented edges. As an example of performance, if α = 1/9, the rating of a vertical detected edge offset by one pixel is set at R = 0.90, and a two pixel offset gives a rating of R = 0.69. With α = 1/9, a smeared edge of three pixels width centered about the true vertical edge yields a rating of R = 0.93 and a five pixel wide smeared edge gives R = 0.84. A higher rating for a smeared edge than for an offset edge is reasonable because it is possible to thin the smeared edge by post-processing.

Some edge detectors provide an indication of edge height, slope angle, and orientation. For these detectors, it is useful to determine the mean square deviation in these quantities from their true values averaged over all true edge points which are coincident with detected edge points.

The performance evaluation methodology described above has been applied to the assessment of edge detection techniques using a test image consisting of a 64 x 64 pixel array over a 0 to 255 amplitude range with a vertically oriented edge of variable contrast and slope placed at its center. Independent Gaussian noise of standard deviation σ_n^2 has been added to the edge image, and the resultant

pixture has been clippel to the maximum display limits (0 to 255). The signal-to-noise ratio is defined as

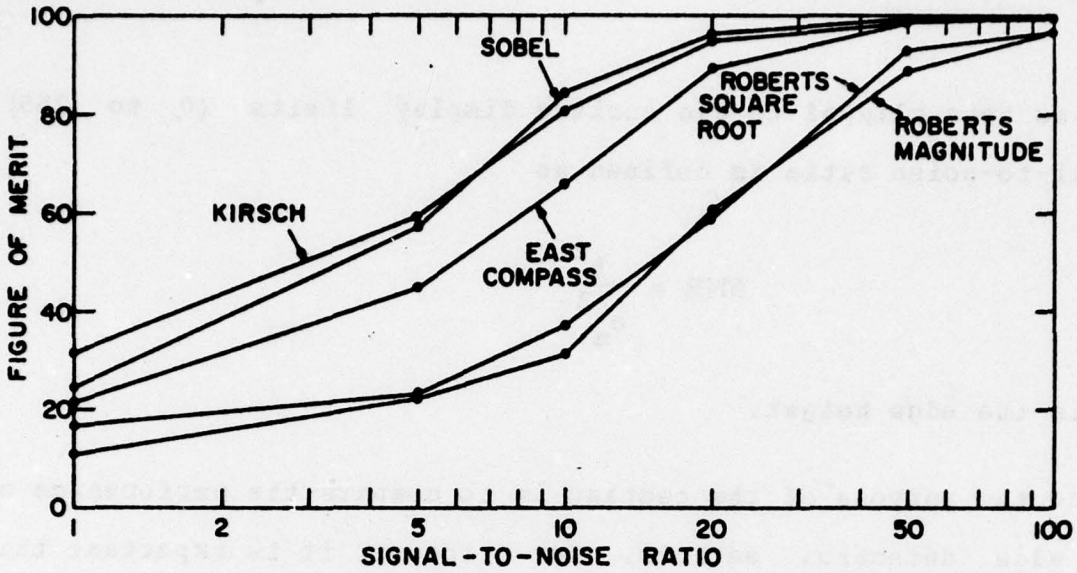
$$SNR = \frac{h^2}{\sigma_n^2} \qquad (2.7\text{-}2)$$

where h is the edge height.

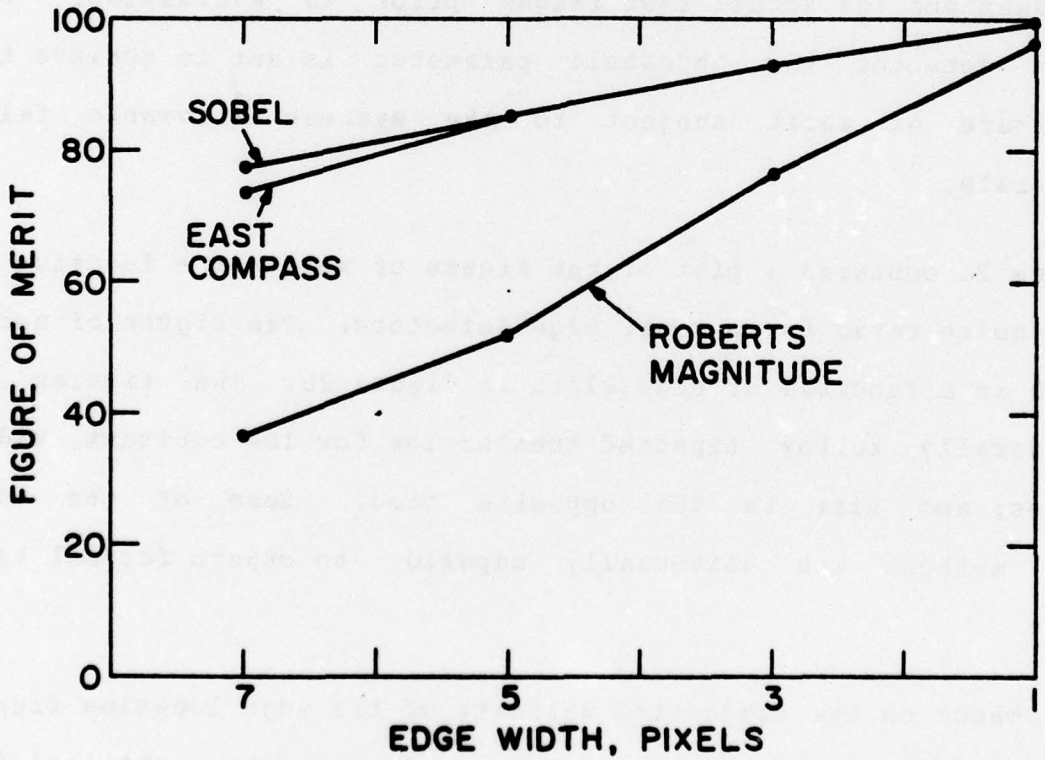
Since the purpose of the testing is to compare the performance of various edge detection methods, for fairness, it is important that each edge detector be tuned to its best capabilities. Consequently, each edge detector is permitted to train both on random noise fields without edges and the actual test images prior to evaluation. For each edge detector the threshold parameter is set to achieve the maximum figure of merit subject to the maximum allowable false detection rate.

Figure 2a contains a plot of the figure of merit as a function of signal-to-noise ratio for several edge detectors. The figure of merit is plotted as a function of edge width in figure 2b. The figures of merit generally follow expected trends: low for low contrast, wide, noisy edges; and high in the opposite case. Some of the edge detection methods are universally superior to others for all test images.

As a check on the subjective validity of the edge location figure of merit, figures 3 and 4 present the edge location maps obtained for several test images for high and low ranking edge detectors. These figures tend to corroborate the utility of the figure of merit. A

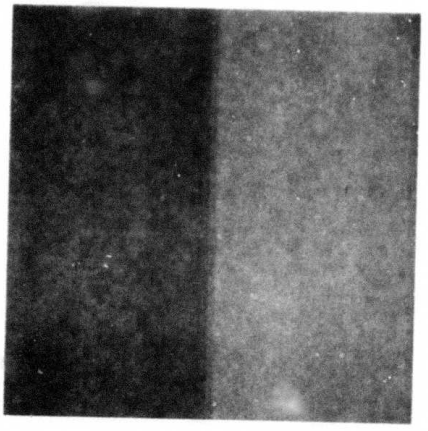


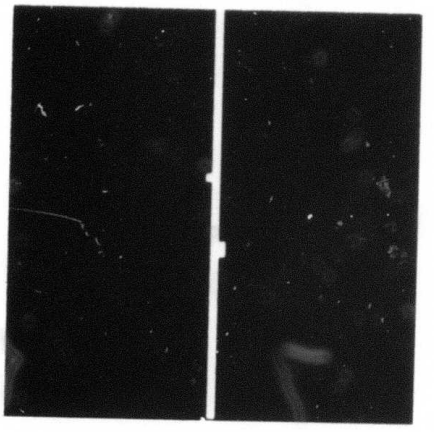
(a) Figure of merit vs. SNR



(b) Figure of merit vs. edge width

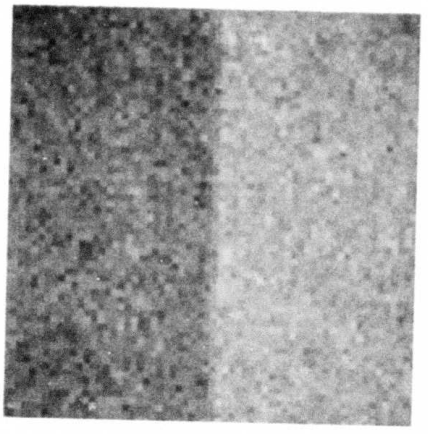
Figure 2.7-2. Edge location figure of merit as a function of signal-to-noise ratio and edge width.

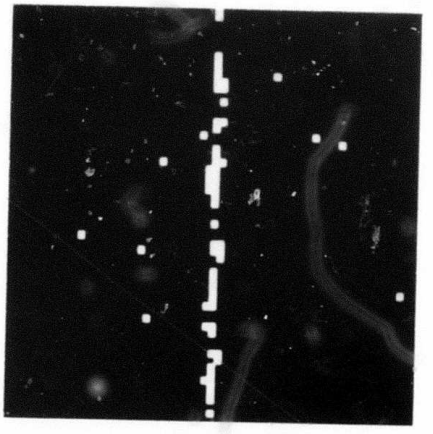




F=99.5%

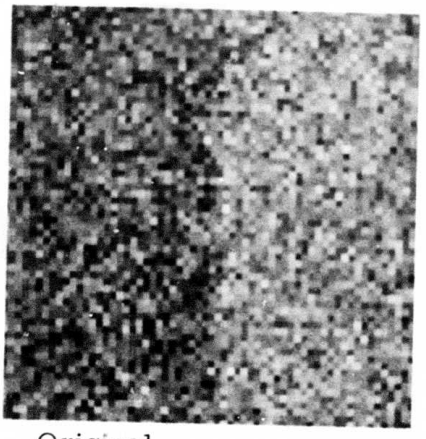
100, 25, 1



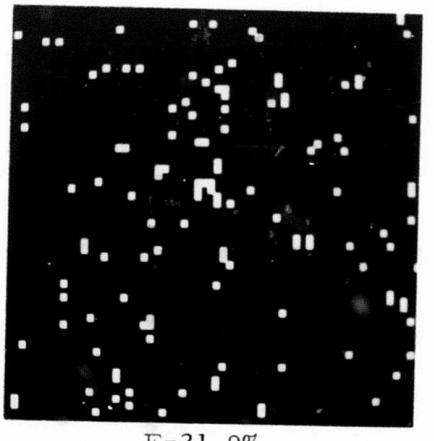


F=66.1%

10, 25, 1



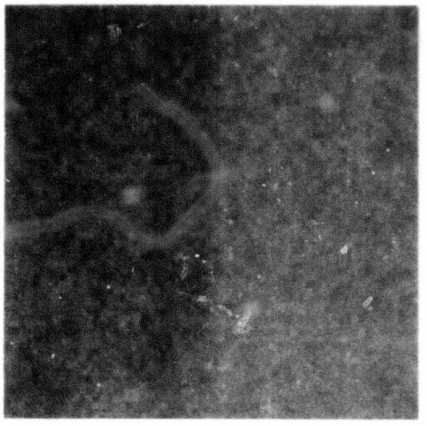
Original



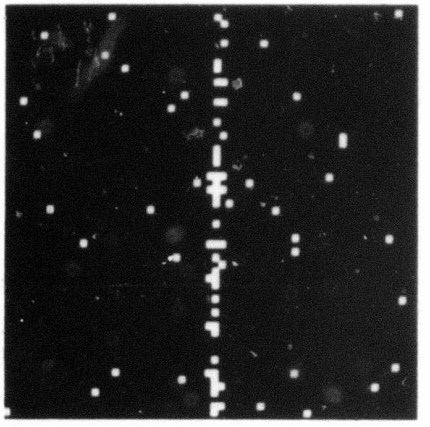
F=21.9%
Edge map

1, 25, 1

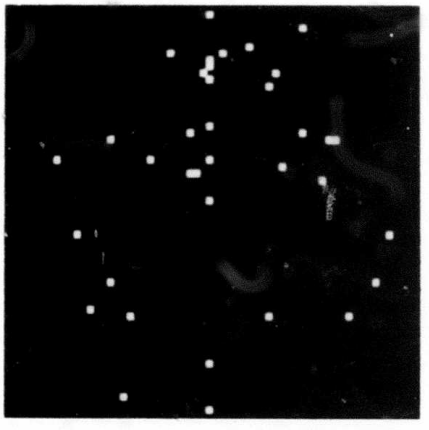
Figure 2.7-3. Edge location performance of Sobel edge detector as a function of signal-to-noise ratio. h=25, w=1.



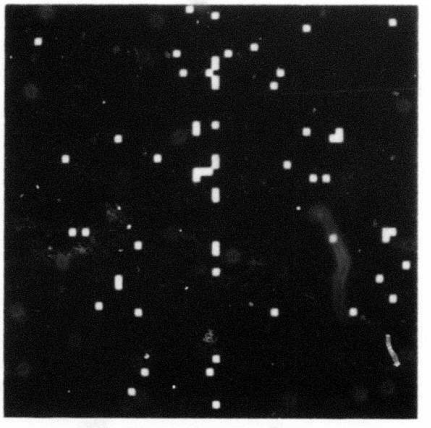
10, 13, 1



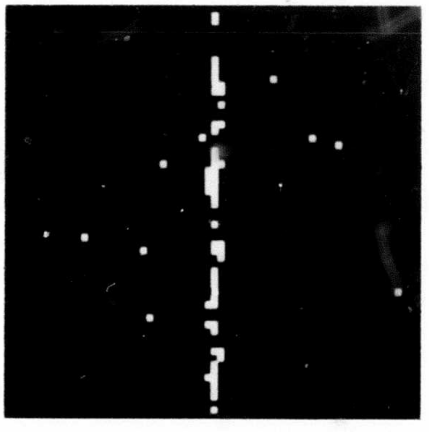
East compass
F=66.1%



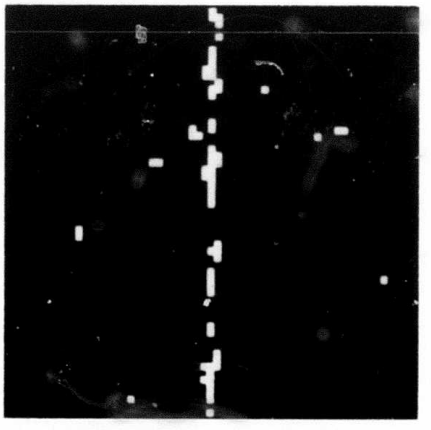
Roberts magnitude
F=31.5%



Roberts square root
F=37.0%



Sobel
F=85.1%



Kirsch
F=80.8%

Figure 2.7-4. Edge location performance of several edge detectors for SNR=10, h=13, w=1.

high figure of merit generally corresponds to a well located edge upon visual analysis and vice versa.

References

1. A. Herskovits and T.O. Binford, "On Boundary Detection," MIT Project MAC, Artificial Intelligence Memo 183, July 1970.

2. J.R. Fram and E.S. Deutsch, "On the Evaluation of Edge Detection Schemes and Their Comparison with Human Performance," IEEE Transactions on Computers, Vol. C-24, No. 6, June 1975, pp. 616-628.

2.8 Extraction of Prominent Features from Aerial Photographs

Lee W. Martin and Ram Nevatia

The identification of "prominent features" in an image yields guideposts for further processing. Prominent features are those that correspond to areas of interest for scene analysis, and are likely to exhibit some global properties. For example, the features may correspond to partial boundaries and be used to direct the scene segmentation process.

The attributes of a useful feature are, in general, unknown. The ideas upon which to base feature extraction are founded on human observations of scenes and, as is presently the case, the human observer is the final arbiter of what constitutes prominent features.

In this report features useful for making correspondences between two images of the same scene for image registration or stereo matching are considered. This approach is an alternative to previously used techniques of bulk area correlation which are very expensive computationally <1>. Extraction of some prominent features will significantly reduce the number of elements to be searched for correspondences. Some criteria for determining features are:

(a) relative invariance under slightly dissimilar imaging conditions and small perspective changes.

(b) reproducibility of the feature.

(c) uniqueness or at least the set of features which have common attributes should be small, to be useful in complement image matching.

(d) correspondence with areas of interest in the image, e.g. object boundaries.

This section describes a technique for extracting features derived from edge data. The features are called edge segments. Local edges consist of links satisfying certain relationships of proximity and orientation. Long edge segments often correspond to object boundaries <2>. Such edge segments satisfy the criteria above, to a degree, and are prominent in the sense that a human observer is likely to note them in describing the scene. A brief description of the program operation follows.

A Hueckel <3> edge operator processes the entire image returning edge data on position, orientation, and type. Edges are typed depending upon whether the discontinuity within the view area of the operator is most like: a step (S); a line (L); or unclassifiable as either (U). The distinctions are described briefly below (for more detail refer to <3>).

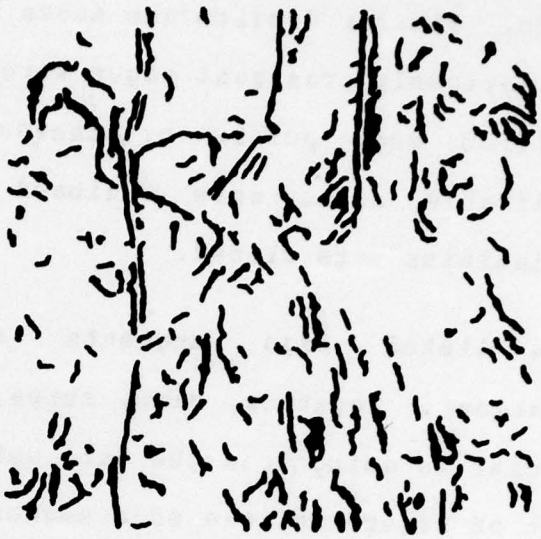
The view area of the edge operator is quasi-circular, and in the idealized circle a step (S) changes from some average brightness b_1 to a second value b_2, only two average brightnesses are considered in the disk; a line (L) occurs when there is a change from b_1 to b_2 and back to b_1 within the disk. The unclassifiable type results from edges which cannot be typed as being an S edge or an L edge.

The edge data obtained is operated upon by a program described in <2>. Briefly, edge points which have an orientation along a given direction (with a given angular tolerance) are considered. Edges which satisfy certain proximity relations are linked into edge segments. These edge segments are ordered sets of edge points which denote connectivity between adjacent edge elements. Only segments which contain more than n links (here, $n \geq 5$) are retained. This process is repeated for several angles over the entire range of directions in small steps (e.g. here, every 10 degrees).

In the present program, linking is performed on any of the three typed edge sets or combinations thereof to create edge segments. Figure 1 shows the digitized image which is 256 x 256 and figure 2 illustrates typical results. The total number of step edge points is

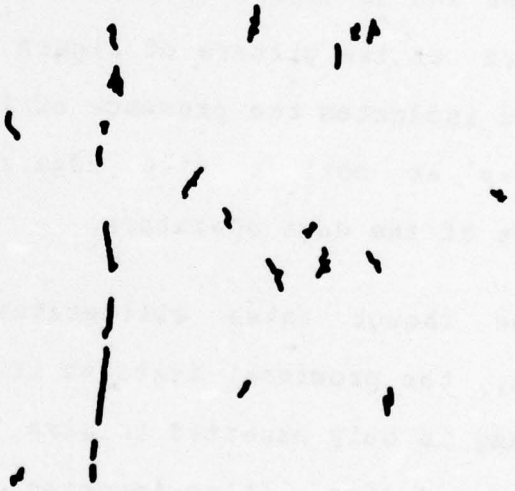
Figure 2.8-1. Digital aerial image
(Resolution 256 X 256)

(a) Step edge points
(3682 edge points)



(b) Step edge segments
(Step edge prominent features)

(c) Line edge points
(1213 edge points)



(d) Line edge segments
(Line edge prominent features)

Figure 2.8-2. Effects of edge linking to extract prominent features.

shown in figure 2a. The linked segments which result are shown in figure 2b. Similar results are shown for line edges in figures 2c and 2d. No obviously prominent edges were observed by linking over the unclassified edge points by themselves, however, when the step and unclassifiable edges were unioned and linked some boundary discontinuities were closed.

The linked edge segments are features with direction (orientation), location, size, shape, and position. Note the effects which occur in going from the data set of figure 2a, the edge points, to that of figure 2b, the edge segments (similarly for figures 2c and 2d). The edge points which are retained in edge segments correspond to the delineation of some of the prominent features, e.g. outlining some river and lakeshore areas, as well as much of the highway in the left part of the picture of figure 2b. The linking of line edges in figure 2d indicates the presence of line type features (the highway is picked up as both a step edge and a line edge depending upon the placement of the edge operator).

Even though noise obliterates some segments that a human observes, the prominent features are still easily recognizable. This processing is only expected to give features that suggest possible object boundaries. Also important is the reduction of the number of elements being considered. Not only have approximately 35% to 50% of the edges been rejected for failure to meet linking criteria, but those retained are in sets of edge points. The search space using edge segments contains considerably fewer elements. Note that a

majority of edges which occur due to texture are eliminated. The edges in textured regions have random orientations and do not form sufficiently long connected segments.

Thus it can be seen that features have been obtained which are prominent in some sense, and which also satisfy our criteria (to a degree). In addition, most texture edges are filtered out. Third, areas of the image where further processing might be more fruitful have been identified. Fourth, the original data set of oriented edges has been changed and reduced to a set of edge segments which contain information of a more global nature.

Finally, the edge segments yield areas from which correspondences between two images can be generated, based on some measure of "sameness", e.g. correlation of segments based on direction, location, size, etc. Spatial relationships between edge segments will further reduce the ambiguity in matching of features in two images.

This program has created features by linking edges along angles to form basically straight line segments along the central angle. In figures 2b and 2d interangle links do not exist, even though they appear to exist. The eye creates the illusion of interangle linking. This first step has considerably reduced the number of elements to be considered for further processing. The next step is to link the edge segments from different angles that intersect. Such linking will create larger edge segments. The threshold for number of edge points could be increased to eliminate some smaller segments. Thus edges that link over enough edge points over one angle but never link up to

anything else would be eliminated. This will result in further reduction in the number of features without eliminating the important ones.

References

1. Hannah, M.J., "Computer Matching of Areas in Stereo Images," Stanford A.I. Laboratory Memo, AIM 239 STAN CS-74-438, July 1974.

2. Nevatia, R., "Object Boundary Determination in Textured Environment," Proceedings of the ACM 1975 Conference, Minneapolis, October 1975 (also in USCIPI Report 620, September 1975).

3. Hueckel, M.H., "A Local Visual Operator Which Recognizes Edges and Lines," Journal ACM, Vol. 20, No. 4, October 1973, pp. 634-647.

2.9 Clustering for Image Segmentation

Harry C. Andrews

This report should be viewed as a broad research outline for an approach to automatic image segmentation. The technique is based upon methodologies which have proven successful in the past in the companion fields of pattern recognition and signal processing of images. Specifically the approach suggested here involves a mathematically generalized method to accept features or attributes of potential segments, to computationally select the best reduced set of attributes or features for successful segmentation, and then to subsequently attempt hardware simplification for ultimate

implementation at the sensor end of an image acquisition system. The generalization of this segmentation method will allow subsequent addition of features defined by new operators (as they are invented) without changing the segmentation technique. This then will allow a quantitative methodology for evaluating new attributes as they come along while correspondingly improving the segmentation results without need of modification of the segmentation procedure.

Before developing the specifics of the proposed segmentation procedure, it is useful to describe the placement of this technique within the heterarchial formalism of USC's approach to image understanding. Essentially the image segmentor is considered a preprocessor to the interpretation element of the image understanding which initially operates as a bottom up processor accumulating evidence as to proper segmentation as it progresses. This is not to say that it will not be responsive to feedback from the higher level world models and interpretation phases, for in fact this feedback will aid in the optimization of the reduction of necessary features for effective segmentation.

Finally one peripheral application of a successful image segmentation procedure consists of novel image coding methods. These methods would fall into the category of "image synthesis via image analysis," the image analysis being automatic segmentation. As a simple example consider a segment which has a property vector in N space in which (for simplicity) each property coordinate is quantized to 5 bits. This implies a total of 32 different textures, 32

intensities, 32 possible means, 32 potential variance values, etc. Thus with 5 bits per property and N properties per segment, it may be possible to obtain very large bandwidth reductions through combination of the attributes of segments with segmentation boundaries. Clearly these possibilities need to be investigated.

Property-Attribute Vector: The philosophy of "clustering for image segmentation" is to describe arbitrary possibly overlapping regions of an image by a vector of attributes or properties which humans feel would be useful descriptors to differentiate between segments. Because of the lack of complete understanding of how segmentation occurs, the property vector will probably be overspecified with highly redundant attributes or coordinates. However this is to be expected since a "significant" feature selection operation will follow the property vector definition. Before describing the feature selection and dimensionality reduction aspects of automatic image segmentation, it will be useful to list a few candidate attributes:

local area mean
local area variance
local area histograms
Robert's operator
Hueckel operator
Baker operator
other edge operators
texture operators (joint density measures)
entropy
correlation
intensities
color, R,G,B
color Y,I,Q
color L,C_1,C_2
ratio images
angle images
depth from stereo
spatial frequencies
other transform coefficients
principal components

The above list is not meant to be all inclusive, but is provided as a set of suggested first alternatives for the property vector.

Clustering and Feature Selection: The N dimensional vector describing the properties characterized above becomes a point in (hopefully Euclidean) N space. If the original image is broken into possibly overlapping small rectangular sections, and if each section is described by an attribute vector, then each section becomes a point in N dimensional space. During the training phase of the clustering procedure, a class of images will be selected in which unambiguous segmentation exists and this class of images will be sectioned, attributes computed, and resulting vectors placed as points in N space. Remembering that the axes in this space represent texture, color, brightness, etc., pattern recognition clustering procedures

will then be applied to get the points to cluster as tightly as possible. Such techniques include convergence algorithms and measures for describing tightness of clusters using between, within, and mixture scatter matrices. Coupled with these will be feature selection procedures to remove those attributes which tend to scatter the data and do not contribute to well defined clusters. Typical measures for effecting such feature selection include divergence and Bhattacharyya distances as useful attributes or property discriminators <1>.

The entire procedure of sectioning, property vector computation, imbedding in N dimensional space, clustering, and feature selection is directed at effective and efficient image segmentation. However, it is not clear what a "properly segmented image" is or what a "better partition of the image" means. Therefore it will be necessary to direct this training procedure with extra-image knowledge, i.e. human intervention. Much was accomplished in the way of the development of texture discriminants for consistency with the human visual process <2>, a similar procedure is anticipated in the definition of "correct partitioning" in the segmentation problem. Specifically, imagery will be selected in which human interpretation results in consistent image segmentation. A segmentation discriminant will then be developed much as in the texture discriminant work <2>. Extrapolation to more difficult imagery will follow with necessary fine tuning of the segment discriminator via additional attributes or properties and more precise clustering and feature selection.

References

1. Andrews, H.C., An Introduction to Mathematical Techniques in Pattern Recognition, Wiley, 1972

2. Thompson, W.B., "The Role of Texture in Computerized Scene Analysis," USCIPI Report 550, December 1974.

2.10 Interpretation and World Knowledge

Erica M. Rounds

The goal of this research is to design the decision-making component of the image-understanding system which will construct a meaningful description of the real-world scene depicted in the image. Research efforts are concentrated on two sub-problems:

(a) the decision procedures which assign meaning to the regions or parts of the image discovered during the segmentation stage (interpretation); and

(b) the efficient representation of objects and their properties and relations in the problem domain (knowledge base).

Interpretation is essentially a decision process designed to select from a set of possible meanings one which fits the given input image. There is an obvious similarity to the classical pattern recognition problem, the main distinction being one of complexity. In image understanding, the physical basis of the image is of major importance, and the descriptions of object classes and segmented

regions in the image require more complex data structures than is customary in pattern recognition. However, with suitable modifications, some of the pattern recognition methodology can be usefully applied to the problem of image understanding <1>.

Decision Process: The underlying strategy for the interpretation is based on the "hypothesis formation and validation" paradigm. This is a two-phase process consisting of a long range strategy (global search) and a short range strategy (local search). During the long range strategy, a coarse evaluation of the regions is attempted to discover some prominent features which suggest tentative hypotheses concerning the objects in the scene. The short range strategy then accumulates evidence for the hypotheses by performing a more detailed analysis of the pictorial data. This paradigm is supported by evidence from human visual processing which suggest that an initial overall scanning of a scene is followed by a close scrutiny of some prominent object <2>.

The model for decision-making is shown in figure 1. Since interpretation is strongly dependent on the output of the segmentation stage and conversely, segmentation into regions may be influenced by partial assignment of regions, the segmentor is explicitly included in the decision mechanism. A full image understanding system will require the integration of all components from low-level pictorial data processing to high-level interpretation, and must provide for multi-directional information flow between the components. During the initial developmental stage attention will be directed to the

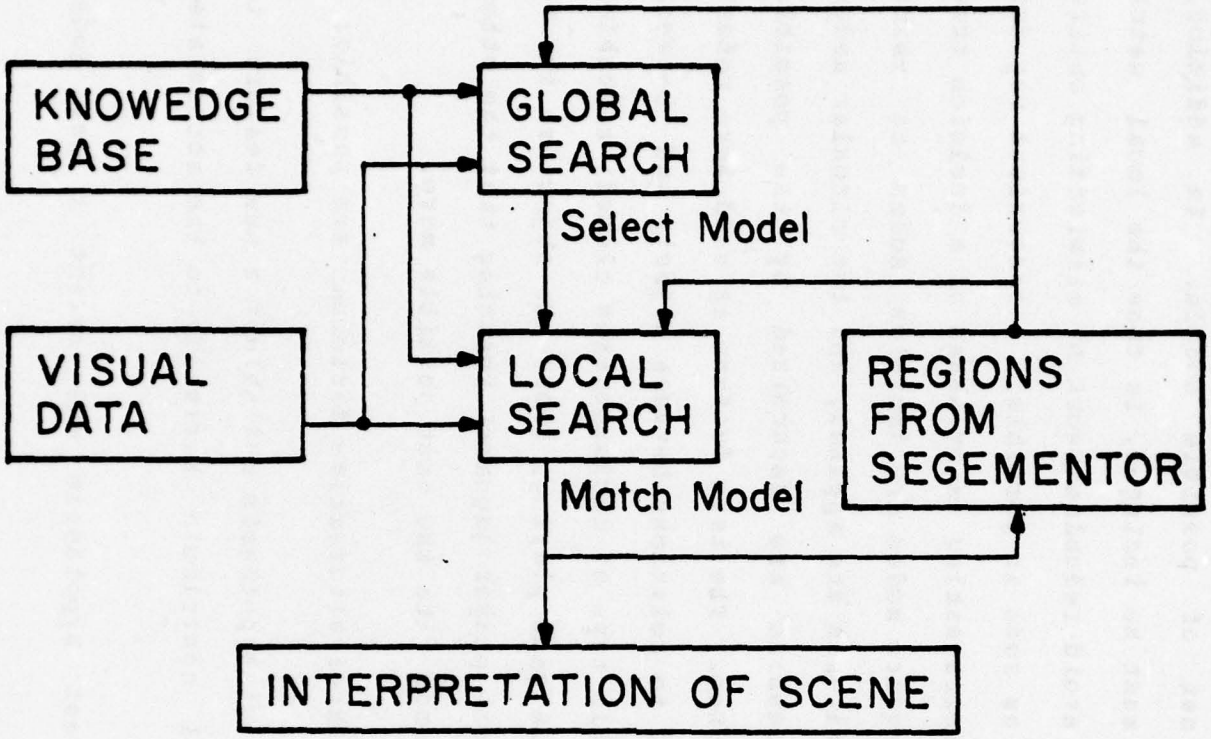


Figure 2.10-1. Decision Mechanism.

interaction between the interpretation and segmentation modules.

The global search requires selection criteria for choosing a candidate model from a set of possible models. In addition, provisions for backtracking must be included, in case the local match fails. Backtracking should avoid redundant work by eliminating entire subtrees of the search tree as soon as possible. A convenient way for describing sequential decision-making is by means of a decision tree as shown in figure 2. The square nodes are decision nodes at which selection or termination criteria are applied, and the circular nodes are chance nodes whose successors are determined by the possible values of the tested features. The term feature as used here refers to attributes of objects and to relations between objects and between components of objects. This type of decision tree closely resembles the so-called AND-OR trees in game playing, where the objective is to optimize the strategy for one player (square) assuming that the other player (circle) always responds with the best possible move.

At each decision node three alternative decisions are possible:

(a) continue with current hypothesis and select a new feature to be tested which will contribute maximally to the accumulated evidence;

(b) discard the current hypothesis and select a new model (backtrack); and

(c) accept the current hypothesis as the final interpretation of the scene (terminate search).

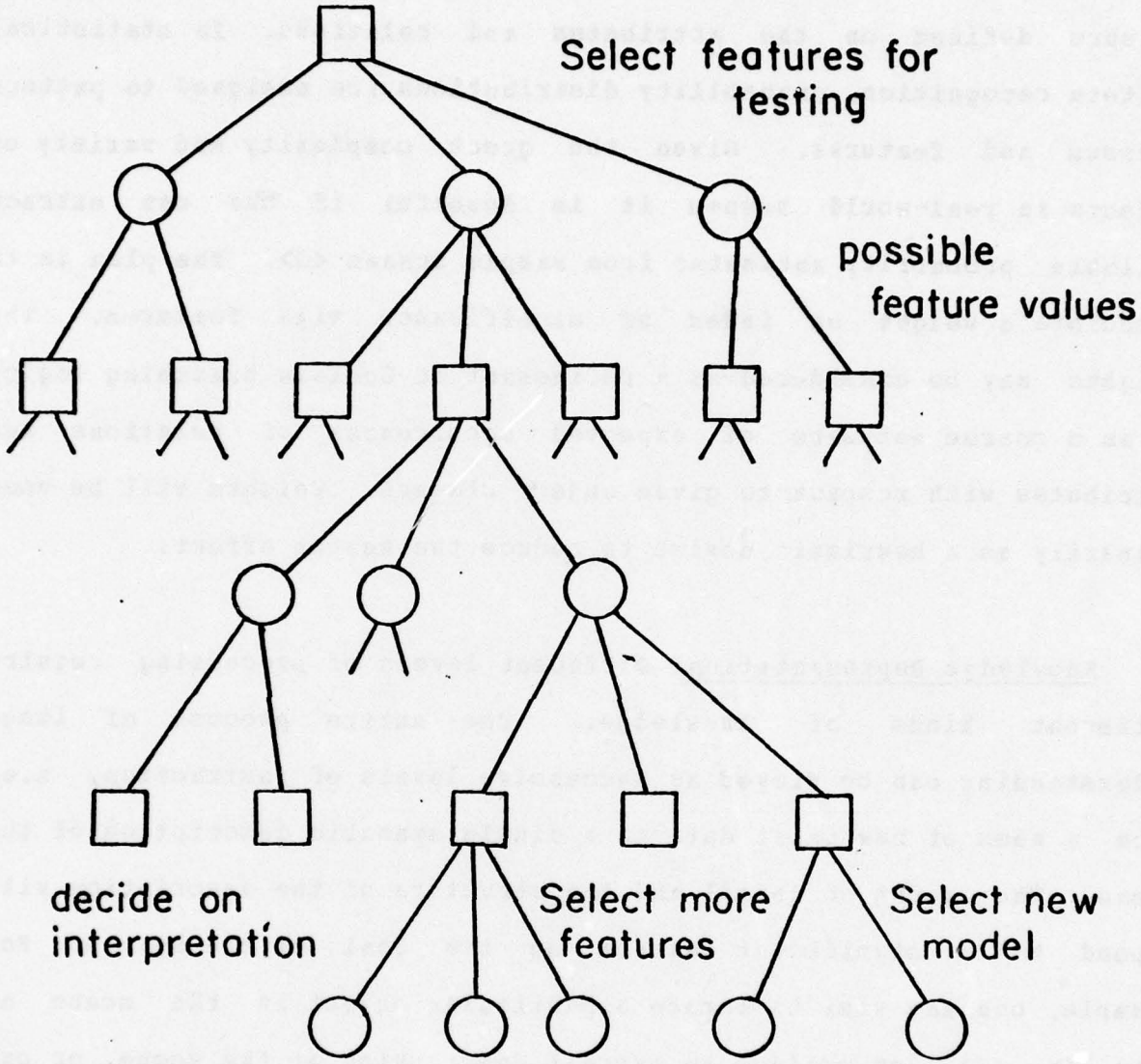


Figure 2.10-2. Partial Decision Tree.

Selection and termination criteria are based on some quantitative measure defined on the attributes and relations. In statistical pattern recognition, probability distributions are assigned to pattern classes and features. Given the great complexity and variety of objects in real-world scenes it is doubtful if one can extract reliable probabilty estimates from sample scenes <3>. The plan is to associate a weight or index of significance with features. The weights may be considered as a refinement of Boolean branching logic, or as a coarse estimate of expected occurrences of relations and attributes with respect to given object classes. Weights will be used primarily as a heuristic device to reduce the search effort.

Knowledge Representation: Different levels of processing require different kinds of knowledge. The entire process of image understanding can be viewed as successive levels of abstraction, i.e. from a mass of raw input data to a single symbolic description of the scene. The amount of detail and the structure of the description will depend to a significant extent on the goal of the system. For example, one may wish to locate a particular object in the scene as done in <3>, or produce an overall description of the scene, or one may wish to extract enough information so that the system can answer meaningful questions about the scene.

The system is to be applied to twenty images of outdoor scenes with a combination of man-made and natural objects. The first problem domain is derived from aerial photographs showing various vehicles in a desert-like environment. Object classes consist of vegetation,

bushes, sand, tracks, and vehicles of type tank, truck and airplane. The goal of the system is to detect and interpret the man-made objects from the background desert and vegetation. The interpretation module will construct a conceptual model of an input image including a description of major objects and structural information relating these objects to the scene as a whole.

As indicated in figure 1, knowledge used by the decision processes is roughly divided into visual data, region descriptions produced by the segmentor, and knowledge about the problem domain retrieved from the knowledge base. The first two kinds of information change dynamically during processing and may be considered part of the short-term memory which stores the image descriptions produced at different levels. The knowledge base is static and constitutes the long-term memory containing models of scenes and objects.

In reference <1> decision trees provide an implicit description of the object classes. A path through the tree specifies which features a region or set of regions must possess in order to be assigned to a given object class. In our system object descriptions are explicitly stored in the knowledge base. This has the advantage that the decision processes are largely independent of the problem domain. Thus the system can achieve a certain degree of generality and may be applied with relatively few modifications to other problem domains by providing a different knowledge base. Similar considerations of flexibility have influenced the choice of a two-part search. The long-range strategy uses a global decision tree to select

a model on the basis of a few discriminatory features. The local search then attempts a matching of region descriptions against model descriptions in the knowledge base. This approach can easily be extended to the case where input images may come for a number of different problem domains. The global strategy would first select the applicable problem domain followed by a selection of a candidate interpretation.

A major problem in developing the knowledge base is the identification of relevant features. Criteria for "good" features include discriminatory power and a certain degree of invariance, e.g. with respect to viewing angle and scale. In addition, semantic knowledge such as nonpictorial attributes may be essential for unambiguous interpretation <5>. Under consideration are the following features which have proved useful at the interpretation level:

(a) optical properties (texture, shadows, highlights, reflectivity)

(b) geometric attributes (relative size, location, shape)

(c) spatial relations (distance, size ratios, orientation)

(d) topological relations (adjacency, containment, occlusion, connectivity)

(e) semantic information (functional, spatial and temporal context)

Initially, scene and object descriptions will be derived by segmenting the digitized images of the example problem domain manually and selecting features using human protocol. To reconstruct objects

seen from different perspectives, objects descriptions will be stored in three canonical views: front, top, and side. Representation of world knowledge is in the form of graph-like structures similar to semantic nets, where the nodes represent concepts, object categories, objects, and components of objects, and the edges represent relations between nodes or properties <6>. The weights discussed above are associated with the edges. The grouping of objects into categories which share common properties allows a more parsimonious description and will reduce the search effort. The major categories are natural objects and man-made objects which are further subdivided by size, shape, texture, etc.

Too much specialized detail stored in the knowledge base will impede matching with region descriptions. The system must be able to tolerate deviations from the expected prototypes. Our approach is related to Minsky's frame concept <7>, i.e. certain features are labelled as essential, which must always be matched, and others are nonessential which may be assigned a default value representing the most likely value.

The foregoing discussion has provided an overview of the approach in the design of an interpretation module. Many details have not been covered because the ideas are not yet fully developed. The interpretation module, for example, will consist of many other subprocesses in addition to the hypothesis generator and matching process. Region descriptions obtained from the segmentor must be analyzed for shape, shadows and related properties. Modular

development of the system will facilitate the addition of new subprocesses.

References

1. Y. Yakimovsky, "Scene Analysis Using a Semantic Base for Region Growing," AIM-209, Stanford University, June 1973.

2. D. Noton, "A Framework for Representing Knowledge," AI Memo No. 306, MIT, June 1974.

3. D.H. Fishman, A.R. Hanson, and E.M. Riseman, "Some Considerations in a Model Building System for Scene Analysis," COINS Technical Report 75C-2, University of Massachusetts, March 1975.

4. J.M. Tenenbaum and S. Weyl, "A Region-Analysis Subsystem for Interactive Scene Analysis," Technical Note 104, Stanford Research Institute, June 1975.

5. M.L. Baird and M.D. Kelly, "A Paradigm for Semantic Picture Recognition," Pattern Recognition, Vol. 6, 1974, pp. 61-74.

6. S. Fahlman, "A System for Representing and Using Real-World Knowledge," AI Memo No. 331, MIT, May 1975.

7. M. Minsky, "A Framework for Representing Knowledge," AI Memo No. 306, MIT, June 1974.

3. Image Processing Projects

The image processing projects comprise an ongoing research activity directed toward image coding, image restoration, vision modelling, and the implementation of image processing systems. In image coding novel ideas based upon the results of the image understanding study are being explored as a means of achieving significantly higher compression ratios than obtainable by conventional coding methods. The image restoration studies are directed toward the solution of two major problems: blind restoration in which a priori information about image degradation is unavailable or incomplete; and constrained restoration which involves the use of luminance bounds and smoothness criteria to improved image restoration. Vision modelling research activities include the extension of previously developed models of the human visual system to encompass higher levels of visual perception in support of the image understanding program. Implementation studies are underway on techniques of nonlinear, two dimensional optical filtering which can be utilized as a form of sensor based image processing.

3.1 Smoothing Splines for Spatially Variable Restoration

M. Javad Peyrovian and Alexander A. Sawchuk

As a continuation of previous work <1>, spline functions have been utilized for locally variable filtering restoration of noisy blurred images <8>. For an unblurred noisy image the image g(x) is given by

$$g(x) = f(x) + n(x) \tag{3.1-1}$$

The restoration operation is designed to minimize the integral

$$\int [f'(x)]^2 dx \tag{3.1-2}$$

among all twice differentiable functions f(x) such that

$$\sum_i \left(\frac{f(x_i) - g(x_i)}{\delta_i} \right)^2 \leq S \tag{3.1-3}$$

where δ_i locally controls a smoothing window at point x_i and S controls the overall extent of smoothing. If available, σ_i the standard deviation of the noise at point x_i can be used for δ_i . In this case, natural values of S lie within the confidence interval of the left hand side of eq.(3) That is

$$N - (2N)^{\frac{1}{2}} \leq S \leq N + (2N)^{\frac{1}{2}} \tag{3.1-4}$$

where N is the number of data points. Reinsch <2> has shown that the solution to eqs.(2) and (3) is a cubic spline and more generally is a spline function of degree 2K-1 for least square minimization of the K th derivative instead of the second derivative. The case K = 2 leads to very simple algorithms for the construction of the function f(x).

Assuming f to be a polynomial of degree 3 or less in each interval, the optimal restoration filter with respect to the conditions of eqs.(2) and (3) can be obtained by calculus of variations <2>. The solution is

$$\underline{f} = \underline{g} - \underline{D}^2\underline{Q}\,(\underline{Q}^T\underline{D}^2\underline{Q} + p\underline{T})^{-1}\underline{Q}^T\,\underline{g} \tag{3.1-5a}$$

where

$$\underline{f} = [f(x_1),\ f(x_2),\ldots,f(x_N)]^t \tag{3.1-5b}$$

$$\underline{g} = [\,g(x_1), g(x_2),\ldots,g(x_N)] \tag{3.1-5c}$$

$$\underline{D} = \text{Diag}\,[\delta_1, \delta_2,\ldots\ \delta_N] \tag{3.1-5d}$$

Matrix $\underline{T}$ is a positive-definite tridiagonal matrix of order N-2, and $\underline{Q}$ is a tridiagonal matrix with N rows and N-2 columns. For a uniformly sampled image, $\underline{Q}$ and $\underline{T}$ have the form

$$\underline{Q} = \begin{bmatrix} 1 & & \\ -2 & 1 & 0 \\ 1 & -2 & -1 \\ 0 & 1 & 2 \\ & & -1 \end{bmatrix} \tag{3.1-6}$$

$$\underline{T} = \begin{bmatrix} 4/3 & 1/3 & & & 0 \\ 1/3 & 4/3 & 1/3 & & \\ & & & & \\ & & & 1/3 & 4/3 \\ 0 & & & & 1/3 \end{bmatrix} \tag{3.1-7}$$

where p is a Lagrangian parameter which, like S, controls the overall extent of smoothing. Since $\underline{T}$ is invertible, by using matrix identities <3>, eq.(5) can be written in the more concise form

$$\underline{f} = (\underline{I} + p^{-1}\,\underline{D}^2\,\underline{Q}\,\underline{T}^{-1}\underline{Q}^T)^{-1}\,\underline{g} \tag{3.1-8}$$

Although eq.(8) appears simpler, it needs two matrix inversions in

comparison to eq.(5) which requires the inverse of a banded matrix. The Cholesky decomposition <4> $\underline{R}^T\underline{R}$ of a positive-definite band matrix $\underline{Q}^T\underline{D}^2\underline{Q} + p\underline{T}$, where $\underline{R}$ is a lower diagonal (triangular) matrix, provides an efficient computational algorithm.

An interesting property of the fidelity criterion of eq.(3) is that the smoothing window can be locally controlled by determination of δ_i. If the noise variance is higher in some regions, δ_i can be set larger at that region. An application of this property is the filtering of images with film-grain noise, modeled <5> by

$$D_r = D_s + D_s^{\frac{1}{2}} n \qquad (3.1\text{-}9)$$

where D_s is the signal density, D_T is recorded density and n is zero mean noise. Equation (9) shows that when the signal has a higher amplitude, noise has a greater variance. To apply the above mentioned filter, the value of δ_i and an estimate of the signal are needed. Hunt <6> has shown that images are well described statistically as a stationary variance about a spatially non-stationary local mean. Assuming ergodicity of similar classes of images, the spatial average may be used as an estimate of the ensemble average. A local spatial average is thus used as a non-stationary estimate of the signal mean.

A noisy blurred image can be expressed in the form

$$\underline{g} = \underline{\underline{H}}\underline{f} + \underline{n} \qquad (3.1\text{-}10)$$

for a discrete model. The following fidelity criteria for image restoration may then be formulated as minimization of

$$\int (\underline{f}'(x))^2 \, dx \tag{3.1-11}$$

among all f(x) such that

$$\|\underline{D}^{-1} (\underline{H}\underline{f} - \underline{g})\|^2 \leq S \tag{3.1-12}$$

where $\underline{D}$, $\underline{f}$ and $\underline{g}$ are previously defined. Using the same procedure as before, the estimate

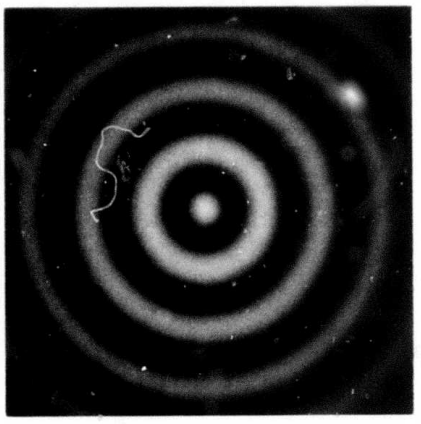
$$\hat{\underline{f}} = (\underline{H}^T \underline{D}^{-2} \underline{H} + \lambda \underline{Q} \underline{T}^{-1} \underline{Q}^T)^{-1} \underline{H}^T \underline{D}^{-2} \underline{g} \tag{3.1-13}$$

is obtained, where $\lambda = 1/p$. Setting $\underline{H} = \underline{I}$ leads to the estimate of eq. (8). Comparison of this restoration filter to the Wiener filter matrix

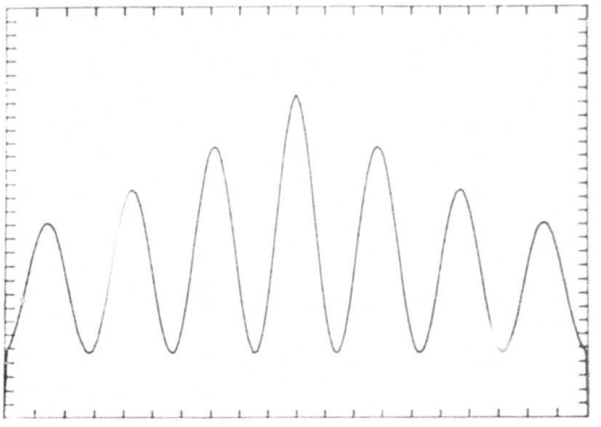
$$\underline{W} = (\underline{H}^T \underline{C}_n^{-1} \underline{H} + \underline{C}_f^{-1})^{-1} \underline{H}^T \underline{C}_n^{-1} \tag{3.1-14}$$

shows the similarity of these two filters with $\underline{C}_f^{-1}$ replaced by $\lambda \underline{Q}\underline{T}^{-1}\underline{Q}^T$. In order to obtain the filter numerically, it is necessary to invert matrix $\underline{T}$ as well as $\underline{H}^T\underline{D}^{-2}\underline{H}+\lambda\underline{Q}\underline{T}^{-1}\underline{Q}^T$. Matrix $\underline{T}$ is a positive-definite tridiagonal matrix; efficient techniques exist for computation of its inverse <7>. Matrix $\underline{H}$ is a rectangular M x N matrix where $M < N$ for non-zero background pictures, thus an exact inverse does not exist for $\underline{H}^T\underline{D}^{-2}\underline{H}+\lambda\underline{Q}\underline{T}^{-1}\underline{Q}^T$, and the pseudo-inverse must be employed.

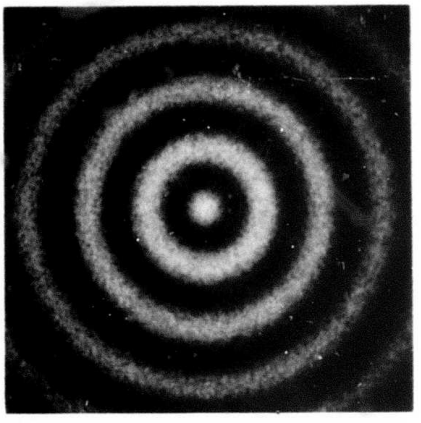
The filter of eq.(5) has been applied to an image corrupted by signal-dependent noise modelled by eq.(9). Figure 1a is the original picture. Figure 1d is a plot of the brightness cross-section of figure 1a, and the subsequent pictures are noisy and filtered images. The noise is Gaussian with variance $\sigma^2 = 4S$ where S is the signal



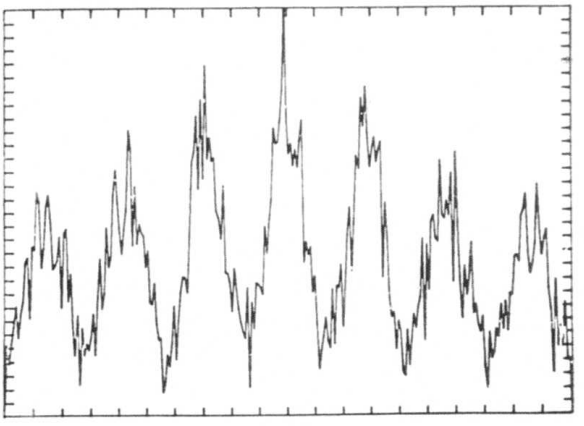
(a) Original



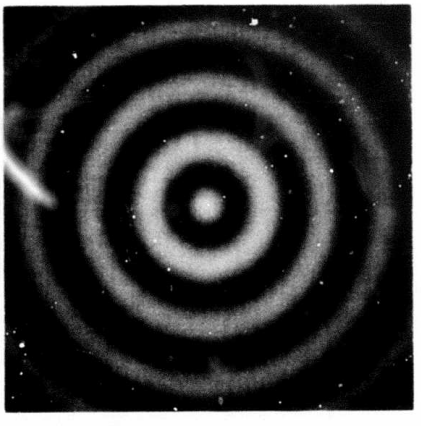
(d) Original



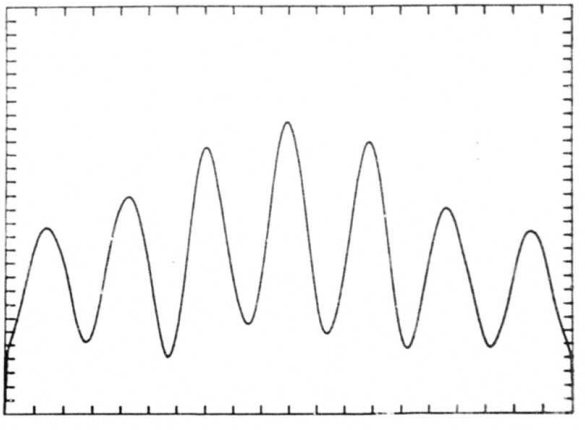
(b) Noisy



(e) Noisy



(c) Filtered



(f) Filtered

Figure 3.1-1. Filtering of signal-dependent noisy images.

intensity. The filter has reduced the mean square error by a factor of ten. The filter of eq. (13) is applied to a blurred noisy image as shown in figure 2 where the point spread function is uniform (motion degradation) and the noise is Gaussian with standard deviations of 1 and 5. The observed image has a nonzero background, and therefore, the system is undetermined. The smoothing parameter λ is set to 0.05 and 1.0 and the images are restored using thresholds $\epsilon = 0.001$ and $\epsilon = 0.01$ for the singular values of the matrix $\underline{H}^T \underline{D}^{-2} \underline{H} + \lambda \underline{Q}^T \underline{Q}^{-1} \underline{Q}^T$. As ϵ decreases, the restored image becomes sharper, but at the same time the amplitude of unwanted high frequency components increases. The quality of the restored image is highly dependant on the proper selection of λ and ϵ.

This work has demonstrated the usefulness of spline functions for digital image processing. An optimal filter with respect to the second derivative has been derived and applied to noisy and blurred images. Although the experimental results are preliminary, they show that spline functions are very useful in digital filtering and restoration. Moreover, since all the matrices are banded, these filters are computationally very efficient.

References

1. M.J. Peyrovian and A.A. Sawchuk, "Image Restoration by Smoothing Spline Functions," USCIPI Report 620, 86, September 1975.

2. C.H. Reinsch, "Smoothing by Spline Functions," Numer. Math. 10, 177, 1967.

(a) Original

(b) Noisy blurred $\sigma = 1$

(d) Noisy blurred $\sigma = 5$

(c) Restored $\lambda = .05, \epsilon = .001$

(e) Restored $\lambda = 1., \epsilon = .01$

Figure 3.1-2. Restoration of noisy blurred images.

3. P.B. Liebelt, An Introduction to Optimal Estimation, Addison-Wesley, Reading, Massachusetts, 1967.

4. R.S. Martin and J.H. Wilkinson, "Symmetric Decomposition of Positive Definite Band Matrices," Numer. Math 7, 355, 1965.

5. J.F. Walkup and R.C. Choens, "Image Processing in Signal-Dependent Noise," Optical Engineering, 13, 258, 1974.

6. B.R. Hunt, "Digital Image Processing," Proceedings IEEE, 63, 693, 1975.

7. E.L. Allgower, "Exact Inverses of Certain Band Matrices," Numer. Math. 21, 279, 1973.

8. M.J. Peyrovian and A.A. Sawchuk, "Image Processing by Smoothing Spline Functions," Proceedings of the Optical Society of America Topical Meeting on Image Processing, Asilomar, California, February 24-26, 1976.

3.2 Least Squares Image Restoration Using Spline Basis Functions

Steve Hou and Harry C. Andrews

This contribution presents a theoretical analysis and computational technique for constrained least squares image restoration using spline basis functions. A realistic continuous-discrete physical imaging model has been adopted throughout the formulation. The optical system is assumed to be incoherent, and the general problem of image restoration with space-variant or

space-invariant point-spread function degradations has been studied.

A normal equation has been formulated in a finite dimensional spline space. The magnitude of the smoothing parameter in the normal equation provides freedom to control resolution in a trade-off of smoothing the restored object. This effect has been demonstrated with experimental results. Constraints that characterize the physical properties of the restored object have been formulated so that they can be imposed on the solution of the normal equation. These constraints require the restored object pixels be positive and the energy of the object be equal to that of the degraded image.

Iterative methods for both unconstrained and constrained solutions of the normal equation have been studied in detail. Among the unconstrained methods the conjugate gradient method has been successfully simulated on a computer and strikingly good results have been obtained.

The needs of digital image restoration arise from the fact that an image in the real world is either corrupted by noise or degraded by various physical phenomena during formation, such as the diffraction limit of the optical system, sensor and display non-linearities, optical system aberrations, atmospheric turbulence, image motion blur, geometrical distortion and film grain or sensor noise.

The mathematical approach consists of finding a function f such that a known operator transforms f into a given function g. In the context of image restoration, an estimate $\hat{f}$ of the object is sought to

be as close as possible to the ideal f that produced g. There are many difficulties associated with direct minimization of $\|f-\hat{f}\|$. First the inverse of the operator may not exist or even if it exists, it is usually very hard to find. Second in most physical imaging systems the ideal object does not exist in a measurable form, but simply is the conceptual object that would be produced if there were no degradation and noise present in the imaging system. Third g is only partially known because it results from measurements. To avoid these difficulties, a certain fidelity criterion of image quality is mathematically defined in order to judge the degree of image restoration. The image restoration model is based upon continuous object-discrete image description and a Fredholm integral in two variables given by

$$\underline{G} = \int\int_{-\infty}^{\infty} \underline{H}(\xi,\eta) f(\xi,\eta) d\xi\, d\eta + N \qquad (3.2\text{-}1)$$

where $\underline{G}$ is an I x J matrix with element g_{ij}, $\underline{H}$ is an I x J matrix with element $h_{ij}(\xi,\eta)$ and $\underline{N}$ is also an I x J matrix with noise samples n_{ij} as its elements. This model describes an object scene as a continuous function in two dimensional space, as it should be in the real world, but the image at the sensor output (or at the input of a processing computer) is discrete, and has been sampled into I x J points. This is certainly a realistic model for digital image restoration systems, and will be pursued in detail throughout this report. Note that the matrix $H(\xi,\eta)$ is a point spread function matrix whose entries are continuous functions of the object plane (ξ,η) as well as the sampling position in the image plane (i,j). Clearly this point spread matrix

in its most general form can be space-variant.

To proceed it is necessary to establish certain criterion on image quality for restoration that permits control of the resolution versus noise smoothing on the final restored object. The criterion used is a modified least squares objective function. By minimizing this objective function it is possible to formulate a linear, algebraic equation for solving the unknown parameters. However these unknowns must be discrete and have finite numbers, since a digital computer can only handle a limited amount of discrete data. On the other hand an estimation is sought for a continuous object which needs infinite data points for its description. Hence an interpolation scheme must be designed to restore the continuous object from a finite set of discrete data. From a data compression point of view, this interpolation approach is attractive because it is only necessary to solve a linear, algebraic equation in finite-dimensional space instead of the integral equation of eq.(1).

Pictorial image processing usually requires handling a large amount of data. Effective iterative methods must be considered for the solution of the linear, algebraic equations derived from the minimization of the objective function. This iterative algorithm must also be tested for its speed of convergence for its usefulness for image restoration.

In the formulation of the equations for image restoration, three points must be considered: (a), tradeoffs between image resolution and noise smoothing; (b), feasibility of numerical calculations of very

large size matrices; (c), constraints imposed on the equations by a priori knowledge about the object function. To satisfy the first point the objective function consisting of the sum of the noise norm in eq.(1) and a smoothing integral is chosen for minimization. For the second point, a finite-dimensional space is chosen in which the object function $\hat{f}$ is a linear combination of the cubic B-spline functions, i.e., the bases are cubic B-splines. This choice enables conversion of an infinite-dimensional or continuous problem such as that in eq.(1) into a discrete one, to obtain computationally attractive solutions. For the third point, bounded, lossless imaging restoration mentioned above, is utilized. The positive restoration and lossless imaging assumptions add additional constraints to the solution although the solution is still not unique because of ill-conditioning.

To derive a fundamental equation for image restoration in two dimensions, the so called normal equation from the least square criterion, an objective function, which is minimum at $\hat{f}(\xi,\eta)$, is defined as

$$W(\underline{\hat{f}}) = \mathrm{Tr}[(\underline{G}-\underline{\hat{G}})^T(\underline{G}-\underline{\hat{G}})] + \gamma \int_{-\infty}^{\infty}\int [\nabla^4 \hat{f}(\xi,\eta)]^2 d\xi\, d\eta \qquad (3.2\text{-}2)$$

where

$$\nabla^4 = \frac{\partial^2}{\partial \xi^2}\,\frac{\partial^2}{\partial \eta^2} \qquad (3.2\text{-}3a)$$

$$\underline{G} = \int_{-\infty}^{\infty}\int \underline{H} f(\xi,\eta)\, d\xi d\eta \qquad (3.2\text{-}3b)$$

and

$$\hat{\underline{G}} = \int\int_{-\infty}^{\infty} \underline{H} f(\xi,\eta)\, d\xi\, d\eta \qquad (3.2\text{-}3c)$$

is a matrix that has the same size as $\underline{G}$ which is I x J. The elements g_{ij}, for i = 1 to I and j = 1 to J, of matrix $\underline{G}$ are the measured data. $\underline{H}$ is the point spread matrix, with elements $h_{ij}(\xi,\eta)$. The second term in eq. (2) is the desired smoothing measure for the class of estimated objects, $\hat{f}(\xi,\eta)$. The function $\hat{f}(\xi,\eta)$ can be linearly interpolated by a product of two one-dimensional cubic B-splines. Hence let

$$\hat{f}(\xi,\eta) = \sum_{k=1}^{K}\sum_{\ell=1}^{L} c_{k\ell}\, s_k(\xi) s_\ell(\eta) \qquad (3.2\text{-}4)$$

$$= \sum_{k=1}^{K}\sum_{\ell=1}^{L} c_{k\ell}\, \phi_{k\ell}(\xi,\eta)$$

where $s_k(\xi)$ and $s_\ell(\eta)$ are one-dimensional cubic B-splines and $\phi_{k\ell}$ is the product of s_k with s_ℓ.

The next step is to find an $\hat{f}(\xi,\eta)$ i.e., a particular set of coefficients $\{c_k\}^2$ in eq. (4) such that the objective function $W(\hat{f})$ in eq. (2) attains its minimum. This can be achieved by solving the normal equation

$$(\underline{A} + \gamma \underline{B})\, \underline{c} = \underline{d} \qquad (3.2\text{-}5)$$

where

$$\underline{A} = \int\int_{-\infty}^{\infty} \int\int_{-\infty}^{\infty} \Phi(\xi,\eta)\, \mathrm{Tr}[\underline{H}^T(\xi,\eta)\, \underline{H}(\xi'\eta')]\, \underline{\Phi}^T(\xi'\eta')\, d\xi\, d\eta\, d\xi'\, d\eta' \qquad (3.2\text{-}6a)$$

$$\underline{B} = \int\int \nabla^4 \underline{\Phi}(\xi,\eta)\, \nabla^4 \underline{\Phi}^T(\xi,\eta)\, d\xi d\eta \qquad (3.2\text{-}6b)$$

$$\underline{d} = \iint \Phi(\xi, \eta)\, \mathrm{Tr}[\underline{H}^T(\xi, \eta)\underline{G}]\, d\xi\, d\eta \qquad (3.2\text{-}6c)$$

Equation (5) deserves special attention because of its general form. It should be observed that eq. (5) can be further reduced to the form

$$\boxed{(\underline{\tilde{A}}^T\underline{\tilde{A}} + \gamma \underline{B}_x \otimes \underline{B}_y)\,\underline{c} = \underline{\tilde{A}}^T\underline{g}} \qquad (3.2\text{-}7)$$

where

$$\underline{\tilde{A}} = \left(\int\int_{-\infty}^{\infty} h_{ij}(\xi, \eta) s_k(\xi) s_\ell(\eta)\, d\xi\, d\eta \right)_{IJ \times KL} \qquad (3.2\text{-}8a)$$

$$\underline{B}_x = \left(\int_{-\infty}^{\infty} s''_k(\xi)\, s''_m(\xi)\, d\xi \right)_{K \times K} \qquad (3.2\text{-}8b)$$

$$\underline{B}_y = \left(\int_{-\infty}^{\infty} s''_\ell(\eta)\, s''_n(\eta)\, d\eta \right)_{L \times L} \qquad (3.2\text{-}8c)$$

and $\otimes$ denotes the Kronecker (direct or tensor) product.

The following mathematical properties exist for the matrices $\underline{A}$ and $\underline{B}$:

(a) $\underline{A}$ and $\underline{B}$ are both real, symmetric.

(b) $a_{pp} > a_{pq}$ and $b_{pp} > b_{pq}$ for $p \neq q$.

(c) $\underline{A}$ is non-negative definite with all positive elements.

(d) $\underline{B}$ is positive definite.

In addition, from eqs. (5a) and (8a) it is clear that $\underline{A}$ is essentially the amount of correlation or overlap of the point spread function represented in spline space. Thus the effect of the blur degradation

manifests itself in the rank or degree of $\underline{A}$ (or $\underline{\tilde{A}}^T\underline{A}$). Thus $\underline{A}$ behaves as a Gramian describing the imaging model, but in spline space. Hence $\underline{A}$ is non-negative definite. The elements $\tilde{a}_{pq}$ and a_{pq} are always non-negative because of the non-negativeness of $h_{ij}(\xi,\eta)$, $\Phi_p(\xi,\eta)$ and $\Phi_q(\xi,\eta)$. The value of the elements of $\underline{A}$ in eq.(8) depends on the PSF overlap area. For finite extended $h_{ij}(\xi,\eta)$ sparseness of matrix $\underline{A}$ can be expected. Matrix $\underline{B}$ represents the correlation or overlap of the second derivative of the spline basis functions with themselves. As a result of the separability of the basis functions, $\underline{B}$ can be represented as a kronecker product as in eq.(7). Again $\underline{B}$ represent the "Gramian" of the second derivative basis functions, but is deterministically known because of the expansion and interpolation on cubic B-spline space. This results in a banded structure. The matrices $\underline{B}_x$ and $\underline{B}_y$ are non-negative definite banded matrices.

Unconstrained Restoration: The formulation of $\underline{A}+\gamma\underline{B}$ in eq.(5) is often referred to as the Tikhonov regularization method <1>. Much Russian literature <2-5> has been devoted to the special case of $\underline{B} = \underline{I}$. Yet none has presented a complete solution for the general case. A simple minded approach for arriving at a solution is to invert $\underline{A}+\gamma\underline{B}$ directly. However there are two reasons for avoiding this direct approach. One reason is that the matrix $\underline{A}+\gamma\underline{B}$ is usually very large in image processing practice. Its inversion may require a huge memory space, which is just not available even in modern computers. The other reason is that the matrix $\underline{A} = \underline{\tilde{A}}^T\underline{A}$ is frequently ill-conditioned and influenced greatly by roundoff errors. A numerical example of this kind has been shown by Golub <6>. Adding a

positive definite matrix $\gamma\underline{B}$ to $\underline{A}$ may lessen this difficulty, but still not completely eliminate it because $\underline{B}$ is an almost periodic matrix.

In this report an iterative technique provided by the conjugate gradient method is considered. The advantages of adopting an iterative approach are the ability to monitor singularity developing during iteration, and permit human participation in the convergence process. The conjugate gradient method is useful for the solution of a system of simultaneous linear equations with an arbitrary symmetric positive definite matrix. It is especially advantageous if the matrix is not full but contains many zero elements. In the problem as posed by eq. (5), the matrix $\underline{A}+\gamma\underline{B}$ is symmetric positive definite if γ is positive and non-zero. Matrix $\underline{A}$ is sparse if the point spread function is localized; matrix $\underline{B}$ is always sparse. In the case $\gamma = 0$ and $\underline{A}$ is singular, the conjugate gradient method will give a least squares solution.

As an introduction to the gradient method consider the quadratic form

$$J(\underline{c}) = \tfrac{1}{2}\,\underline{c}^{T}(\underline{A} + \gamma\underline{B})\underline{c} - \underline{c}^{T}\underline{d} \qquad (3.2\text{-}9)$$

with the assumption that $(\underline{A}+\gamma\underline{B})$ is positive definite and symmetric. Taking the partial derivative of $J(\underline{c})$ with respect to $\underline{c}$ gives

$$\text{Grad } J(\underline{c}) = \underline{r} = (\underline{A} + \gamma\underline{B})\underline{c} - \underline{d} \qquad (3.2\text{-}10)$$

Hence the solution of a symmetric positive definite system of equations of the form $(\underline{A}+\gamma\underline{B})\,\underline{c} = \underline{d}$ is equivalent to the problem of

finding the minimum of a quadratic function of eq.(9). For this reason the gradient vector $\underline{r}$ in eq.(10) is called the residual vector.

Geometrically, the contour surfaces of constant $J(\underline{c})$ represent concentric ellipsoids. The gradient $\underline{r}$ at every $\underline{c}$ is normal to the surface at that point. Finding the minimum of $J(\underline{c})$ by varying $\underline{c}$ is equivalent to moving the vector $\underline{c}$ in a certain direction for finding the common center of the family of ellipsoids of $J(\underline{c})$ = constant.

An iterative implementation requiring no matrix inversions has been discovered. The conjugate gradient algorithm offers the following additional advantages <7,8>.

(a) Simplicity of the computational procedures.

(b) The need for only a limited amount of storage space as compared with the variable metric method <9,10>.

(c) The preservation of the original matrices of coefficients such as $\underline{A}$, $\underline{B}_x$, $\underline{B}_y$, and $\underline{G}$ during the computation and the consequent ability to exploit the advantages of having many zero elements in these matrices.

(d) The ability to start anew at any point in the computation.

(e) The superiority of each approximation to the preceding ones in the sense of being closer to the true solution.

For a more extensive description of this technique, complete details are available in USCIPI Report 650, March 1976. In addition experimental results as well as spline interpolation and constrained optimization are included in that report.

References

1. A.N. Tichonov, "Solution of Incorrectly Formulated Problems and the Regularization Method," Soviet Math. Dokl., Vol. 4, 1963, pp. 1035-1038.

2. A.N. Tichonov, "Regularization of Incorrectly Posed Problems," Soviet Math. Dokl., Vol. 4, 1963, pp. 1624-1627.

3. Y.I. Khudak, "On the Convergence of Regularizing Algorithms," Zh. Vychial. Mat. mat. Fiz., Vol. 11, No. 1, 1971, pp. 29-35.

4. Y.I. Khudak, "On the Convergence of a Family of Regularizing Algorithms," Zh. Vychial. Mat. mat. Fiz., Vol. 12, No. 2, 1972, pp. 497-502.

5. S.D. Dreizin-Dudchenko and M.K. Likht, "Approximation by Polygonal Paths in a Quadratic Metric in the Solution of Integral Equations of the First Kind," Zh. Vychial. Mat. mat. Fiz., Vol. 11, No. 5, 1971, pp. 1301-1306.

6. G.H. Golub, "Matrix Decomposition and Statistical Calculations," in Statistical Computation, edited by R.C. Milton and J.A. Nelder, Academic Press, New York, New York, 1969, pp. 365-367.

7. M.R. Hestenes, "The Conjugate Gradient Method for Solving Linear Systems," Proceedings of Symposium in Applied Mathematics and Numerical Analysis, 1955, pp. 83-102.

8. F.S. Beckman, "The Solution of Linear Equations by the Conjugate Gradient Method," in Mathematical Methods for Digital Computers, Vol. 1, John Wiley, 1967, pp. 62-72.

9. W.C. Davidon, "Variable Algorithm for Minimization," Computer J., Vol. 10, 1968, pp. 405-410.

10. W.C. Davidon, "Variable Metric Method for Minimization," ANL-5990, AEC Research Development Report, 1959.

3.3 Fast Sequential SVD Pseudoinverse Image Restoration

William K. Pratt and Ikram Abdou

Andrews and Patterson <1-3> have recently introduced the singular value decomposition (SVD) of matrices as a means of performing pseudoinverse image restoration for blurred images. Also, Pratt and Davarian <4> have developed a fast computational algorithm for pseudoinverse image restoration. This report proposes an application of the fast pseudoinverse algorithm to SVD image restoration.

Pseudoinverse Image Restoration: An imaging system resulting in the observation of a spatially blurred image in the presence of additive noise is often modelled by a set of linear equations. The vector space representation of these equations is given by

$$\underline{g} = \underline{B}\,\underline{f} + \underline{n} \tag{3.3-1}$$

where $\underline{g}$ is an M x 1 vector of pixels obtained by sampling and column scanning of the observed image, $\underline{f}$ is an N x 1 vector of points of an

ideal image field, $\underline{B}$ is the M x M blur matrix containing elements which are samples of the blur impulse response function, and $\underline{n}$ is an M x 1 vector representing additive observation noise. It will be assumed that the sample spacing on $\underline{g}$ and $\underline{f}$ is the same and that the impulse response is space invariant. Furthermore, for simplicity in explanation, the presentation is limited to a one dimensional blur. Under these conditions the blur matrix assumes the form

$$\underline{B} = \begin{bmatrix} h(L) & \cdots & h(1) & 0 & \cdots & 0 \\ 0 & h(L) & \cdots & h(1) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & h(L) & \cdots & h(1) & \end{bmatrix} \tag{3.3-2}$$

where $\underline{h}$ is an L x 1 vector of points of the impulse response. It should be noted that this system is underdetermined since N = M+L-1.

The pseudoinverse estimate of $\underline{f}$ is given by

$$\hat{\underline{f}} = \underline{B}^{-} \underline{g} \tag{3.3-3}$$

where $\underline{B}^{-}$ denotes the generalized inverse of $\underline{B}$. The generalized inverse provides a minimum least square error, minimum norm restoration of the ideal image. If $\underline{B}$ is a matrix of rank M the generalized inverse assumes the form

$$\underline{B}^{-} = \underline{B}^{T}(\underline{B}\,\underline{B}^{T})^{-1} \tag{3.3-4}$$

Algorithms requiring on the order of N^3 operations have been developed for computation of the generalized inverse.

The fast pseudoinverse algorithm <4> is based upon the development of an adjoint model

$$\underline{q}_E = \underline{C}\,\underline{f}_E + \underline{n}_E \qquad (3.3\text{-}5a)$$

where the extended vectors $\underline{q}_E$ and $\underline{f}_E$ are defined in correspondence with

$$\begin{matrix} M \\ J\text{-}M \end{matrix}\left\{\begin{bmatrix} \underline{q} \\ \hline \underline{0} \end{bmatrix}\right. = \begin{bmatrix} & \underline{C} & \end{bmatrix} \left.\begin{matrix} \\ K \\ \\ J\text{-}K \end{matrix}\right\}\begin{bmatrix} \underline{f}_T \\ \hline \underline{0} \end{bmatrix} + \begin{bmatrix} \underline{n}_T \\ \hline \underline{0} \end{bmatrix} \qquad (3.3\text{-}5b)$$

The vector $\underline{f}_T$ is composed of the center K pixels of the ideal image vector $\underline{f}$ where $K = M-2(L-1)$, and $\underline{C}$ is a circulant matrix of the form

$$\underline{C} = \begin{bmatrix} h(1) & 0 & \cdots & 0 & h(L) & \cdots & h(2) \\ h(2) & h(1) & & & & & \vdots \\ \vdots & & \ddots & & & & \\ h(L) & h(L-1) & & & & & h(L) \\ 0 & h(L) & & & & & 0 \\ \vdots & \vdots & & & & & \vdots \\ 0 & 0 & \cdots & 0 & h(L) & \cdots & h(1) \end{bmatrix} \qquad (3.3\text{-}6)$$

The vector $\underline{q}$ is identical to the image observation $\underline{g}$ over its R center elements where $R = M-2(L-1)$. The outer elements of $\underline{q}$ can be approximated by <4>

$$\underline{q} \approx \tilde{\underline{q}} = \underline{W}\,\underline{g} \qquad (3.3\text{-}7)$$

where $\underline{W}$ is a windowing matrix. Combining eqs. (5) and (7), an estimate of $\underline{f}_T$ can be obtained from

$$\hat{\underline{f}}_E = \underline{C}^{-1}\,\tilde{\underline{q}}_E \qquad (3.3\text{-}8a)$$

in correspondence with

$$\begin{matrix} K \\ J\text{-}K \end{matrix} \left\{ \begin{bmatrix} \hat{\underline{f}}_T \\ \text{---} \\ \end{bmatrix} \right. = \underline{C}^{-1} \begin{bmatrix} \tilde{\underline{q}} \\ \text{---} \\ \underline{0} \end{bmatrix} \tag{3.3-8b}$$

where $\underline{C}^{-1}$ is the inverse of the circulant blur matrix.

The restoration operation of eq. (8) can be performed indirectly by Fourier domain processing. Letting $\underline{A}$ denote the Fourier transform matrix, eq. (8) can be rewritten as

$$\hat{\underline{f}}_E = \underline{A}^{-1}\underline{\mathcal{C}}^{-1}[\underline{A}\,\tilde{\underline{q}}_E] \tag{3.3-9}$$

where both

$$\underline{\mathcal{C}} = \underline{A}\,\underline{C}\,\underline{A}^{-1} \tag{3.3-10a}$$

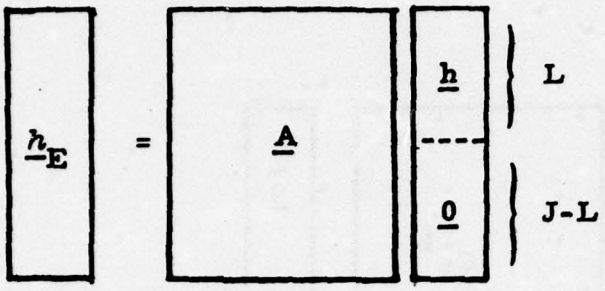
and its inverse

$$\underline{\mathcal{C}}^{-1} = \underline{A}\,\underline{C}^{-1}\underline{A}^{-1} \tag{3.3-10b}$$

are found to be diagonal matrices since $\underline{C}$ and $\underline{C}^{-1}$ are circulant. The terms of $\underline{\mathcal{C}}$ can be obtained from <4>

$$\underline{\mathcal{C}} = \text{diag}[h_E(1), \ldots, h_E(J)] \tag{3.3-11}$$

where



(3.3-12)

With the fast pseudoinverse algorithm, the brute force matrix operations of eq.(8) can be replaced by a pair of one-dimensional Fourier transforms and J scalar multiply-add operations. The computational saving is usually at least a factor of 1000:1!

SVD Image Restoration: The blur matrix $\underline{B}$ of eq.(1) can be decomposed in the product form

$$\underline{B} = \underline{U}\,\underline{\Lambda}^{\frac{1}{2}}\,\underline{V}^{T} \tag{3.3-13}$$

by a singular value matrix decomposition where $\underline{U}$ and $\underline{V}$ are unitary matrices composed of the eigenvectors of $\underline{B}\underline{B}^{T}$ and $\underline{B}^{T}\underline{B}$, respectively, and $\underline{\Lambda}$ is a diagonal matrix whose diagonal terms are the eigenvalues of $\underline{B}\underline{B}^{T}$ or $\underline{B}^{T}\underline{B}$. Since $\underline{U}$ and $\underline{V}$ are unitary matrices the generalized inverse of $\underline{B}$ can be written directly as

$$\underline{B}^{-} = \underline{V}\,\underline{\Lambda}^{-\frac{1}{2}}\,\underline{U}^{T} \tag{3.3-14}$$

and the SVD pseudoinverse estimate assumes the form

$$\hat{\underline{f}} = \underline{B}^{-}\,\underline{g} = \underline{V}\,\underline{\Lambda}^{-\frac{1}{2}}\,\underline{U}^{T}\underline{g} \tag{3.3-15}$$

The operations of eq.(15) may be viewed as a linear transformation $(\underline{U}^{T}\underline{g})$ followed by a scalar weighting (multiplication by $\underline{\Lambda}^{-\frac{1}{2}}$), followed

by another linear transformation (multiplication by $\underline{V}$). The transformations are into the eigen-space of the blur matrix $\underline{B}$. Estimation by eq. (15) requires N^2+N multiply and add operations.

As a consequence of the orthogonality of $\underline{U}$ and $\underline{V}$ it is possible to express the blur matrix in the series form

$$\underline{B} = \sum_{i=1}^{R_B} [\lambda(i)]^{\frac{1}{2}} \underline{u}_i \underline{v}_i^T \tag{3.3-16}$$

where the λ (i) are diagonal elements of $\underline{\Lambda}$, $\underline{u}_i$ and $\underline{v}_i$ are the i th rows of $\underline{U}$ and $\underline{V}$, respectively, and R_B is the rank of the matrix $\underline{B}$. The generalized inverse of $\underline{B}$ can also be written in the series form

$$\underline{B}^- = \sum_{i=1}^{R_B} [\lambda(i)]^{-\frac{1}{2}} \underline{v}_i \underline{u}_i^T \tag{3.3-17}$$

Hence, the SVD pseudoinverse estimate can be expressed as

$$\hat{\underline{f}} = \sum_{i=1}^{R_B} [\lambda(i)]^{-\frac{1}{2}} \underline{v}_i \underline{u}_i^T \underline{g} \tag{3.3-18}$$

Equation (18) can be manipulated to obtain a sequential SVD pseudoinverse estimate. The k th estimate is equal to

$$\hat{\underline{f}}_k = \hat{\underline{f}}_{k-1} + [\lambda(k)]^{-\frac{1}{2}} (\underline{u}_k^T \underline{g}) \underline{v}_k \tag{3.3-19}$$

for $1 \leq k \leq T$ where T represents the number of terms employed in the series expansion. One of the principal advantages of the sequential

formulation is that problems of ill-conditioning generally occur only for the higher order singular values. Thus, it is possible to terminate the expansion before numerical problems occur. The major difficulty associated with the sequential SVD algorithm is the large amount of computation involved in generation of the orthogonal vectors $\underline{u}$ and $\underline{v}$ and the number of operations required to implement eqs. (14) or (17).

Fast SVD Algorithm: The SVD pseudoinversion technique can be applied directly to the adjoint model of eq. (5) by the SVD expansion

$$\underline{C} = \underline{X}\,\underline{\Delta}^{\frac{1}{2}}\,\underline{Y}^{*T} \tag{3.3-20}$$

where $\underline{X}$ and $\underline{Y}$ are unitary matrices defined by

$$\underline{X}(\underline{C}\underline{C}^{T})\,\underline{X}^{*T} = \underline{\Delta} \tag{3.3-21a}$$

$$\underline{Y}(\underline{C}^{T}\underline{C})\underline{Y}^{*T} = \underline{\Delta} \tag{3.3-21b}$$

Since $\underline{C}$ is circulant, $\underline{C}\,\underline{C}^{T}$ is also circulant. Therefore, $\underline{X}$ and $\underline{Y}$ must be equivalent to the Fourier transform matrix $\underline{A}$ (or $\underline{A}^{-1}$) since the Fourier matrix performs a diagonalization of a circulant matrix. For purposes of standardization let $\underline{X} = \underline{Y} = \underline{A}^{-1}$. As a consequence, the eigenvectors $\underline{x}_i = \underline{y}_i$, which are rows of $\underline{X}$ and $\underline{Y}$, are actually the complex exponential basis functions of a Fourier transform. That is,

$$x_k^{*}(j) = \exp\left\{\frac{2\pi i}{J}(k-1)(j-1)\right\} \tag{3.3-22}$$

It is easy to show that

$$\underline{\Delta} = \underline{\mathcal{C}}\,\underline{\mathcal{C}}^{*T} \tag{3.3-23}$$

where $\underline{\mathcal{C}} = \underline{A}\,\underline{C}\,\underline{A}^{-1}$ is computed from eqs. (11) and (12).

Then in correspondence with eq. (15) the SVD pseudoinverse estimate of the extended ideal image vector becomes

$$\hat{\underline{f}}_E = \underline{Y}\,\underline{\Delta}^{-\frac{1}{2}}\,\underline{X}^{*T}\,\tilde{\underline{q}}_E \tag{3.3-24a}$$

or

$$\hat{\underline{f}}_E = \underline{A}^{-1}\underline{\Delta}^{-\frac{1}{2}}\,\underline{A}\,\tilde{\underline{q}}_E \tag{3.3-24b}$$

Equation (24b) should be recognized as exactly equivalent to the Fourier domain pseudoinverse procedure defined in eq. (9). Hence, it may be concluded that the SVD eigen space for a circulant matrix is the Fourier domain. Computation of the fast SVD estimate by eq. (24b) requires $J \log_2 J$ operations each for the forward and inverse Fourier transforms using a fast Fourier transform (FFT) algorithm plus J scalar multipliers for a total of $J(1 + 2\log_2 J)$ operations.

A series form of the fast SVD pseudoinverse estimate can be written directly in correspondence with eq. (18) as

$$\hat{\underline{f}}_E = \sum_{i=1}^{R_C} [\delta(i)]^{-\frac{1}{2}}\,\underline{x}_i^*\,\underline{x}_i^T\,\tilde{\underline{q}}_E \tag{3.3-25}$$

where R_C is the rank of $\underline{C}$. Similarly in correspondence with eq. (18) the sequential pseudoinverse estimate assumes the form

$$\hat{\underline{f}}_{E_k} = \hat{\underline{f}}_{E_{k-1}} + [\delta(k)]^{-\frac{1}{2}}\,[\underline{x}_k^T\,\tilde{\underline{q}}_E]\,\underline{x}_k^* \tag{3.3-26}$$

for $1 \leq k \leq T$ where T represents the number of terms employed in the series expansion. Using an FFT to pre-compute and store the product $[\underline{x}_k^T \tilde{\underline{q}}_E]$ results in a total of $TJ(1+\log_2 J)$ operations required for the computation. Greater computational efficiency can be realized if the series is sequentially bisected until the point of ill-conditioning is reached. This strategy requires about $J \log_2 J[1+\log_2 J]$ operations.

Fast Modified SVD Algorithm: The series and sequential SVD pseudoinverse estimates of eqs.(24) and (25) can be expressed in equivalent vector space form as

$$\hat{\underline{f}}_{E_T} = \underline{A}\,\underline{\Delta}_T^{-\frac{1}{2}}\,\underline{A}\,\tilde{\underline{q}}_E \tag{3.3-27}$$

where

$$\underline{\Delta}_T^{-\frac{1}{2}} = \begin{bmatrix} [\delta(1)]^{-\frac{1}{2}} & & & & \\ & [\delta(2)]^{-\frac{1}{2}} & & 0 & \\ & & \ddots & & \\ & & & [\delta(T)]^{-\frac{1}{2}} & \\ & 0 & & & 0 \end{bmatrix} \tag{3.3-28}$$

This estimate inherently provides a truncation of the high frequency terms of the Fourier domain pseudoinverse operator. Complete truncation to avoid ill-conditioning is often unnecessary. As an alternative to truncation the diagonal zero elements could be replaced by unity, or the value of $[\delta(k)]^{-\frac{1}{2}}$, or perhaps by some sequence that declines in value as a function of frequency.

Experimental Results: Figures 1 and 2 contain examples of the modified fast SVD pseudoinverse image restoration method. A blurred image has been formed by computer simulation of horizontal blurring with an impulse response of the form

$$h(\ell) = \exp\left\{-\tfrac{1}{2}\left[\frac{\ell - \left(\frac{L+1}{2}\right)}{\sigma_b}\right]^2\right\} \qquad (3.3\text{-}29)$$

where σ_b is the blur standard deviation and L is the length of the impulse response. The restoration algorithm involves repeated restoration with eq.(26) in which the J-T zero terms along the diagonal are replaced by $[\delta(T)]^{-\frac{1}{2}}$.

In figure 1 for blurring with $\sigma_b = 2$, there is a sudden change in the ill-conditioning effect between restorations obtained with T = 89 and T = 90 modified Fourier coefficients. The onset of ill-conditioning is not quite so pronounced in the set of restorations of figure 2 in which $\sigma_b = 3$.

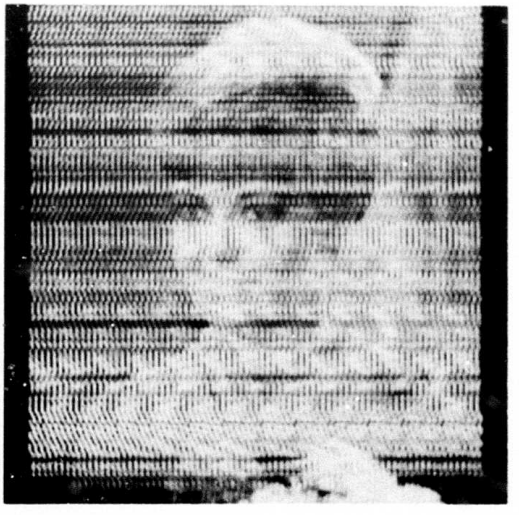
References

1. H.C. Andrews and C.L. Patterson, "Outer Product Expansions and Their Uses in Digital Image Processing," American Mathematical Monthly, Vol. 1, No. 82, January 1975, pp. 1-13.

2. H.C. Andrews and C.L. Patterson, "Outer Product Expansions and Their Uses in Digital Image Processing," IEEE Transactions on

(a) blurred observation

(b) restoration
T = 89



(c) restoration
T = 90

Figure 3.3-1. Examples of sequential SVD pseudoinverse image restoration for horizontal Gaussian blur with $\sigma_b = 2$, L=15, J=256.

(a) Blurred observation

(b) Restoration
T=58

(c) Restoration
T=60

Figure 3.3-2. Examples of sequential SVD pseudoinverse image restoration for horizontal Gaussian blur with $\sigma_b = 3$, L=23, J=256.

Computers, Vol. C-25, No. 2, February 1976, pp. 140-148.

3. H.C. Andrews and C.L. Patterson, "Singular Value Decompositions and Digital Image Processing," IEEE Transactions on Acoustics, Speech, and Signal Processing," Vol. ASSP-24, No. 1, February 1976, pp. 26-53.

4. W.K. Pratt and P. Davarian, "Fast Computational Techniques for Pseudoinverse and Wiener Image Restoration," IEEE Transactions on Computers (to be published).

3.4 Nonstationary Wiener Filtering of Film-Grain Noise

Pirouz Naderi and Alexander A. Sawchuk

Taking into account the chemical and optical degrading factors as well as the granularity noise results in a subjectively reasonable mathematical model for the formation and recording of photographic images. This model represents highly nonlinear observations, making the subsequent restoration difficult. When images are scanned with very small apertures on a microdensitometer, the noise is so severe that conventional estimators do not perform well. In these cases detection schemes suited to individual degraded pictures are much more effective. When the noise is not so severe, an adaptive minimum mean square error filter can be applied. This filter explicitly includes the nonlinear image formation effects and does not require the assumption of stationary image statistics.

Model: Basic to all estimation problems are a signal $\underline{s}$ to be

estimated, a noise $\underline{n}$ which has made the estimation necessary and an observation which is a function of the signal and the noise. Denoting the observation by $\underline{y}$ let

$$\underline{y} = g(\underline{s}, \underline{n}) \tag{3.4-1}$$

where eq. (1) is the observation model. The function g in the observation model may or may not be linear. If nonlinear, the nonlinearity may be in terms of the signal only or it may include some cross-multiplication terms involving the signal and the noise.

The statistics of the noise may depend on some parameter of the signal, thus making it signal-dependent noise. The complexity of different types of estimators designed to estimate $\underline{s}$ from the observation $\underline{y}$ is very much dependent on the complexity of the observation model. To illustrate this point, consider the minimum mean square error (MMSE) estimate of $\underline{s}$. Denoting this estimate by $\hat{\underline{s}}$ it is well known that

$$\hat{\underline{s}} = E(\underline{s} \mid \underline{y}) = \int \underline{s}\, P(\underline{s}|\underline{y})\, d\underline{s} \tag{3.4-2}$$

where $P(\underline{s}|\underline{y})$ is the a posteriori probability density function of $\underline{s}$. Clearly evaluation of the MMSE estimate of $\underline{s}$ requires evaluation of this probability density, which may prove to be quite difficult if the function g in eq. (1) is highly nonlinear in nature.

Thus, before attempting any restoration of images corrupted by film-grain noise, it is imperative to establish an analytical model

which reasonably reflects the formation and recording of images and yet is mathematically tractable.

In developing a mathematical model for any real world physical process, one faces the tradeoff between the utility of the simplified models versus the accuracy of the more complex ones. In the case of images, the degradation in the recorded image is not merely due to granularity noise, but is also caused by complex chemical and optical effects.

The model chosen for analysis, shown in figure 1, is basically an augmented version of Kelly's model <1>. The model of figure 1 has been subjectively evaluated by comparing a real world image with a similar target simulated according to this model. The results are shown in figure 2. Figure 2a shows a Kodak Pan-X film digitized using a $2(\mu m)^2$ aperture. Figures 2b and 2c show simulated targets of almost the same size, shape and contrast. Figure 2c was simulated according to the the model in figure 1, while in figure 2b only the effects of granularity were considered.

Restoration: In scanning photographic films, it is sometimes necessary to use very small apertures (on the order of 4 to 25 $(\mu m)^2$) on the microdensitometer in order to detect fine details in the image. For apertures of this size the film-grain noise is so severe that in regions of low contrast, conventional restoration techniques employing linear or nonlinear estimators will have little effect.

In such cases algorithms suited to a specific degraded photograph

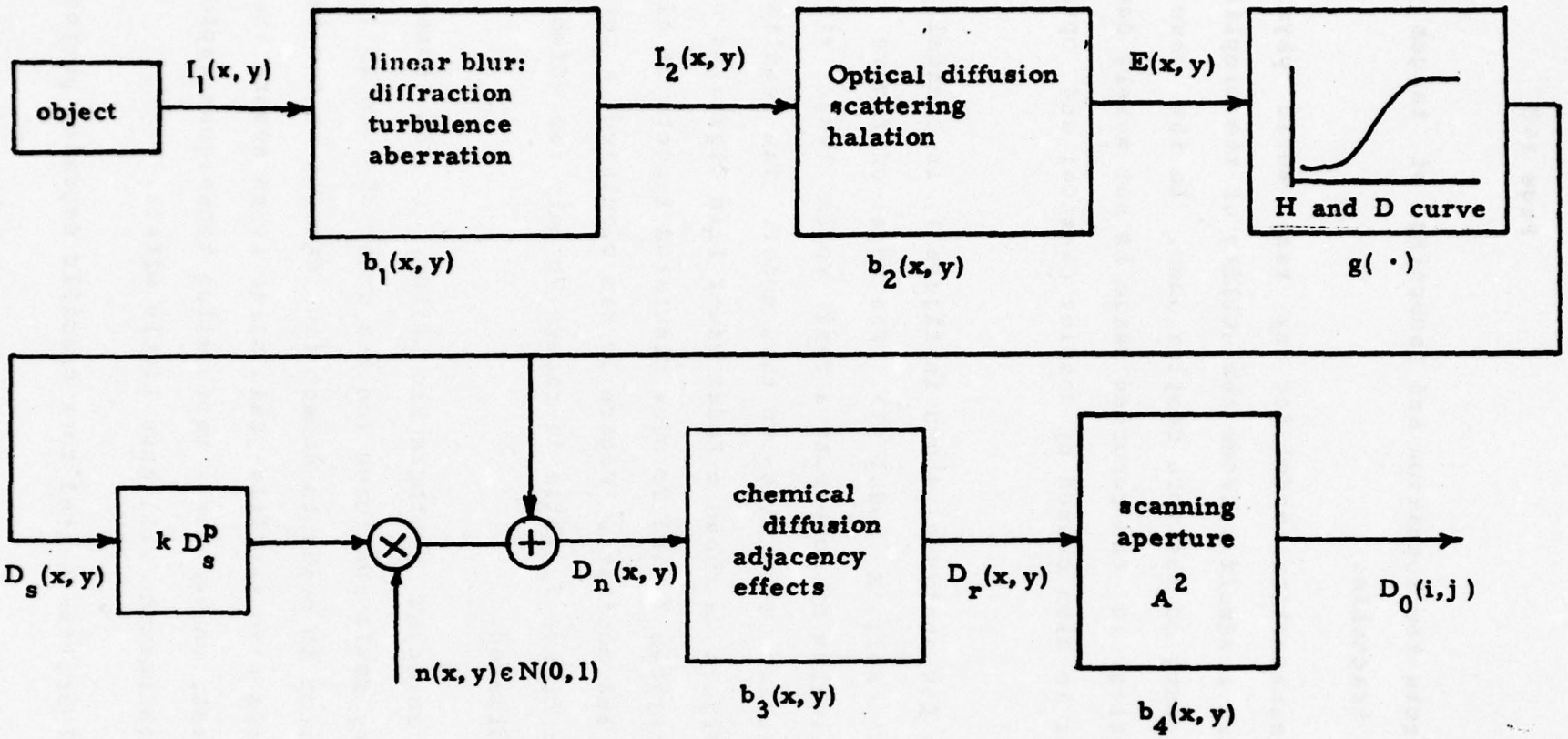
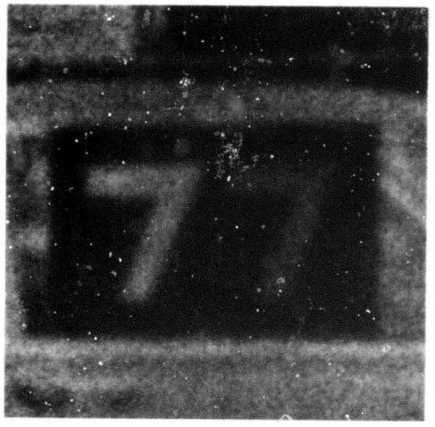
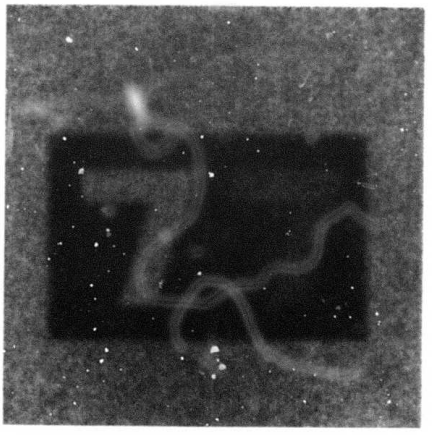


Figure 3.4-1. Suggested model for the image formation and recording system.



(a)

(b)



(c)

Figure 3.4-2. Subjective evaluation of imaging model of figure 1.

can be developed to yield very good results <3-4>. One such scheme is the noise-cheating algorithm devised by Zweig, Barrett, and Hu <2>. It has been shown that this algorithm is a sub-optimal maximum-likelihood detection scheme, but is useful as the basis of a more flexible sub-optimal Bayesian detection algorithm <4,7-8>. Although the development of this detector will not be discussed here in detail, its effectiveness is demonstrated in reference <7>. If this detection scheme is to be successful, the parameters used in designing the detector need to be accurately calculated based on each individual picture to be processed. This is in contrast with some purely statistical estimators such as Wiener filtering in which the estimator parameters are calculated in terms of the ensemble statistics of a class of images so that all pictures belonging to that class can be processed by the same estimator. The tradeoff is that these estimators have little effect when the noise is severe.

Another disadvantage of Wiener filters as applied to images is the assumption of stationarity. In general, most images of interest are not statistically stationary. Hunt <5> points out that in most cases images have a stationary second moment about a non-stationary mean.

A linear MMSE estimator based on the model of figure 1 in which the image is not required to be stationary will now be derived. Figure 3 shows the discrete version of the model in figure 1 where $\underline{H}_1$ is an M x M blur matrix which accounts for the combined effects of b_1 and b_2. Similarly, $\underline{H}_2$ is a R x M blur matrix describing blurs b_3 and

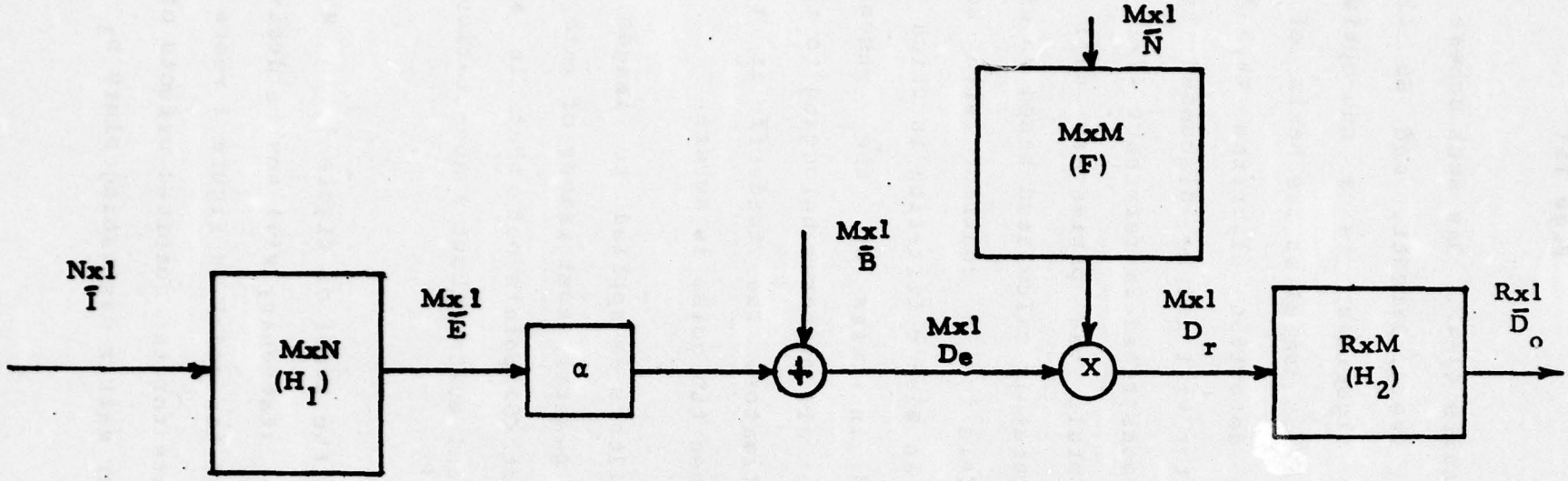


Figure 3.4-3. A discrete model for the imaging system.

b_4 of figure 1. Operation in the linear region of the H and D curve is assumed. Thus the function g in figure 1 is approximated by a linear function whose parameters are α and $\underline{B}$. Matrix $\underline{F}$ is structured so that it reflects the nonlinearity of the observation model and is of the form

$$\underline{F} = k \begin{bmatrix} (de_1)^p & & & \\ & (de_2)^p & & \\ & & \ddots & \\ & & & (de_M)^p \end{bmatrix} \qquad (3.4\text{-}3)$$

where k is a constant related to the grain size and the scanning aperture, exponent p is a constant which different experimentors have reported to be anywhere in the range 0.2 - 0.5, and $de_1,\ldots,de_M$ are elements of the non-stationary mean vector D of figure 3. Finally, vector N is taken to be a zero-mean, unit-variance multivariate normal vector which is statistically white and independent of the signal.

As it is well known, the linear MMSE estimator has the form

$$\underline{W} = (\underline{C}_{ID_o})(\underline{C}_{D_oD_o})^{-1} \qquad (3.4\text{-}4)$$

regardless of the underlying observation model, where $\underline{C}_{ID_o}$ and $\underline{C}_{D_oD_o}$ are the cross-covariance of the signal and observation, and the auto-covariance matrix of the observation, respectively. Evaluating the above two matrices from figure 3, it can be shown that

$$\underline{W} = (\alpha \underline{C}_{II} \underline{H}_1^T + \underline{I}_m \underline{B}^T + \underline{B})(\alpha^2 \underline{H}_1^T + \underline{C}_{N'N'})^{-1} \underline{H}_2^{-1} \qquad (3.4\text{-}5)$$

The form of eq.(5) is greatly simplified by setting the scaling factor

$\alpha = 1$ and $\underline{\beta} = 0$. In this case

$$\underline{W} = \underline{C}_{II}\underline{H}_1^T(\underline{H}_1\underline{C}_{II}\underline{H}_1^T + \underline{C}_{N'N'})^{-1}\underline{H}_2^{-1} \tag{3.4-6}$$

In the above equation $\underline{C}_{II}$ is the auto-covariance of the signal I which is to be estimated and $\underline{C}_{N'N'}$ is the auto-covariance of the effective noise term which is a diagonal matrix whose diagonal is related to the (2p)th moment of the signal. It has the form

$$\text{Diagonal } (\underline{C}_{N'N'}) = k^2(\underline{H}_1)\ (\underline{I})^{2p} \tag{3.4-7}$$

where

$$(\underline{I})^{2p} \equiv E[i_1^{2p},\ i_2^{2p},\ .....,\ i_N^{2p}\ ] \tag{3.4-8}$$

and where $i_1, i_2, \ldots, i_N$ are the N elements of the signal vector I. Note that if p = 0.5 then

$$\underline{I}^{2p} = \underline{I}_m = E(\underline{I}) \tag{3.4-9}$$

where $\underline{I}_m$ is the non-stationary mean of the signal vector $\underline{I}$. In this case the filter $\underline{W}$ is completely defined in the terms of the first two moments of the signal, the blur matrices $\underline{H}_1$ and $\underline{H}_2$, and the scanning constant k. If p is taken to be a more reasonable value of less than 0.5 then the (2p)th moments of the signal are required to construct the filter. It should be noted that since no restriction on stationarity is placed on the signal, then vector $\underline{I}_m$, the mean of the signal, is not a constant vector.

Application: This filter has been applied to one dimensional vectors of length 123 x 1 as shown in figure 4. Figure 4a is the

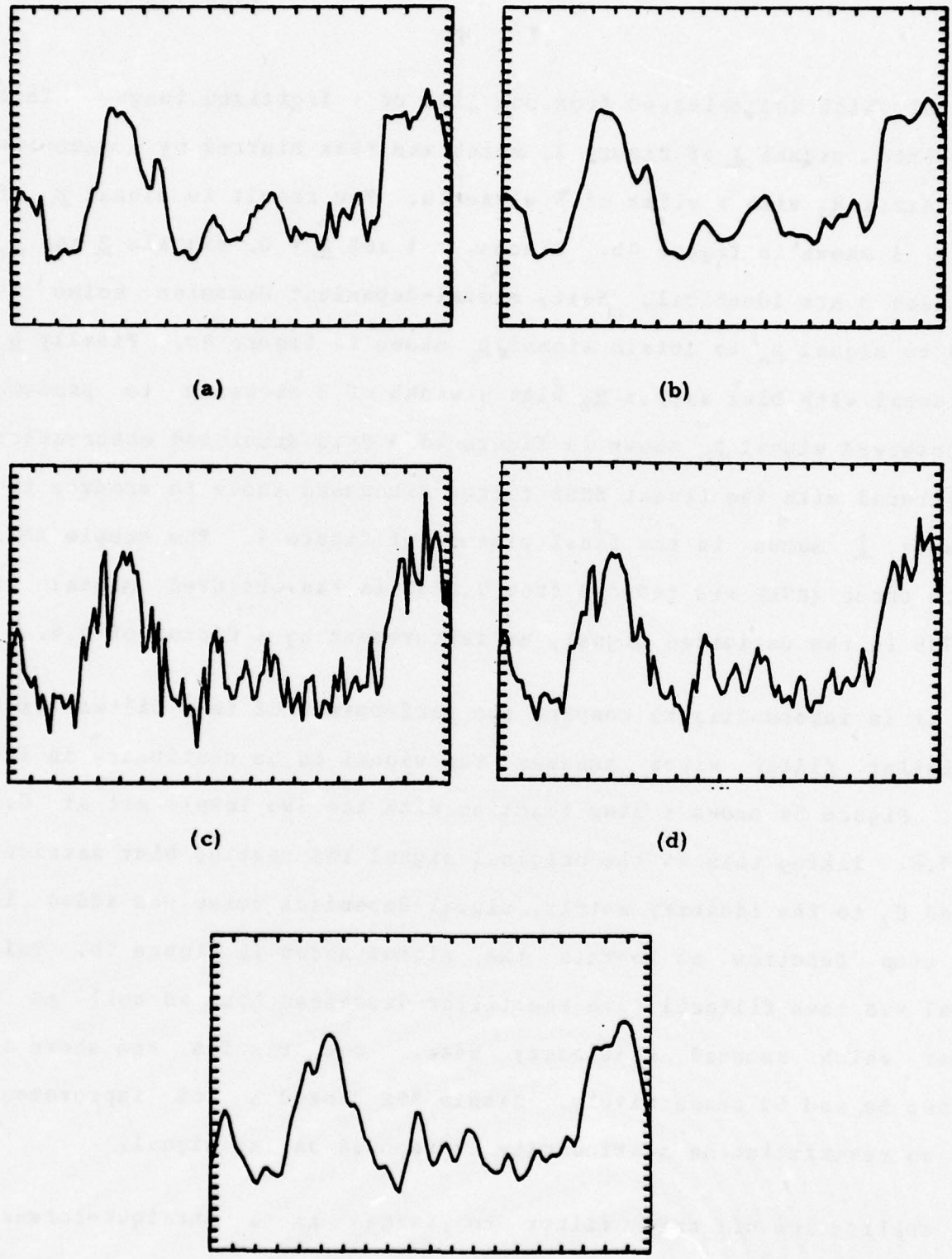
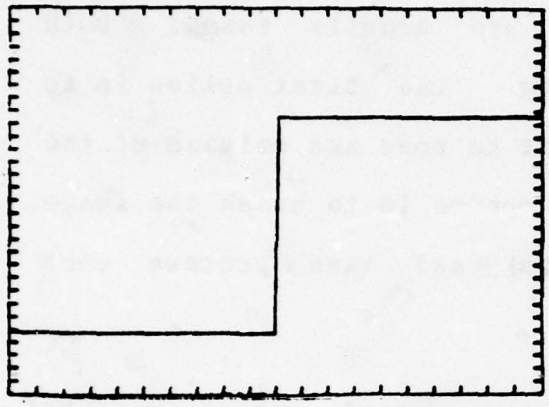


Figure 3.4-4. Application of MMSE filter to one dimensional signal.

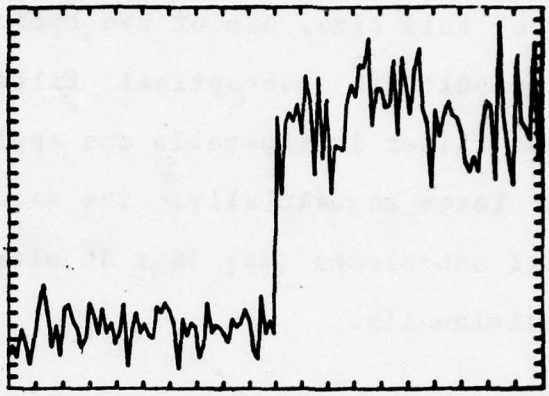
original which was selected from one line of a digitized image. This represents signal $\underline{I}$ of figure 3, which was then blurred by a Gaussian blur matrix $\underline{H}_1$ with a width of 5 elements. The result is signal $\underline{E}$ of figure 3 shown in figure 4b. Since $\alpha = 1$ and $\underline{B} = 0$, signals $\underline{E}$ and $\underline{D}_e$ in figure 3 are identical. Next, signal-dependant Gaussian noise is added to signal $\underline{D}_e$ to obtain signal $\underline{D}_r$ shown in figure 4c. Finally $\underline{D}_r$ is blurred with blur matrix $\underline{H}_2$ with a width of 3 elements to produce the observed signal $\underline{D}_o$ shown in figure 4d. This simulated observation is filtered with the linear MMSE filter discussed above to produce the estimate $\underline{\hat{I}}$ shown in the final picture of figure 4. The sample mean square error (MSE) was reduced from 0.0433 in the observed signal to 0.01798 in the estimated signal, an improvement by a factor of 2.4.

It is interesting to compare the performance of this filter with a similar filter which assumes the signal to be stationary in the mean. Figure 5a shows a step function with the two levels set at 0.4 and 1.8. Taking this as the original signal and setting blur matrices $\underline{H}_1$ and $\underline{H}_2$ to the identity matrix, signal-dependant noise was added to the step function to obtain the signal shown in figure 5b. This signal was then filtered with the filter discussed here as well as a filter which assumed stationary mean. The results are shown in figures 5c and 5d respectively. Sample MSE showed a 10% improvement when no restriction on stationarity is imposed on the signal.

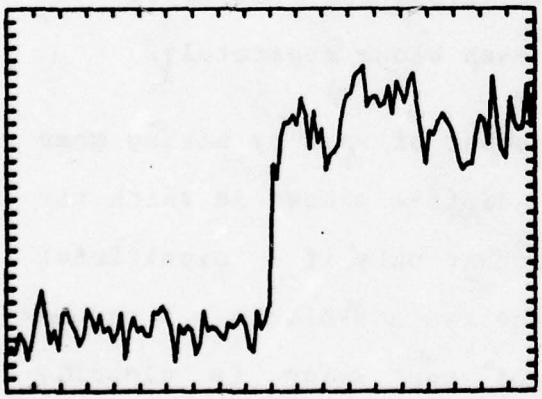
Application of this filter to images is a straight-forward extension of the one-dimensional case. The image is first lexicographically arranged in a column vector to obtain a



(a) Ideal signal



(b) Noisy signal



(c) Estimated signal not assuming stationarity

(d) Estimated signal assuming stationarity

Figure 3.4-5. Comparison of MMSE filters with and without assumption of stationarity.

one-dimensional signal <6> and then is processed by the filter. In practice, a reasonable size image of 128 x 128 pixels would correspond to a vector of size 16,384. To avoid manipulating vectors and matrices of this size, one of two options are usually taken. Both options result in sub-optimal filtering. The first option is to assume the filter is separable and apply it to rows and columns of the degraded image sequentially. The second option is to break the image into small sub-blocks (say 16 x 16 elements) and then process each block individually.

A few computational aspects for the second option can be considered at this point. Since the mean of the image is assumed non-stationary, the mean of each individual 16 x 16 sub-block differ from one another. Examining eqs.(6), (7) and (9) it is clear that the matrix $\underline{C}_{N'N'}$ will differ for each 16 x 16 sub-block and therefore the filter $\underline{W}$ of eq.(6) must be calculated for each block separately.

The computation can be reduced in a number of ways by making some simple assumptions. One way is to use an adaptive scheme in which the filter is updated from one sub-block to another only if a significant difference exists between the means of the two sub-blocks. A second method is to assume the overall mean of the image is globally non-stationary, but is locally stationary within each elemental 16 x 16 block. This means that within each block the mean can be taken to be a constant. Equations (7) and (9) indicate that if this is taken to be the case, then matrix $\underline{C}_{N'N'}$ of eq.(6) can be written in the form

$$\underline{C}_{N'N'} = \gamma(\underline{I}) \qquad (3.4\text{-}10)$$

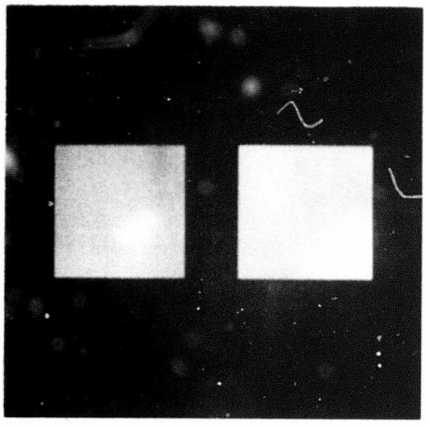
where $\underline{I}$ is the identity matrix and γ is a constant. In this case the only change in filter $\underline{H}$ of eq.(6) from one block to another is caused only by the change in γ. Inspecting eq.(6) it is noted that most of the calculation in forming filter $\underline{H}$, is due to the inversion

$$\underline{K} \equiv (\underline{H}_1 \underline{C}_{II} \underline{H}_1^T + \underline{C}_{N'N'})^{-1} \qquad (3.4\text{-}11)$$

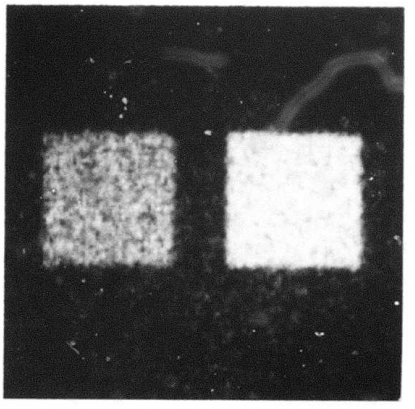
The other inversion to form $\underline{H}_2$ needs to be calculated only once since $\underline{H}_2$ does not change from one block to another.

If the assumption of local stationarity is made such that eq.(10) holds, then a simple way exists to perform the inversion of eq.(11). This is done by using a similarity transformation on matrix $(\underline{H}_1 \underline{C}_{II} \underline{H}_1^T)$ to reduce it to a diagonal form. Since matrix $\underline{C}_{N'N'}$ is of the form given in eq.(10) the similarity transformation would leave it diagonal and eq.(11) is reduced to the inversion of a diagonal matrix at each block. This filter was tested on a simulated image and the results are shown in figure 6.

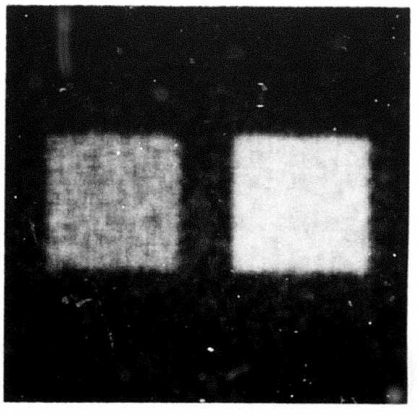
Figure 6 represents one of the first attempts in applying the discussed MSE filter to two dimensional signals. In obtaining the noisy image of figure 6 the two matrices $\underline{H}_1$ and $\underline{H}_2$ were set to the identity matrix. As dictated by the non-stationary mean of the signal, three different MMSE filters were obtained to operate on the rows of the noisy image. The estimated image is shown in figure 6c.



(a) Ideal image



(b) Noisy image



(c) Estimated image

Figure 3.4-6. Application of MMSE filter to two dimensional signal.

More results in application of this filter to images are forthcoming. It is expected that if filtering is done on sub-blocks rather than rows of the degraded image to take advantage of image correlation in both horizontal and vertical direction better results can be obtained.

Conclusions: It has been demonstrated that when dealing with low contrast images and severe noise, some picture dependent detection schemes perform better than filters designed solely on ensemble statistics. It has also been shown that there is an advantage in designing MMSE filters with no stationarity of the mean imposed on the image.

References

1. D.H. Kelly, "Systems Analysis of Photographic Processing I.A. Three-Stage Model," Journal Optical Society of America, Vol. 50, 1960, p. 259.

2. H.J. Zweig, E.B. Barrett and P.C. Hu, "Noise-Cheating Image Enhancement," Journal Optical Society of America, Vol. 65, 1975, p. 1347.

3. H.J. Zweig, E.B. Barrett and P.C. Hu, "Comparison of Noise-Cheating and Wiener Filtering," Journal Optical Society of America, Vol. 65, 1975, p. 1203A.

4. F. Naderi and A.A. Sawchuk, "Nonlinear Digital Processing of

Images Corrupted by Film-Grain Noise," Journal Optical Society of America, Vol. 65, 1975, p. 1202A.

5. B.R. Hunt, "Digital Image Processing," Proceedings IEEE, Vol. 63, 1975, p. 693.

6. H.C. Andrews, Computer Techniques in Image Processing, Chapters 5 and 6, Academic Press, New York, 1970.

7. F. Naderi and A.A. Sawchuk, "Detection and Estimation of Images Degraded by Film-Grain Noise," USC Image Processing Institute, Semiannual Technical Report 620, September 30, 1975, pp. 92-100.

8. F. Naderi and A.A. Sawchuk, "Nonlinear Detection and Estimation of Images Degraded by Film-Grain Noise," Proceedings Optical Society of America Topical Meeting on Image Processing, Asilomar, California, February 1975, Journal Optical Society of America, Vol. 66, 1976, p. 167.

3.5 Degrees of Freedom in Image Formation

Dennis McCaughey and Harry C. Andrews

The concept of degrees of freedom (DOF) has arisen in the imaging literature in an attempt to quantify the number of truly independent samples of data one gathers with photographic or other image sensing devices <1,2>. As sensor technology grows, the quantity of data gathered increases, and it becomes reasonable to ask what the true increase in information content is as one increases image samples. One useful model, in which the degrees of freedom (DOF) concept fits

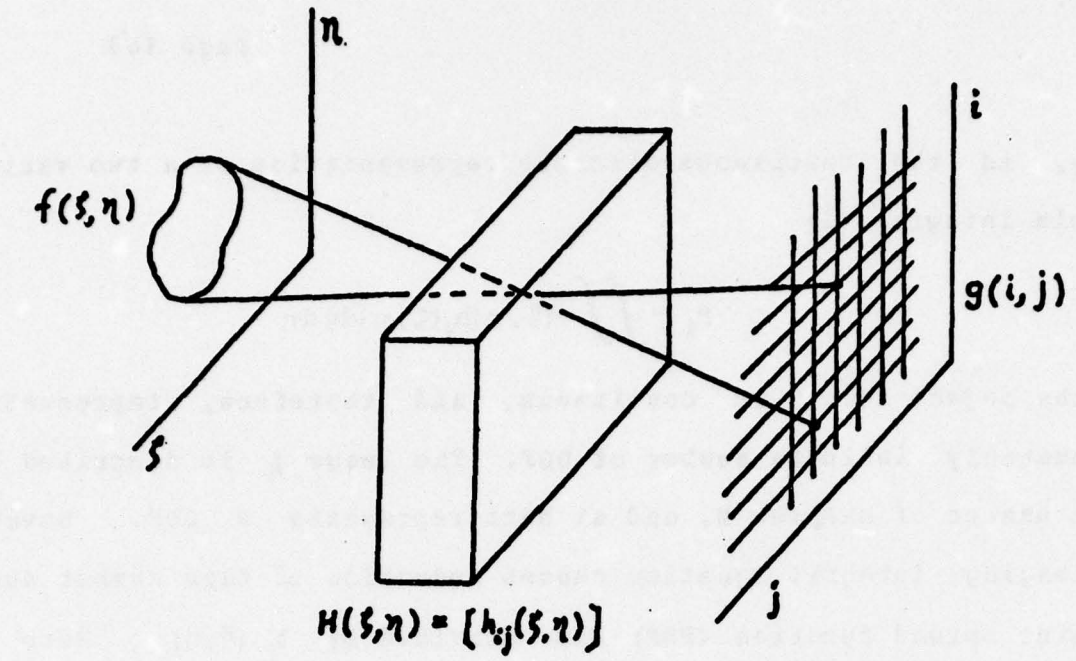
nicely, is the continuous-discrete representation of a two variable Fredholm integral <3>

$$g_i = \iint f(\xi,\eta)h_i(\xi,\eta)d\xi\,d\eta \qquad (3.5\text{-}1)$$

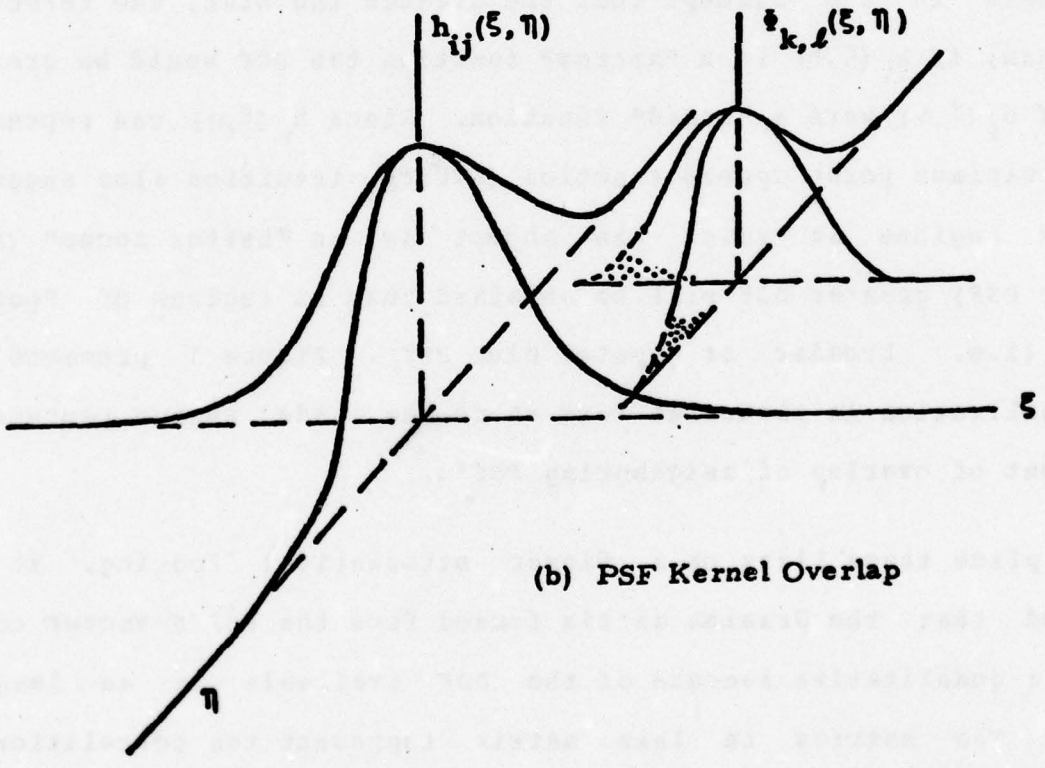
Here the object $f(\xi,\eta)$ is continuous, and therefore, represents a non-countably infinite number of DOF. The image g_i is described by a finite number of samples N, and at best represents N DOF. However, the imaging integral equation causes reduction of this number due to the point spread function (PSF) blur provided by $h_i(\xi,\eta)$. Here the PSF is described by an N-vector whose elements are continuous functions of the object coordinate system (ξ,η). Intuition might serve here in the concept that the greater the blur, the fewer the DOF. Thus, if $h_i(\xi,\eta)$ is a "narrow" function the DOF would be greater than if $h_i(\xi,\eta)$ were a "broad" function. Since $h_i(\xi,\eta)$ can represent a space variant point spread function (SVPSF), intuition also suggests that in regions in which the object is in "better focus" (i.e. narrower PSF) greater DOF will be obtained than in regions of "poorer focus" (i.e. broader or greater blur PSF). Figure 1 presents the conceptualization in pictorial form where the shaded region represents the amount of overlap of neighboring PSF's.*

To place these ideas on a firmer mathematical footing, it is suggested that the Gramian matrix formed from the PSF N-vector could provide a quantitative measure of the DOF available in an imaging system. The entries in this matrix represent the correlation or

*This figure has been provided by Dr. H.S. Hou <4> whose research aided the authors in developing some of these concepts.



(a) Continuous Discrete Model



(b) PSF Kernel Overlap

Figure 3.5-1. Continuous-Discrete Model with PSF Kernel Linear Dependence

overlap of each PSF with its neighbors, and the rank and/or eigenvalue map of such a matrix could be used in the definition of the DOF <1,2,5>. Twomey has equated the degrees of freedom of an imaging system with the number of effectively non-zero eigenvalues of the Gramian of the imaging system <2>. Twomey's techniques were developed for one-dimensional systems, but are applicable to two-dimensional systems and will be suitably extended and applied to projection imaging in what follows.

Continuous-Discrete Model and the Gramian: In applying two-dimensional linear systems methods to imaging systems the assumption is made that an image g is related to the original object f by a superposition integral

$$g(x,y) = \int\int_R h(x,y;\xi,\eta) f(\xi,\eta)\, d\xi\, d\eta \qquad (3.5\text{-}2)$$

where $h(x,y;\xi,\eta)$ represents the weighting function of the imaging system and R is the region of integration over the input coordinate system. The aim of any reconstruction process is to recover f as best as possible in some sense, knowing h and g. In a practical situation g is known only for a finite set of points in the output plane. Let g_i denote $g(x,y)$ sampled in the image coordinate plane (x,y) and similarly for h, then the vector form of equation (1), the continuous-discrete imaging model

$$\underline{g} = \int\int_R \underline{h}(\xi,\eta) f(\xi,\eta)\, d\xi d\eta + \underline{\delta} \qquad (3.5\text{-}3)$$

results where $\underline{\delta}$ is an error term and all vectors are N x 1 column vectors.

In obtaining an estimate $\hat{f}$ for f in eq.(3), consideration must be taken of the fact that the PSF kernel vector $\underline{h}$ is comprised of kernel functions that are not necessarily independent. Neglecting the error term for the moment, this dependency of kernel functions implies that some of the output samples can be predicted by linear combinations of the others, are thus superfluous, and serve to reduce the DOF of the imaging system. The error term aggravates this situation since, if an output element can be predicted by a linear combination of the others to within an accuracy better than the measurement error, that measurement adds no new information to what is already known. Twomey applied this reasoning to systems involving one-dimensional integral equations with significant results <2>.

To obtain a better feeling for how the DOF of the system might be reduced due to this linear dependence amongst PSF kernels, an arbitrary unit length vector is defined such that its dot product with the imaging equation yields

$$<\underline{\alpha},\underline{g}> = <\underline{\alpha}, \int\int_R \underline{h}(\xi,\eta) f(\xi,\eta) d\xi\, d\eta> + <\underline{\alpha},\underline{\beta}> \qquad (3.5\text{-}4)$$

If the first term on the right is much smaller than the second, then its contribution can be considered minimal. Taking this into account along with the fact that $\|\alpha\| = 1$, it is possible to reduce the degree of independence or number of DOF by one for every independent α such that

$$\left\| <\underline{\alpha}, \int\int_R \underline{h}(\xi,\eta) f(\xi,\eta) d\xi\, d\eta> \right\|^2 \leq \|\underline{\beta}\|^2 \qquad (3.5\text{-}5)$$

where $\|\beta\|^2$ can be considered the sensor noise which contributes to the

definition of the signal to noise ratio of the imaging system. The implications of eq.(5) can be further investigated by noting that the vector inner product on the left can be related to the continuous inner product of f with h by

$$\left\| <\underline{\alpha}, \iint_R \underline{h}(\xi,\eta) f(\xi,\eta) d\xi\, d\eta > \right\|^2 = <f(\xi,\eta), <\underline{h}(\xi,\eta), \underline{\alpha}>>^2 \qquad (3.5\text{-}6)$$

and applying the Schwartz inequality

$$<\underline{f}(\xi,\eta), <\underline{h}(\xi,\eta), \underline{\alpha}>>^2 \leq <\underline{f}(\xi,\eta), \underline{f}(\xi,\eta)><<\underline{h}(\xi,\eta), \underline{\alpha}>, <\underline{h}(\xi,\eta), \underline{\alpha}>>$$

$$= \|f(\xi,\eta)\|^2 \, \underline{\alpha}^* \underline{\Gamma}\, \underline{\alpha} \qquad (3.5\text{-}7)$$

Here $\underline{\Gamma}$ is known as the Gramian of the PSF kernels and is defined as

$$\underline{\Gamma} = \left[\iint_R h_i(\xi,\eta) h_j^*(\xi,\eta) d\xi\, d\eta \right] \qquad (3.5\text{-}8)$$

which is Hermitian having real eigenvalues only. Since $f(\xi,\eta)$ can be considered bounded, $|f(\xi,\eta)| \leq E$, eq. (5) becomes

$$E^2 \, \underline{\alpha}^* \underline{\Gamma}\, \underline{\alpha} \leq \|\underline{g}\|^2 \qquad (3.5\text{-}9)$$

Expanding $\underline{\Gamma}$ into its eigenvector decomposition $\underline{\Gamma} = \underline{U}\underline{\Lambda}\underline{U}^*$ and letting $\underline{\beta} = \underline{U}\underline{\alpha}$, another unit vector, gives

$$E^2 \underline{\beta}^* \underline{\Lambda}\, \underline{\beta} \leq \|\underline{g}\|^2 \qquad (3.5\text{-}10)$$

Clearly $\underline{\beta}^* \underline{\Lambda}\, \underline{\beta}$ is minimized by the allowing $\underline{\beta}$ to be the eigenvector associated with the smallest eigenvalue of $\underline{\Gamma}$; which leads to the conclusion, that for every eigenvalue of $\underline{\Gamma}$ such that

$$E^2 \lambda_i < \|\underline{g}\|^2 \qquad (3.5\text{-}11)$$

one degree of freedom is lost.

Attention can now be directed to the estimate of the reconstruction of the object $\hat{f}$ by expanding this estimate in a set of N orthogonal functions $\{\phi(\xi,\eta)\}$ such that

$$\hat{f}(\xi,\eta) = \sum_{i=1}^{N} a_i \varphi_i(\xi,\eta) = \langle \underline{\varphi}(\xi,\eta), \underline{a} \rangle \tag{3.5-12}$$

In order that $\hat{f}$ be a solution up to the degree of freedom of $\underline{\Gamma}$ the orthogonal set will be defined as

$$\underline{\varphi}(\xi,\eta) = \underline{U}^{*}\underline{h}(\xi,\eta) \tag{3.5-13}$$

Then

$$\iint_{R} \underline{\varphi}(\xi,\eta)\underline{\varphi}^{*}(\xi,\eta)d\xi\, d\eta = \iint_{R} \underline{U}^{*}\underline{h}(\xi,\eta)\underline{h}^{*}(\xi,\eta)\underline{U}\, d\xi\, d\eta = \underline{U}^{*}\underline{\Gamma}\,\underline{U} = \underline{\Lambda} \tag{3.5-14}$$

Thus, the set is orthogonal. Solving for the imaging equation for $\hat{f}$ without noise yields

$$\underline{g} = \iint_{R} \underline{h}(\xi,\eta)\underline{\varphi}^{*}(\xi,\eta)\underline{a}\, d\xi d\eta \tag{3.5-15a}$$

$$\underline{g} = \iint_{R} \underline{h}(\xi,\eta)\underline{h}^{*}(\xi,\eta)\underline{U}\,\underline{a}\, d\xi\, d\eta \tag{3.5-15b}$$

$$= \underline{\Gamma}\,\underline{U}\,\underline{a} = \underline{U}\,\underline{\Lambda}\underline{U}^{*}\underline{U}\,\underline{a} = \underline{U}\,\underline{\Lambda}\,\underline{a} \tag{3.5-15c}$$

Consequently the coefficients of the estimate f from eq. (12) become

$$\underline{a} = \underline{\Lambda}^{-1}\,\underline{U}^{*}\,\underline{g} \tag{3.5-16}$$

Thus,

$$\hat{\underline{f}}(\xi,\eta) = \underline{g}^{*}\,\underline{\Gamma}^{-1}\underline{h}(\xi,\eta) \tag{3.5-17}$$

and the estimate of the object is shown to be directly related to the image and expansion of the PSF kernel in the space of the Gramian (weighted by the inverse of the Gramian eigenvalues). Clearly, the larger the condition number of $\underline{\Gamma}$, the poorer the estimate of the object. A large condition number on $\underline{\Gamma}$ implies low eigenvalues which implies an effective linear dependence in the point spread function kernels which implies overlapping point spread functions in the imaging model.

The results of eqs. (16) and (17) indicate an ill-conditioning problem in which an upper bound on the relative error $\|\delta\underline{a}\|/\|\underline{a}\|$ can be found to be

$$\frac{\|\delta\,\underline{a}\|}{\|\underline{a}\|} \leq \frac{|\lambda_{max}|}{|\lambda_{min}|}\;\frac{\|\delta\underline{g}\|}{\|\underline{g}\|} \tag{3.5-18}$$

Thus the condition number of $\underline{\Gamma}$ and the noise to signal ratio in the data define an upper bound on the relative error in the reconstruction process.

Heretofore consideration has been given to the case of a non-singular Gramian, and attention must be given to the case where $\underline{\Lambda}^{-1}$, and thus $\underline{\Gamma}^{-1}$, do not exist. Letting $\underline{C}$ be a non-singular constraint

matrix this problem can be postulated as a suitable minimization as follows <13>:

$$\text{minimize: } \underline{a}^* \underline{C}\, \underline{a} \tag{3.5-19a}$$

$$\text{subject to: } \|\underline{g} - \underline{U}\, \underline{\Lambda}\, \underline{a}\|^2 = 0 \tag{3.5-19b}$$

The criterion to be minimized becomes

$$\underline{a}^* \underline{C}\, \underline{a} + \gamma \|\underline{g} - \underline{U}\, \underline{\Lambda}\, \underline{a}\|^2 \tag{3.5-20}$$

where γ is the usual Lagrange multiplier. Performing this minimization results in a least squares solution of

$$\hat{\underline{a}} = (\gamma^{-1} \underline{C} + \underline{\Lambda}^2)^{-1}\, \underline{\Lambda}\, \underline{U}^* \underline{g} \tag{3.5-21a}$$

and

$$\hat{f}(\xi, \eta) = \underline{a}\, \underline{U}^* \underline{h}(\xi, \eta) \tag{3.5-21b}$$

the above result should be compared with eqs. (15) and (17) to see the implications of this least squares result. Proper conditions on $\underline{C}$ and γ result in the pseudoinverse reconstruction.

References

1. Twomey, S., "The Application of Numerical Filtering to the Solution of Integral Equations Encountered in Indirect Sensing Measurements," Journal of the Franklin Institute, Vol. 279, 1965, pp. 95-109.

2. Twomey, S., "Information Content in Remote Sensing," Applied

Optics, Vol. 13, No. 4, 1974, pp. 942-945.

3. Andrews, H.C. and Patterson, C.L., "Singular Value Decompositions and Digital Image Processing," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-24, No. 1, February 1976, pp. 26-53.

4. Hou, H.S., "Least Squares Image Restoration Using Spline Interpolation," Ph.D. dissertation, USCIPI Report No. 650, University of Southern California, Los Angeles, California, March 1976.

5. Hou, H.S. and Andrews, H.C., "Fundamental Limits and Degrees of Freedom of Imaging Systems," Proceedings of the Optical Society of America Topical Meeting on Image Processing, Asilomar, California, February 1976.

3.6 Texture Image Coding

William K. Pratt

Many images can be segmented into a relatively few number of regions. Some regions, called busy regions, contain objects possessing fine detail. Other regions relatively devoid of detail, can often be described as being of similar composition or texture. Textured regions are characterized by some quasi-periodic repetition of a basic pattern or structure. Examples include grass, farm fields, gravel, tiles, brick walls, and textiles.

Coding Concept: Segmentation of an image into busy regions and textured regions can be utilized as the basis of an image coding concept. The coding process is as follows:

(a) an image is segmented into two region classes: textured and busy;

(b) busy regions are identified, and pixels within the region are coded by some high quality coding method such as predictive or transform coding;

(c) texture is characterized according to some measure within each textured region;

(d) at the image destination busy regions are reconstructed from their coded representations;

(e) the textured regions are reconstructed at the receiver by inserting artificially generated texture within each texture segment.

This texture coding process seeks to exploit the limitations and adaptability of the human visual process. In most scenes textured regions simply provide the spatial context for fine detail busy regions. The visual system does not closely examine textured regions; they are perceived on some global basis. In effect the fine structure of a textured region is redundant to the human visual systems. This visual property has been exploited by artists for centuries. In many paintings textured regions are represented by relatively few brush strokes in comparison to regions of fine detail.

Segmentation: The texture coding system requires segmentation of an image into busy and textured regions prior to coding of each region. Image segmentation for purposes of scene analysis is presently an active field of research. Techniques under study include:

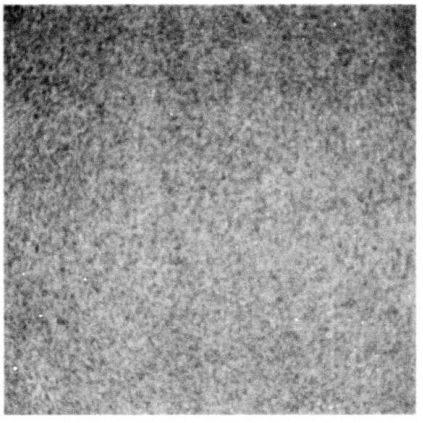
thresholding
histogram division
clustering
region growing
edge linking

Thresholding is a simple form of segmentation in which individual pixels are classified according to their amplitude value. The process is limited to simply structured images for which the textured background possesses an average brightness level significantly different from the busy regions. Histogram division is a multi-dimensional form of thresholding based upon the formation of histograms of image attributes such as luminance, tristimulus value, edge density, etc. If a histogram is multi-modal, the division levels between the modes are used to define classification cells in the multi-dimension image attribute space. Clustering techniques of pattern recognition are a generalization of histogram division in which classification cells are formed from the joint histograms of the image attributes rather than the marginal histogram of each attribute. The histogram and clustering approaches often perform quite well in segmentation, but the associated processing task is formidable.

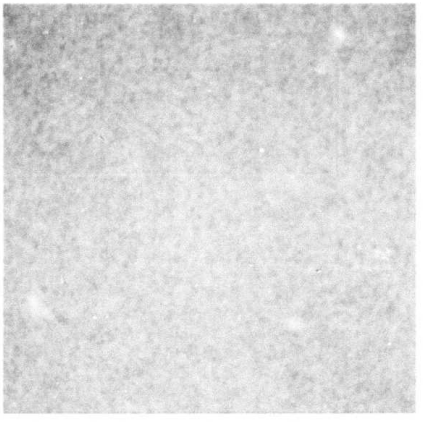
Region growing is a simple approach to segmentation in which neighboring pixels are grouped into a region if they possess similar attributes. The method works well for simple scenes, but must be augmented with semantic knowledge for reasonably complex scenes. Edge linking is a boundary formation process in which edge points are individually detected and subsequently linked to their neighbors if they possess similar gradients and orientations. The major drawback of the technique is that the detected boundaries may not be closed. Further work is needed toward the development of segmentation methods which can be utilized in the texture coding system.

Texture Synthesis: There are two basic approaches to texture synthesis: deterministic and statistical. With the deterministic method a basic pattern of pixels is periodically repeated in two dimensions over an image field. The pattern may be an artificial arrangement of pixel amplitude values or some small segment of a natural scene. The statistical approach involves the generation of a spatially correlated two dimensional random process.

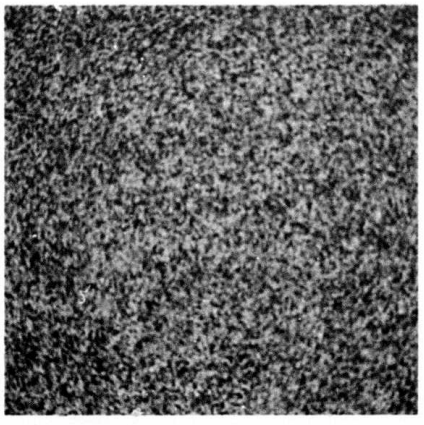
Figure 1 contains examples of artificially generated textures whose average luminance level is 200 over a scale of 0 to 255. The first example in figure 1a consists of a spatially uncorrelated array generated by Gaussianly distributed random numbers with standard deviations of 32. In figure 1b the array of figure 1a has been low pass filtered by convolution with an impulse response function



(a) Gaussian density
$\overline{Y}=200, \sigma_y=16$



(b) Low pass filtering of (a)



(c) High pass filtering of (a)
$\rho_C = \rho_R = 0.99$

Figure 3.6-1. Artificially generated textures.

$$\underline{H} = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{3.6-1}$$

The array of figure 1c results from high pass filtering with an impulse response function

$$\underline{H} = \begin{bmatrix} \rho_C\rho_R & -\rho_C(1+\rho_R^2) & \rho_C\rho_R \\ -\rho_R(1+\rho_C^2) & (1+\rho_C^2)(1+\rho_R^2) & -\rho_R(1+\rho_C^2) \\ \rho_C\rho_R & -\rho_C(1+\rho_R^2) & \rho_C\rho_R \end{bmatrix} \tag{3.6-2}$$

where $\rho_C = 0.5$ and $\rho_R = 0.9$.

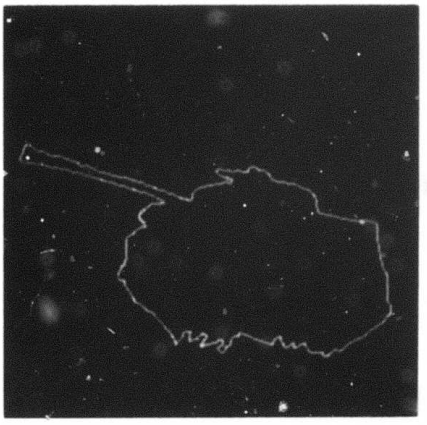
Experimental Results: The basic concept of texture coding has been explored for the coding of the tank image of figure 2a. The outline of the tank, shown in figure 2b, has been obtained manually to provide the basis for the segmentation of the tank and the grass. In figure 2c the tank segment has been combined with a constant luminance background of value 200. The reconstruction of figure 2d utilizes the low pass filtered Gaussian texture array.

The results presented here are merely a preliminary example of the general concept of texture coding. Effort is continuing on several fronts to completely develop the technique. Research is being performed on texture segmentation, measurement, and synthesis.

3.7 Algorithms and Non-ideal Effects in Nonlinear Optical Processing

Stephen R. Dashiell and Alexander A. Sawchuk

(a) Tank



(b) Tank outline

(c) Reconstruction with grey background

(d) Reconstruction with texture background

Figure 3.6-2. Example of texture coding.

The use of halftone screen preprocessing to realize nonlinear operations in a coherent optical system is well established <1-7>. Various operations such as logarithm, exponential, level slicer, notch filter, quantization, and analog-to-digital conversion have been presented. When halftone screens are used to realize monotonic or non-monotonic nonlinear point transfer functions, the screens are usually composed of periodically repeated cells which are themselves monotonic in density. This is not necessary: an iterative computer algorithm for generating a non-monotonic cell shape to yield an arbitrary transfer function in the first order has been developed. There is no unique solution to the non-monotonic design problem. The cell generated by the algorithm is only one of many possibilities. Some operations, specifically the notch filter and quantizer <2,4,6,7> are realized with non-monotonic halftone cells.

Omitting wavelength and geometrical factors for clarity, the bar structure of period a on the copy film resulting from using a one-dimensional halftone screen can be analyzed by representing one period U(x) as a sum of small shifted rectangle functions. If there are ℓ small rectangle functions, then

$$U(x) = \sum_{k=1}^{\ell} I(k, I_{in}) \operatorname{rect}\left(\frac{x - ka/\ell + a/2\ell}{a/\ell}\right) \qquad (3.7\text{-}1)$$

where ℓ can be arbitrarily large, $I(k,I_{in}) = 0$ if the specific sub-region is opaque, and $I(k,I_{in}) = 1$ if it is clear. Equation (1) is convolved with a comb function <8> to produce the complete periodic transmittance function input to the optical processor. After Fourier

transforming, diffraction order selection, and discarding of complex constants, the inverse transform is formed by the second lens in the system and the magnitude squared yields the output intensity

$$I_{out} = g(I_{in}, n) = \left| \frac{1}{\ell} \sum_{k=1}^{\ell} I(k, I_{in}) \exp(-j2\pi kn/\ell) \right|^2 \qquad (3.7\text{-}2)$$

where n represents the diffraction order selected. For the first order, n = 1. In both eqs. (1) and (2), the term $I(k, I_{in})$ reflects the fact the the sub-regions depend on I through the halftone preprocessing.

The synthesis problem for an arbitrary diffraction order can be stated: choose a function $I(k, I_{in})$ so that the square of the n th Fourier component of $I(k, I_{in})$ is equal to the desired $g(I_{in}, n)$ with the restriction that if $I(k, I_{in})$ equals one for any I_{in}, it must be equal to one for all larger I_{in}. The constraint arises because no sub-region can return to clear after being exposed.

Considering only the first order, eq. (2) becomes

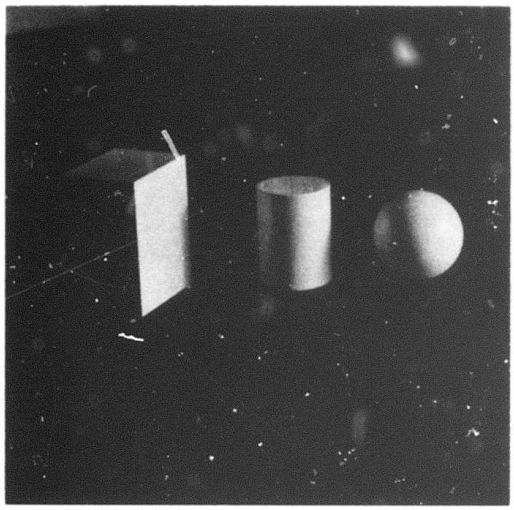
$$I_{out} = g(I_{in}, 1) = \left| \frac{1}{\ell} \sum_{k=1}^{\ell} I(k, I_{in}) \exp(-j2\pi k/\ell \right|^2 \qquad (3.7\text{-}3)$$

This can be regarded as a sum of ℓ vectors of length 0 or $1/\ell$ with the angular orientation of the k th vector given by

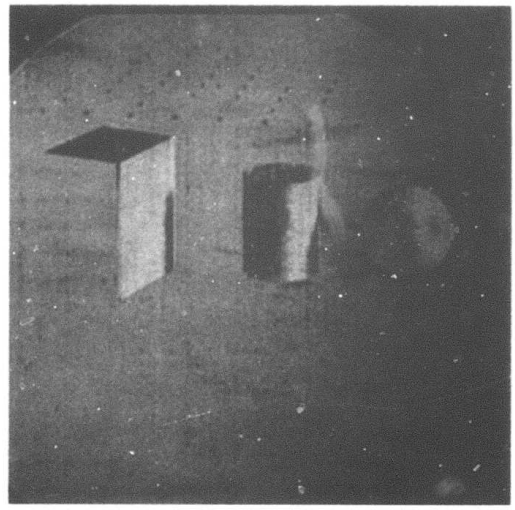
$$\theta(k) = \frac{2\pi k}{\ell} \qquad (3.7\text{-}4)$$

These are directly equivalent to positions ka/ℓ in the periodic cell.

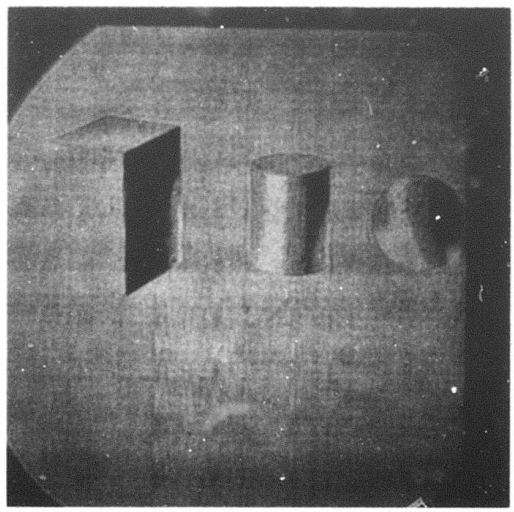
A synthesis algorithm for non-monotonic cells to achieve arbitrary non-monotonic nonlinear transfer functions can be formulated. The procedure is to break the vectors into two families, those for which $1 \leq k \leq \ell/2$ and those for which $(\ell/2)+1 < k < \ell$, and regard the first family as the one which increases the output, and the second as the one which decreases the output. To begin, add in vectors starting with $k = 1$ until the sum squared equals the desired I_{out} for the I_{in}. To increase I_{out} for a larger I_{in}, add in more vectors one at a time and continue until the sum squared equals the desired I_{out}. To decrease I_{out} for a larger I_{in}, add in vectors starting with $k = \ell/2+1$ or, if some of the second family have already been used, start at the vector adjacent to the one last used. Add these in until the sum squared has decreased to the desired value. Continue this process until the complete nonlinear transfer function has been mapped out. One of two possibilities will occur: (1) all of the vectors in the first family will be used before the complete transfer function has been mapped out; or (2), the function will be completely mapped and there will be vectors remaining which could have been used. If the first possibility occurs, increase the length of all the vectors and start over. If the second possibility occurs, decide if a large enough fraction of the available vectors (cell area) have been used, and stop. If not, decrease the vector length and begin again. If the tolerances are chosen wisely, this procedure will arrive at a satisfactory solution fairly rapidly. Figure 1a shows the input picture, and figure 1b shows the output picture for a first order intensity notch filter with a non-monotonic cell halftone screen.



(a) Original input picture



(b) Notch filtered output



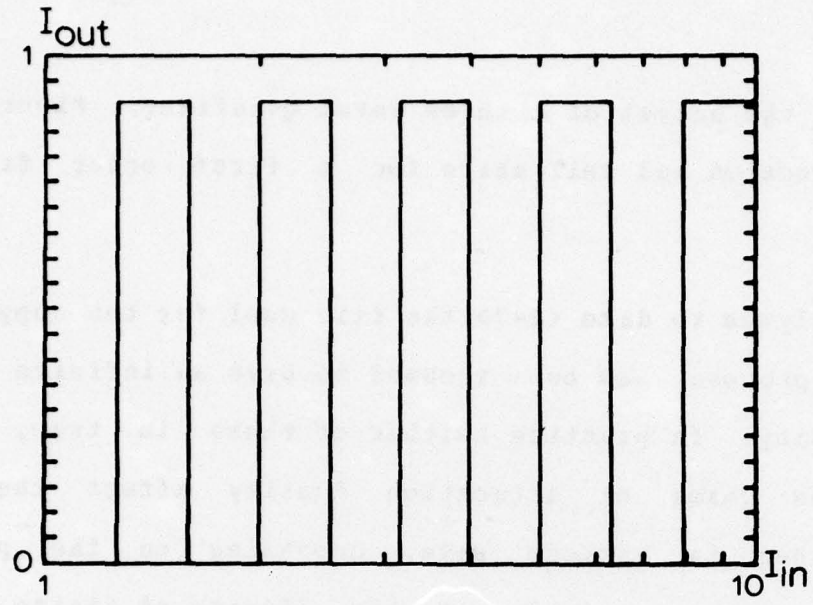
(c) Three level quantizer

Figure 3.7-1. Nonlinearities from non-monotonic halftone cells.

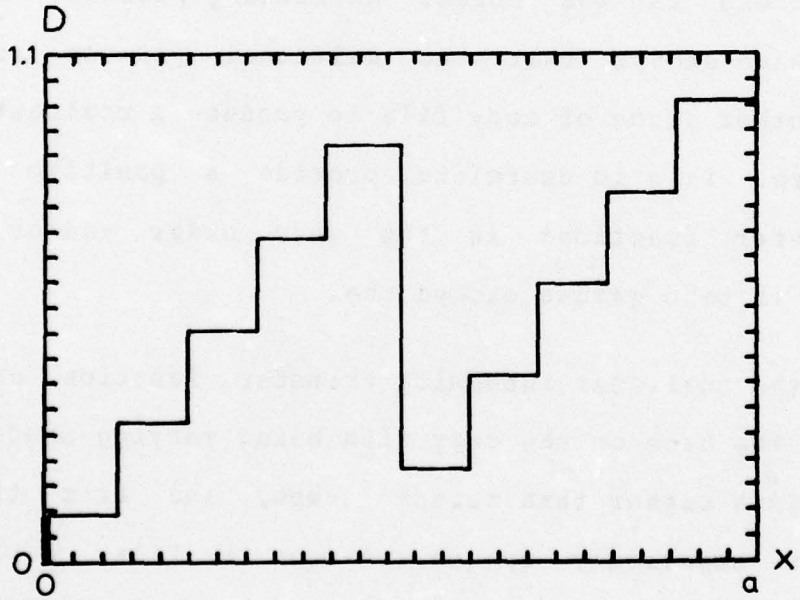
Figure 1c shows the output of a three level quantizer. Figure 2 shows the transfer function and cell shape for a first order five level slicer.

In the analyses to date <1-7> the film used for the copy step in the halftone process has been assumed to have an infinite gamma and saturation density. In practice neither of these is true, and the effects of low gamma or saturation density affect the transfer functions obtained in various ways, depending on the particular halftone screen. An analysis of the effects of finite gamma and saturation density for one-film and two-film processes will be given. A one-film process is the normal halftoning process. A two-film process is the same except that the halftoned picture is contact printed onto another piece of copy film to produce a contrast reversed halftoned picture. This is useful to provide a positive slope on monotonic transfer functions in the zero order and/or increase effective gamma if both gammas exceed one.

Errors in the nonlinear intensity transfer function arise from the edges of the bars on the copy film being varying shades of grey with a finite width rather than abrupt jumps, and from the finite transmission of supposedly opaque areas on the film. A plot of the film density across one bar will show sloping rather than abrupt sides, and a maximum density equal to the saturation density for the film. If this bar density is regarded as consisting of a stack of k trapezoids each of height ΔD, the output intensity resulting from using an arbitrary n th order component of the Fourier transform of a



(a) Desired transfer function



(b) Nonmonotonic cell for operation in the first diffraction order.

Figure 3.7-2. Five level slice nonlinearity.

periodic array of such bars spaced a apart in a two lens optical processing system is

$$I_{out}(D_1, D_{sat}, \gamma, n) = \left| DC + \sum_{i=1}^{k} \left(\sqrt{10^{\frac{(-i-1)D'}{k}}} - \sqrt{10^{\frac{-iD'}{k}}} \right) \cdot \left(\frac{a - f^{-1}\left(D_1 - \frac{(i-1)D'}{\gamma k}\right) - f^{-1}\left(D_1 - \frac{iD'}{\gamma k}\right)}{a} \right) \right.$$

$$\operatorname{sinc} \frac{n\left(f^{-1}\left(D_1 - \frac{(i-1)D'}{\gamma k}\right) - f^{-1}\left(D_1 - \frac{iD'}{\gamma k}\right)\right)}{a} \cdot$$

$$\left. \operatorname{sinc} \frac{n\left(a - f^{-1}\left(D_1 - \frac{(i-1)D'}{\gamma k}\right) - f^{-1}\left(D_1 - \frac{iD'}{\gamma k}\right)\right)}{a} \right|^2 \qquad (3.7\text{-}5)$$

where D_1 = halftone cell density corresponding to this bar width

D_{sat} = film saturation density

$D' = D_{sat}$ or γD_1, whichever is smaller

k = number of stacked trapezoids used

n = diffraction order used

f^{-1} = inverse of f where $f(x) = D$ is the cell profile description

$DC = \sqrt{10^{-D'}}$ for n = 0 and 0 for n = 0.

Realizing that $I_{in} = \log_{10} D_1$, eq. (5) can be evaluated for various D_1 and a picture of the actual I_{out} for various D_{sat} and γ values obtained.

Equation (5) describes the effects for the one-film case. For the two-film case

$$I_{out}(D_1, D_{sat2}, D_{max}, \gamma_1, \gamma_2, n) = \left| DC + \sum_{i=1}^{k} \left(\sqrt{10^{-\gamma_2 (i-1)\Delta D}} + \sqrt{10^{-\gamma_2 i \Delta D}} \right) \cdot \frac{1}{n\pi} \cdot \right.$$

$$\left(\frac{f^{-1}\left(D_1 - \frac{\left(\frac{D'}{\gamma_2} - (i-1)\Delta D\right)}{\gamma_1} \right) + f^{-1}\left(D_1 - \frac{\left(\frac{D'}{\gamma_2} - i\Delta D\right)}{\gamma_1} \right)}{a} \right)$$

$$\operatorname{sinc}\left(\frac{n\left(f^{-1}\left(D_1 - \frac{\left(\frac{D'}{\gamma_2} - i\Delta D\right)}{\gamma_1} \right) - f^{-1}\left(D_1 - \frac{\left(\frac{D'}{\gamma_2} - (i-1)\Delta D\right)}{\gamma_1} \right) \right)}{a} \right)$$

$$\left. \operatorname{sinc}\left(\frac{n\left(f^{-1}\left(D_1 - \frac{\left(\frac{D'}{\gamma_2} - (i-1)i\Delta D\right)}{\gamma_1} \right) + f^{-1}\left(D_1 - \frac{\left(\frac{D'}{\gamma_2} - i\Delta D\right)}{\gamma_1} \right) \right)}{a} \right) \right|^2 \qquad (3.7\text{-}6)$$

where $\Delta D = D'/K\gamma_2$

D_{sat2} = saturation density of film no. 2

$D_{max} = \log_{10}(I_{wpy}/I_2')$

I_{copy} = copying intensity for copying film no. 1 onto film no. 2

I_2' = threshold intensity on film no. 2

γ_1 = film no. 1 gamma

γ_2 = film no. 2 gamma

n = diffraction order selected

D' = D_{sat2} or D_{max}/γ_2, whichever is smaller

$DC = \sqrt{10^{-D'}}$ for n = 0 and 0 for n = 0

Equation (6) is used to calculate a point by point plot of I_{out} for various I_{in} values, given the other specified parameters.

Equations (5) and (6) have such a large number of variables that a comprehensive series of graphs showing all possible behavior modes is impossible in the space available. The plots shown assume high saturation density and low gammas. Figures 3 to 6 show examples of one and two film transfer functions.

It is possible to minimize certain degradations caused by finite gamma and saturation densities by precompensation of the halftone cell. The precompensation can be accomplished for portions of the transfer function which are neither flat nor an abrupt jump. The technique is as follows:

(a) Calculate or specify the ideal transfer function.

(b) Using the specified film parameters, calculate point by point what the required bar width is to obtain the desired output using eq.(5) or (6) as appropriate, with the restriction that the calculated cell must be monotonic, i.e., $f(x) = D$ is a monotonically increasing function of x.

Note that when the maximum allowable bar width is reached, further increases in I serve simply to increase the overall film density.

Figures 7 and 8 show some results from this technique. Note from figure 8 that even with low gamma and saturation density, a good match to the desired transfer function is possible.

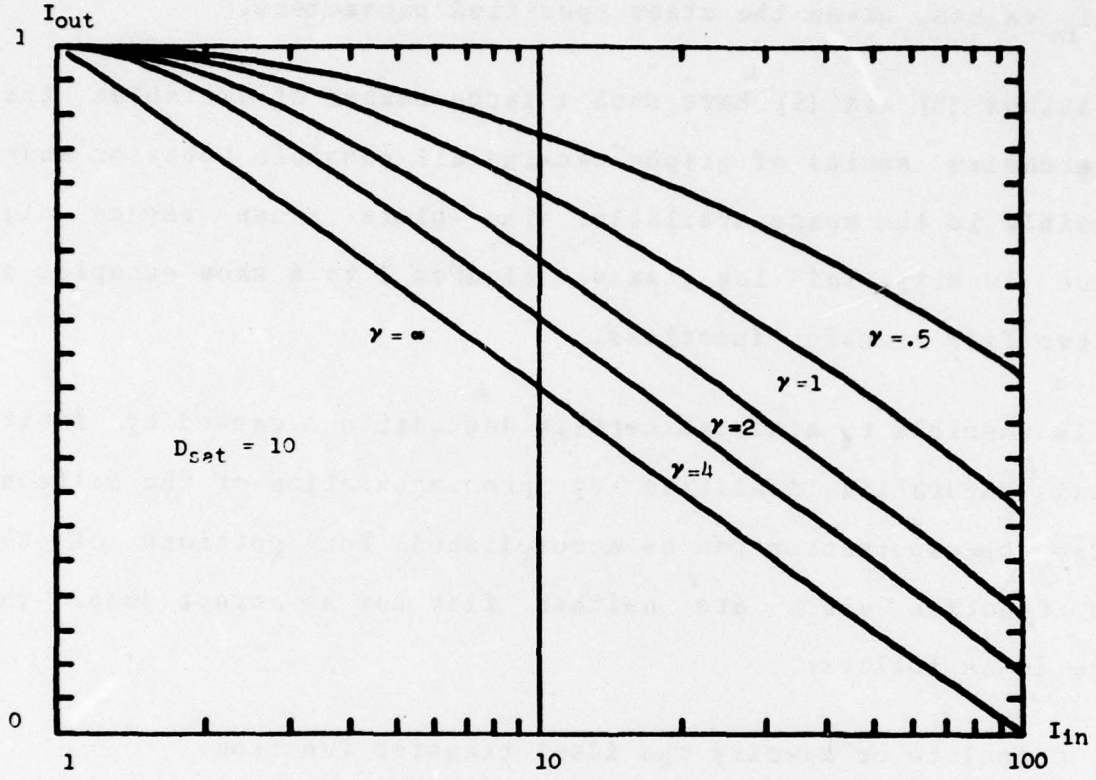


Figure 3.7-3. One-step process monotonically decreasing logarithm transfer function.

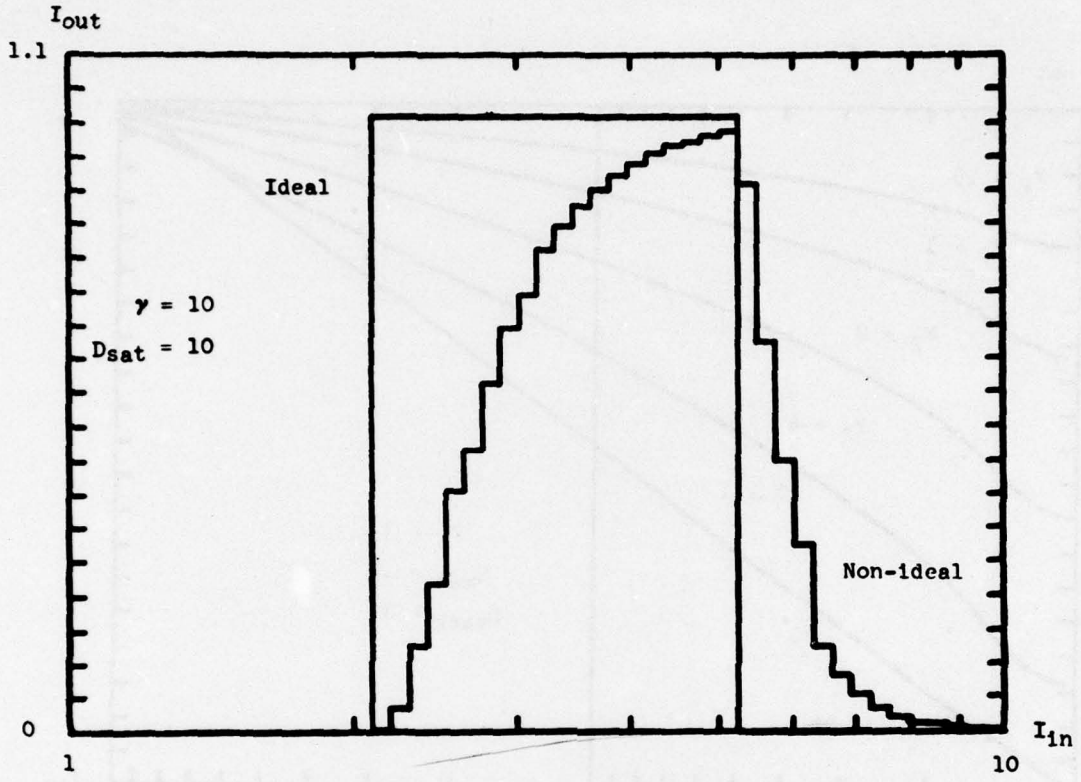


Figure 3.7-4. One-step level slice, $\gamma = 10$.

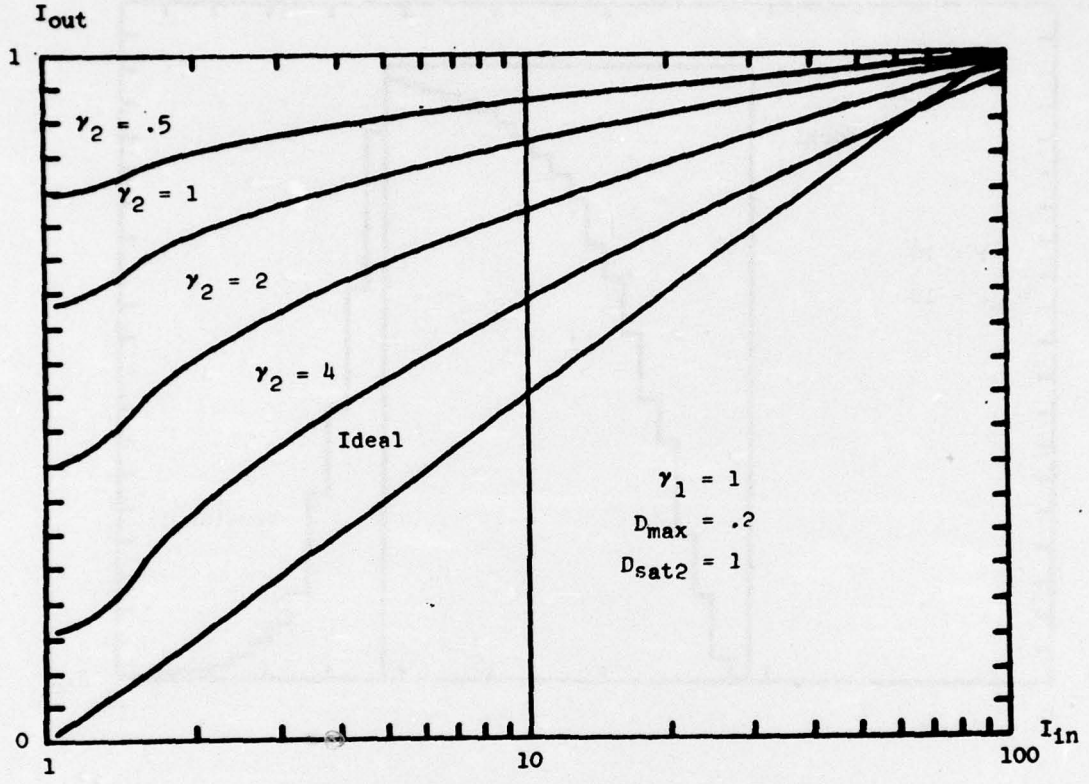


Figure 3.7-5. Two-step process monotonically increasing logarithm transfer function.

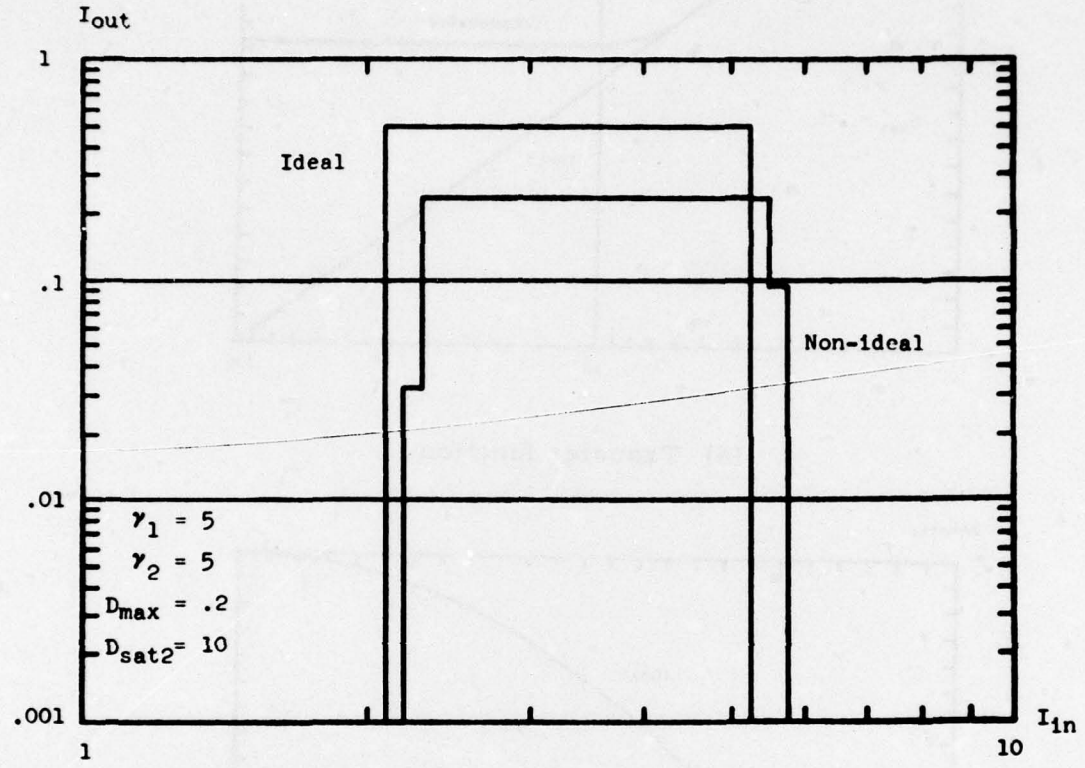
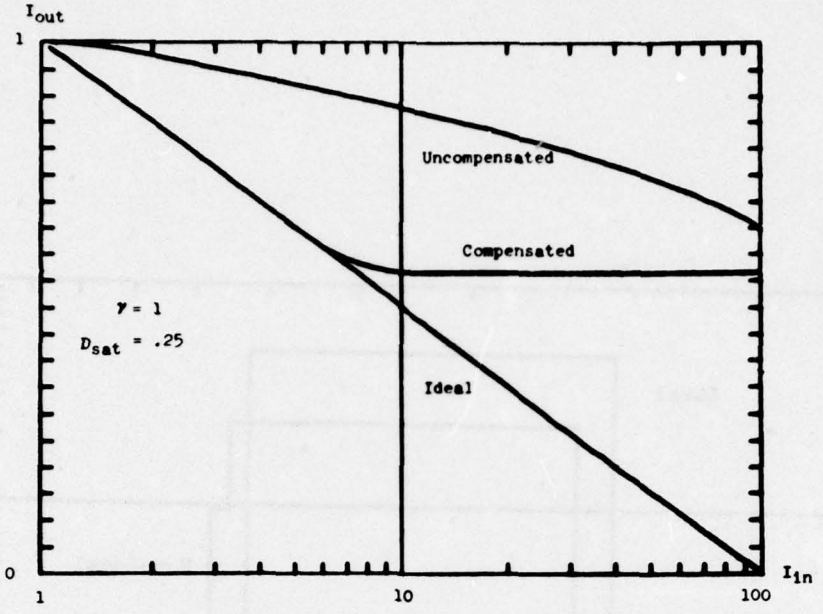
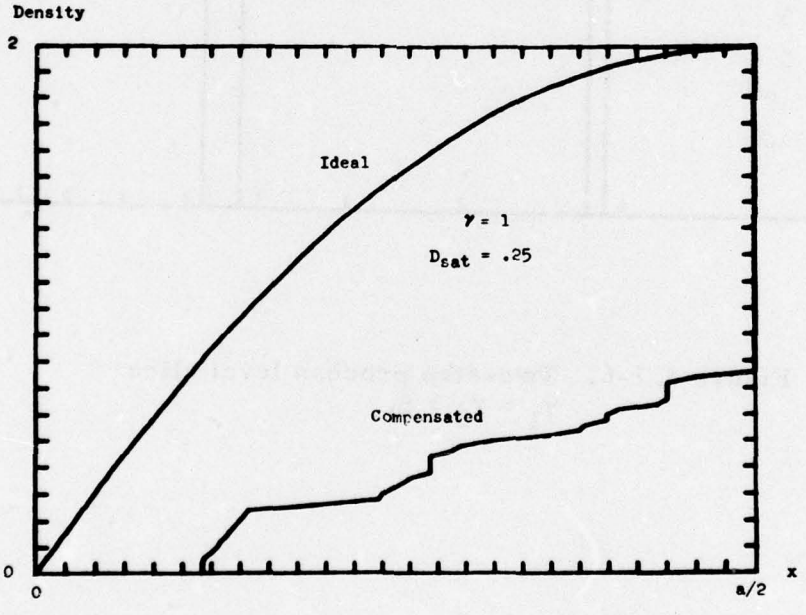


Figure 3.7-6. Two-step process level slice $\gamma_1 = \gamma_2 = 5$.

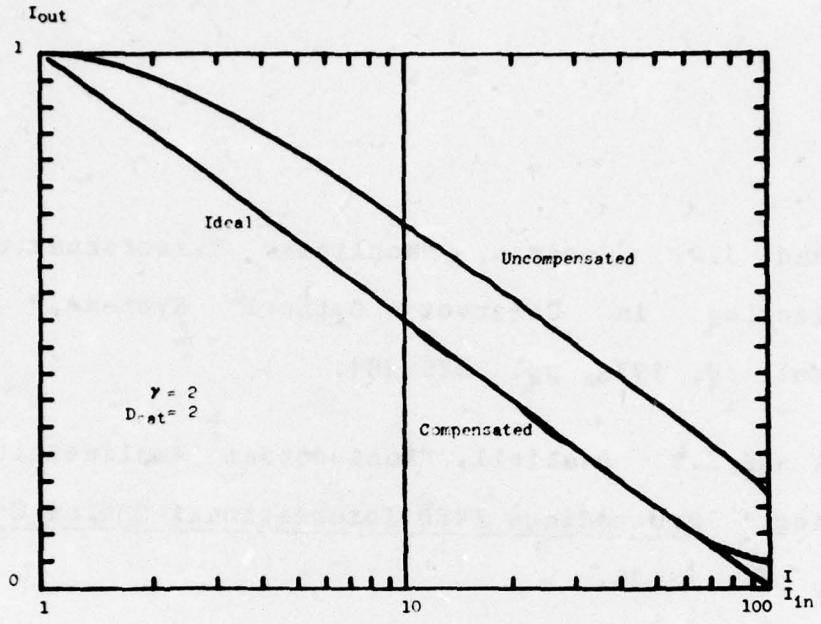


(a) Transfer function

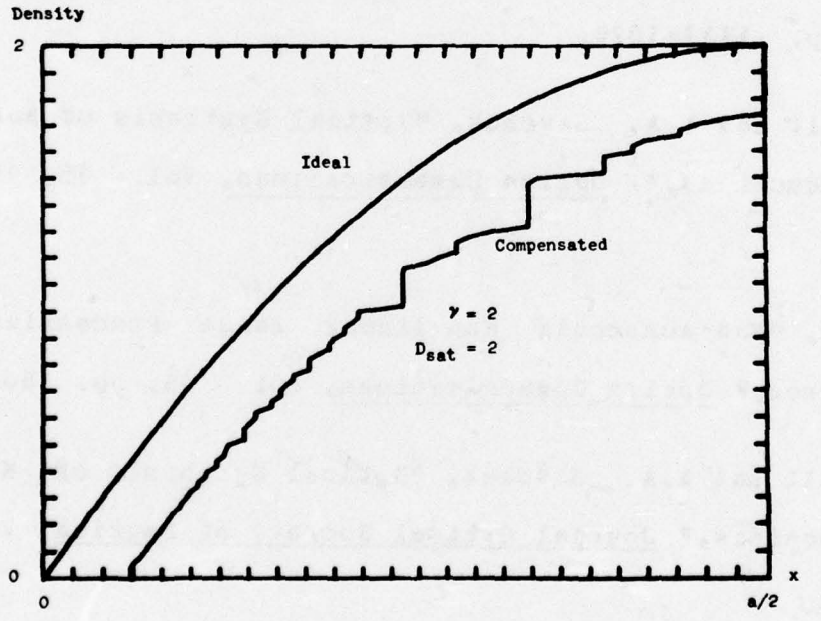


(b) Halftone cell

Figure 3.7-7. Low γ and D_{SAT} compensation.



(a) Transfer function



(b) Halftone cell

Figure 3.7-8. Moderate γ and D_{SAT} compensation.

References

1. H. Kato and J.W. Goodman, "Nonlinear Transformations and Logarithmic Filtering in Coherent Optical Systems," Optics Communications, Vol. 8, 1973, pp. 378-381.

2. A.A. Sawchuk and S.R. Dashiell, "Nonmonotonic Nonlinearities in Optical Processing," Proceedings IEEE International Optics Computing Conference, 1975, pp. 73-76.

3. H. Kato and J.W. Goodman, "Nonlinear Filtering in Coherent Optical Systems Through Halftone Screen Processes," Applied Optics, Vol. 14, 1975, p. 1813-1824.

4. S.R. Dashiell and A.A. Sawchuk, "Optical Synthesis of Non-linear Non-monotonic Functions," Optics Communications, Vol. 15, 1975, pp. 66-70.

5. T.C. Strand, "Non-monotonic Non-linear Image Processing Using Halftone Techniques," Optics Communications, Vol. 15, pp. 60-65.

6. S.R. Dashiell and A.A. Sawchuk, "Optical Synthesis of Nonlinear Nonmonotonic Functions," Journal Optical Society of America, Vol. 65, 1975, p. 1177.

7. S.R. Dashiell and A.A. Sawchuk, "Synthesis of Nonlinear Non-monotonic Functions in Optical Image Processing," Proceedings Optical Society of America Topical Meeting on Image Processing, Asilomar, California, February 24-26, 1976.

8. J.W. Goodman, Introduction to Fourier Optics, McGraw-Hill, New York, 1968.

3.8 Artificial Stereo from Image Features

Alexander A. Sawchuk and Harry C. Andrews

It is well known that stereo image information is an absolute necessity for certain image processing applications such as photointerpretation and cartography. This section describes some preliminary experiments in artificially generating stereo image pairs from single pictures. In these artificial stereo images, the depth information can be obtained from many different types of image features, including pixel brightness, edge information, texture, or a combination of these. One intent of this study is to evaluate the depth illusion created by artificial stereo as an aid to human perception. An important reference for the study of textures in stereo perception is the work of Julesz <1>.

A program has been written to accept any height information about a two-dimensional scene and create a left-right stereo pair from a single image. The program is quite general and the height information can be obtained from a priori knowledge of the scene height, or artificially created from image features as described. The height information is supplied as a non-negative number which proportionately controls the perceived height in the left-right stereo pair. For the left image of the pair, the program locally shifts pixels to the right depending on the height information. In the local shift operation

certain pixels are obliterated by the shift and other surrounding pixels must be moved in to fill in the missing area of the shifted pixels. The program accomplishes this automatically and efficiently keeps only the needed image line in core. A scale factor can be also chosen to control the gain on the perceived depth. For the right image, the operation is similar except that the local pixel shift is to the left. The scale factor is also used with unnormalized height data to control the maximum pixel shift. For input images and artificial stereo pairs of 256 x 256, a maximum local shift of 10 pixels is all that is needed for an exaggerated depth illusion.

It is important that the input information be locally smoothed. Unsmoothed height data produces jagged edges on objects and severe distortions of image shapes. Although some amount of distortion in each picture is tolerable because the eye tends to fuse images, any severe distortion detracts from the stereo illusion. It should also be mentioned that local features such as objects, spots, lines, etc. must be preserved simply for the eye to follow in perceiving depth. If a uniformly featureless flat field were locally shifted, there would be no perceived depth because there are no features for the eye to place in register.

An experiment to test these ideas have been performed. A Roberts gradient edge detector was used to create the height information. Denoting an image array by $f(i,j)$, the Roberts operator produces an edge map

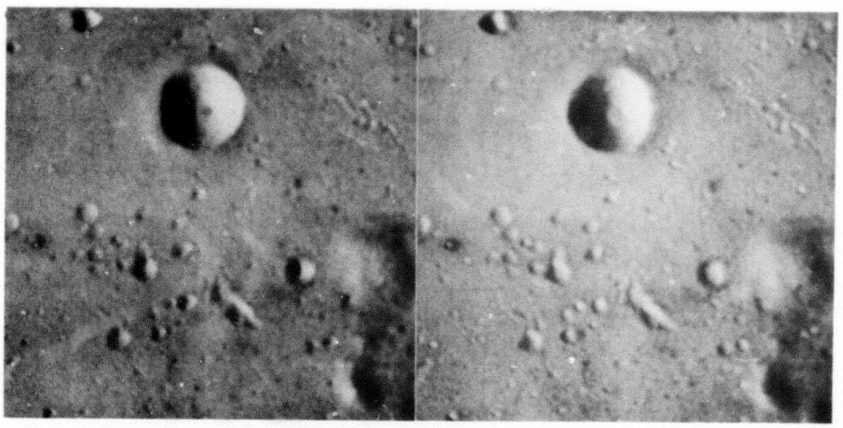
$$e(i,j) = |f(i,j) - f(i+1,j+1)| + |f(i+1,j) - f(j+1,i)| \qquad (3.8\text{-}1)$$

by operating in a 2 x 2 array of pixels. The Roberts operator is very affective at locally isolating edges, but using the edge output e(i,j) directly for height information produced poor results. The Roberts information was smoothed using a simple uniformly weighted convolutional averaging window. Various window sizes have been tried, but a window of 3 pixels horizontally and 5 pixels vertically seems to work well. The best averaging window probably depends on the amount of local detail in the input image.

The smoothed Roberts edge data was used as the input to the artificial stereo program for an experimental example using a moon picture. The left-right stereo pair is shown in figure 1. These pictures are best viewed with a binocular viewer or with a prism held close to one eye so that the image will fuse. The stereo impression from these pictures is very striking; particularly in the lower center of the image where the edge information from the craters makes them appear to be floating at different heights.

This experiment is only preliminary, and considerable further work is suggested by these results. Different preprocessing of height information is a possibility; Julesz <1> has shown that stereo perception with left-right pairs is remarkably unaffected by blurring, small size reductions or noise in one image.

Reference



(a) Left image (b) Right image

Figure 3.8-1. Artificial stereo images - edge information.

1. B. Julesz, "Texture and Visual Perception," Scientific American, February 1965.

4. Real Time Implementation of Image Processing Techniques

In the past several years a number of device concepts and technologies, such as charge coupled devices (CCD) and integrated injection logic (I L), have been developed which are already having significant impact on signal processing capabilities. The purpose of this task is to study means of performing proven image processing algorithms with such technologies in integrated circuit form. The benefits of performing the processing in silicon integrated circuits, which might be incorporated into an image sensor, for example, are clear in terms of power, weight and cost. The advantages of both CCDs and I L in terms of their power-delay product make them particularly attractive, as shown in figure 1. However, it is not intended that the study be limited to these two technologies. If, for example, the speed of GaAs FET logic can add significantly to the capabilities of the processor that technology will be included and the cost and benefits examined.

Since this is a new task some discussion of the program goals and schedules is warranted. The inherent flexibility of the device technologies under consideration implies that a very large number of algorithms might usefully be studied. However, at the present time it is our intention to emphasize the following five classical operations:

(a) Convolutional processing (both spatial and via chirp transformation techniques)

(b) Edge detection via the Roberts magnitude operator

(c) Edge detection via the Sobel operator

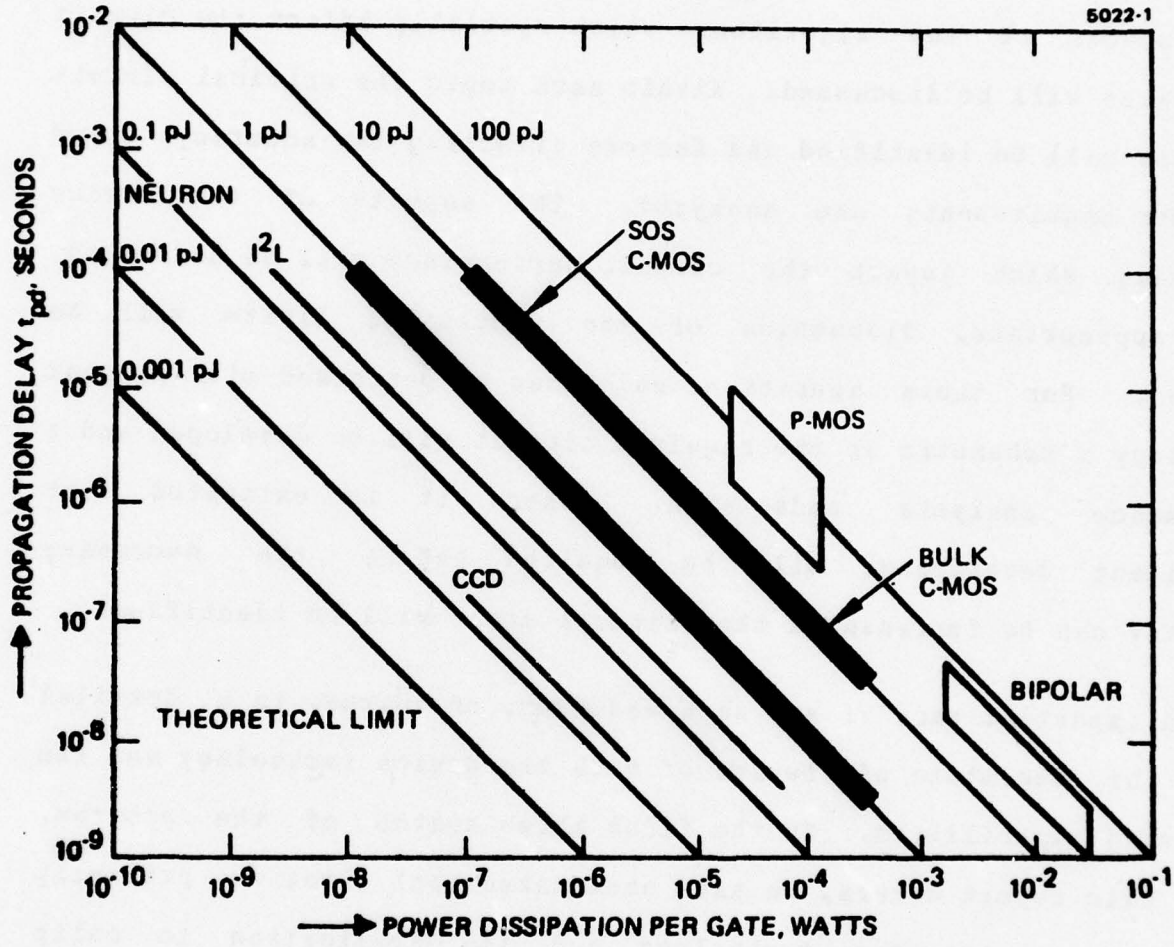


Figure 4.1-1. Power delay for different types of logic.

(d) Edge detection via the Hueckel operator

(e) Successive histogram operations.

Each of these operations is described in the literature and only such aspects of the algorithms which specially affect the circuit performance will be discussed. Within each topic the critical circuit functions will be identified and factors affecting the accuracy, speed and power requirements are analyzed. The aspects of the device processing which impact the circuit performance will be addressed. Where appropriate, discussion of the processing limits will be included. For those operations which can be performed with present technology a schematic of the required circuit will be developed and a performance analysis undertaken. Where it is estimated that significant development will be required before the necessary circuitry can be implemented the critical areas will be identified.

An important part of any such endeavor, of course, is a detailed review of the state of the art of both the device technology and the processing capabilities. In the first three months of the program, which this report covers, we have undertaken such a review, primarily of the charge transfer technology and its application to chirp transform processing. A summary of this work is included here.

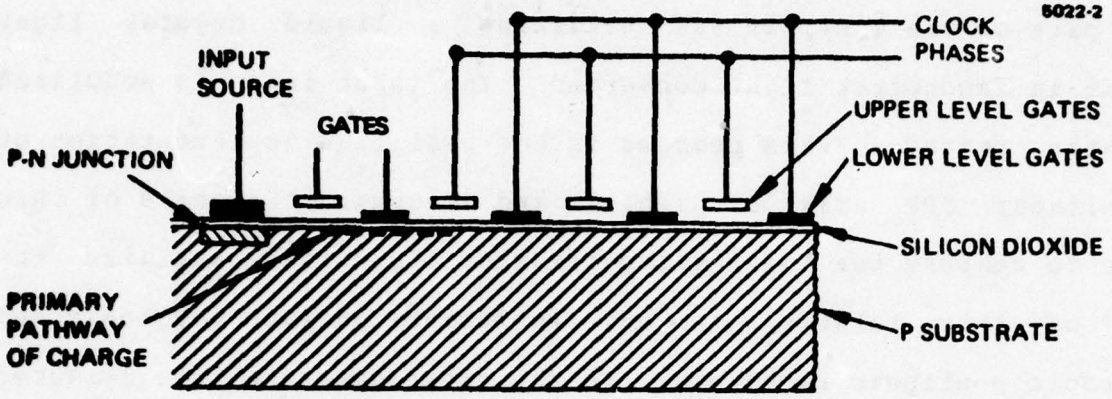
During the next quarter we intend to extend the analysis of CCDs to include the other four functions described above and provide block schematic diagrams and circuit layouts where possible, prior to analysis of possible I^2L implementations.

Another objective of this task is to experimentally study the real time implementation of nonlinear optical data processing (ODP). In this part of the work, we are utilizing a liquid crystal light valve as an incoherent light converter. The input image is modulated by a grating pattern. This process is the real time implementation of the nonlinear ODP work of Dashiell and Sawchuk. The goals of this work are to measure the light valve transfer function, maximize the "gamma" of this relationship and implement various monotonic and nonmonotonic nonlinear functions. In the first quarter, we measured the light valve transfer function with different drive conditions and demonstrated a nonmonotonic nonlinearity. This work is reported in Section 4.3

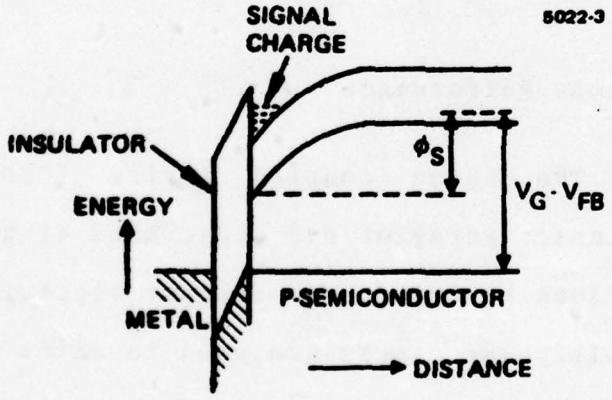
4.1 CCD Technology Review

4.1.1 Charge Transfer Devices Performance

CCD Device Operation: The charge coupled device (CCD) concept <1,2> consists of a linear array of MOS capacitors with input and output semiconductor junctions to inject and extract electrical charge as shown in figure 2. Multiphase clocks are used to shift the charge from one capacitor to the next. Three basic types of CCDs exist; surface channel (SCCD), buried channel (BCCD), and nonuniformly doped devices such as the paristaltic (PCCD). In the surface channel device the MOS capacitor is formed by a thin (typically $\simeq 0.1\mu$ m) silicon dioxide layer formed on a silicon substrate. The control gates, usually of aluminum or heavily doped polysilicon are eposited on the



(a) Surface channel CCD



(b) Potential profile of SCCD

Figure 4.1-2. Surface channel CCD and potential profile of SCCD.

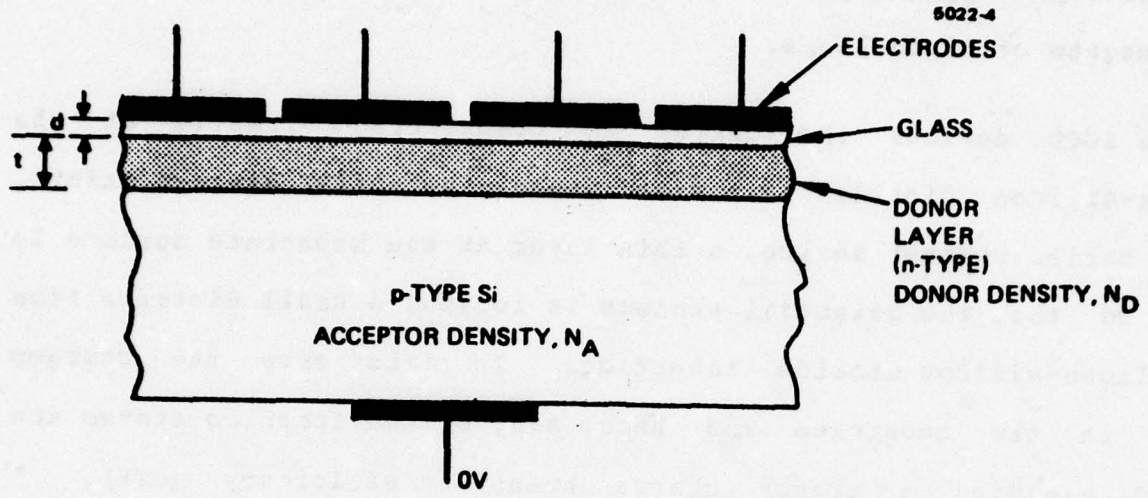
SiO_2 surface. Typical geometries, using photolithography, have gates with lengths of 10 to 25 μm.

In SCCD devices the charge is transferred directly at the silicon-silicon dioxide interface where a potential minimum exists. In the buried channel device, a thin layer at the substrate surface is doped so that the potential minimum is located a small distance from the silicon-silicon dioxide interface. In this case the charges travel in the substrate and hence many of the trapping states are avoided, yielding a higher charge transfer efficiency (CTE). A schematic of a buried channel device is shown in figure 3.

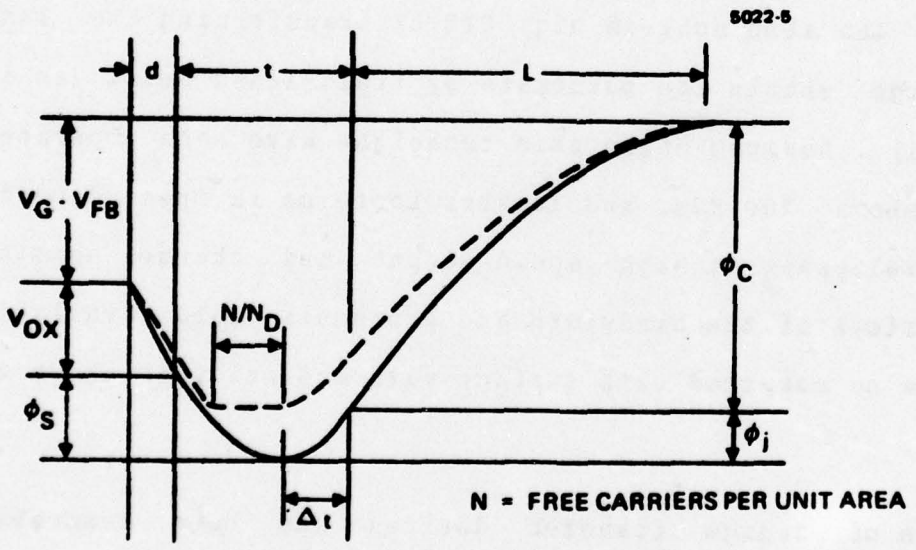
Peristaltic <3> devices rely on varying the doping density within the buried channel to achieve high charge storage (as in surface channel operation) and also achieve high CTE by transferring the last fraction of charge within the substrate by field-aided drift (as in the buried channel). Devices using this technique have been operated at frequencies above 100 MHz, and further increase in speed depends largely on the development of high speed input and charge sensing devices. A comparison of the bandwidth and processing delay available with these devices as compared with surface wave acoustic, is shown in figure 4.

The operation of charge transfer devices has been described elsewhere <4,5>, but for the sake of completeness the concept is reviewed briefly here.

Figure 2 shows a single cell of a typical p-type surface channel



(a) Buried channel CCD (BCCD)



(b) Potential profile in BCCD

Figure 4.1-3. Buried channel CCD.

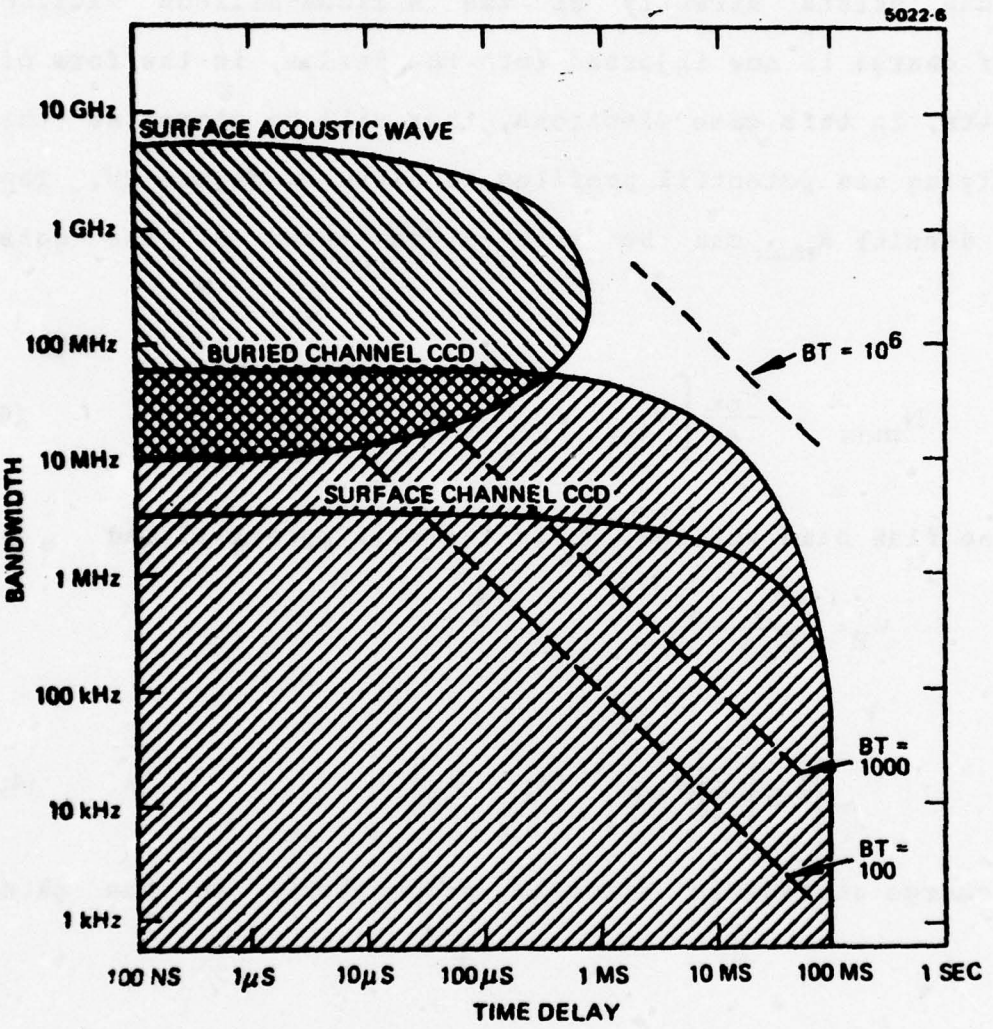


Figure 4.1-4. Present limits of time-bandwidth product for CCD and surface wave devices.

device. A positive electrode voltage applied to the device repels the majority carriers creating a depletion region directly beneath the gate. The energy bands in the semiconductor then become bent and a potential minimum exists directly at the silicon-silicon dioxide interface. If charge is now injected into the device, in the form of majority carriers, in this case electrons, they will be stored at the interface modifying the potential profiles as shown in figure 2b. The maximum charge density N_{max} can be written in terms of the gate voltage V_G as

$$N_{max} = \frac{C_{ox}}{e}\left[V_G - V_{FB} - 2\phi_s - V_B\right] \qquad (4.1\text{-}1)$$

where V_{FB} is the flat band voltage, ϕ_s is the well potential and

$$V_B = \frac{(4eN_a s\phi_s)^{1/2}}{C_{ox}}$$

Typically

$$\left|V_G\right| > \left|V_{FB} - 2\phi_s - V_B\right| \qquad (4.1\text{-}2)$$

and hence the charge storage is directly proportional to the gate voltage, i.e.

$$N_{max} = \frac{C_{ox}\, V_G}{e} A \qquad (4.1\text{-}3)$$

where A is the charge storage area.

The charge transfer is effected by clocking adjacent gates so as to create increasing well depth in the direction of desired charge

travel. Numerous clocking techniques have been employed ranging from a single cycle to full four-phase clocking. The stepped oxide technique is commonly employed to achieve two-phase operation where two adjacent gates are connected together and the variation in oxide thickness provides the different well depths.

For SCCD devices the charge travels at the silicon SiO_2 interface where the density of trapping states is high, typically of the order of $10^9 cm^{-2} eV^{-1}$. The CTE is therefore necessarily lower than for BCCD. However, CTE better than 0.999 have been obtained with SCCD devices by using a "fat zero" which involves constantly clocking through a small background charge. The ratio of the charge remaining $N_T(\tau)$ after time, to the original charge, N_o, in the absence of field aid drift is given by

$$\frac{N_\tau(t)}{N_o} \approx \left(1 + \frac{\pi \mu_n e N_o t}{2L^2 C}\right)^{-1} \qquad (4.1\text{-}4)$$

where μ_n is the electron mobility, L is the distance between adjacent gates, and C is the effective capacitance per unit area.

If fringing fields exist as in the BCCD the relationship becomes

$$\frac{N_\tau(t)}{N_o} \approx \exp\left(\frac{-3L^3 C}{\mu_n 2\Delta V \pi \epsilon_s} t\right) \qquad (4.1\text{-}5)$$

where ΔV is the voltage difference between the initial and final sites. An analysis of the relationship between these two mechanisms shows that the field-induced drift is dominant in the final stage of

the charge transfer and is essential for high CTE at high frequencies. Since most of the electrostatic field is normal to the direction of charge transfer, "complete" charge transfer can be accomplished only at relatively low clock rates (of the order of 100 kHz). In the buried channel concept, a layer of n-doped material is introduced on the surface of the p-type silicon by epitaxial or ion implantation. When a positive voltage is applied, a depletion region is formed under the electrode and the resulting potential profile is as shown in figure 3b. The potential minimum ϕ_{min} is determined by the gate voltage V_G as

$$\phi_{min} = \left\{ - \left(\frac{e\epsilon_s}{2} \frac{N_A (N_D + N_A)}{N_D} \right)^{\frac{1}{2}} \left(\frac{d}{\epsilon_{ox}} + \frac{t}{\epsilon_s} - \frac{\frac{N}{N_D}}{\epsilon_s} \right) \right.$$

$$+ \left[\frac{e\epsilon_s}{2} \frac{N_A (N_D + N_A)}{N_D} \left(\frac{d}{\epsilon_{ox}} + \frac{t}{\epsilon_s} - \frac{\frac{N}{N_D}}{\epsilon_s} \right)^2 \right.$$

$$\left. \left. + \frac{N_D + N_A}{N_D} \left(V_G - V_{FB} + V_I - eN \cdot \left(\frac{d}{\epsilon_{ox}} + \frac{t}{\epsilon_s} - \frac{\frac{N}{N_D}}{2\epsilon_s} \right) \right) \right]^{\frac{1}{2}} \right\}^2 \quad (4.1\text{-}6)$$

where, using the notation shown in figure 3b

$$V_I = e\, N_D t\left(\frac{d}{\epsilon_{ox}} + \frac{t}{2\epsilon_s}\right)$$

This function is plotted in figure 5, (which is taken from <5>) showing that the potential is approximately a linear function of V_G over a wide range of gate voltages. When charge is injected into the device it resides at the potential minimum which is typically several hundred angstroms from the interface. This allows more effective charge transfer and CTE as high as 0.9999. Further the nonlinear or transverse fields are considerably larger at these depths and account for a considerable part of the charge transfer. This in turn allows the higher speeds to be achieved, as illustrated in figure 2.

The peristaltic concept combines the high charge storage capabilities of the SCCD with high charge transfer efficiency and large bandwidth capabilities of the BCCD. It achieves this basically by storing the charge at the surface of the silicon while transferring the last charges deep in the silicon substrate where the fringing fields are strongest. Of the three transfer mechanisms, self-induced drift, thermal diffusion and field aided drift the former accounts for 90 to 99% of the transfer and is accomplished in a few nanoseconds. The thermal diffusion is usually much slower and is significant only in the final stages of the SCCD transfer process. Peristaltic operation stores the charge close to the silicon surface so as to achieve high storage as in the SCCD but the doping profile is modified so as to force the last fraction of charge into the bulk during the transfer process where field aided drift accounts for the final

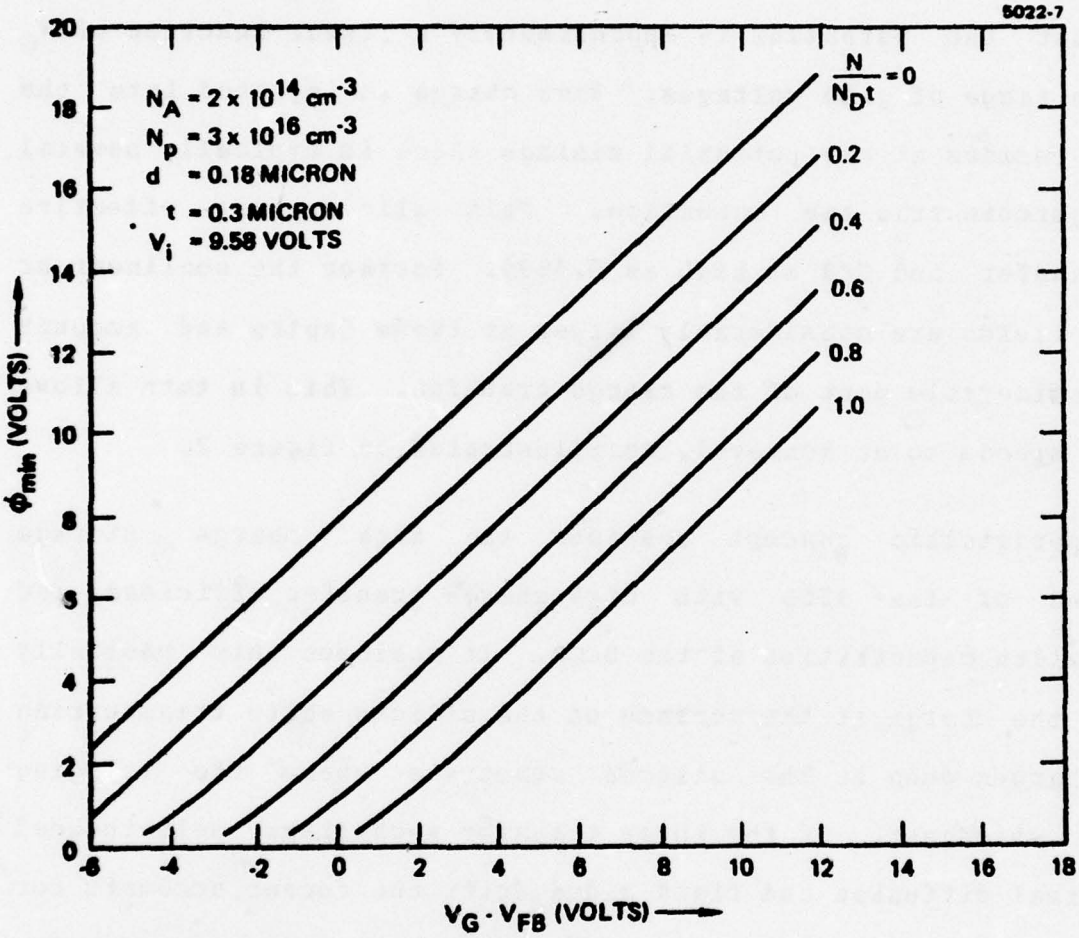


Figure 4.1-5. Potential minimum as a function of gate voltage and charge density N (from Ref.< 5>)

transfer.

Input Circuitry: The speed and efficiency of the CCD input and output circuitry is, of course, an important consideration in the ultimate performance of the final device. Two basic concepts exist to convert the input voltage signal into the required charge for the CCD. They involve either integrating a current for a fixed time interval or charging a fixed capacitor. The integrating input requires less peripheral circuitry, provides voltage gain, and generally has lower noise. The capacitor input (Tompsett input or threshold correcting input) is more linear and is less sensitive to bias conditions. Both inputs have approximately the same transfer characteristics.

A schematic of the integrating input is shown in figure 6. The input signal is ac coupled to the V_{IN} gate, thus modulating the amount of current trickling into the CCD. For this particular input, charge continues to trickle regardless of the state of ϕ_1. The amount of charge finally collected in the ϕ_1 transfer bucket is, therefore, proportional to the integral of the input voltage over the full clock period. The input is very sensitive to the bias voltages at the source diffusion and the input electrode, and becomes increasingly nonlinear for large ac values of V_{in} due to the input MOSFET's square law dependence of current on gate-to-source voltage. The sensitivity problem of this input is further heightened by the traditionally unpredictable threshold voltage.

The threshold correcting input, shown in figure 7 solves many of

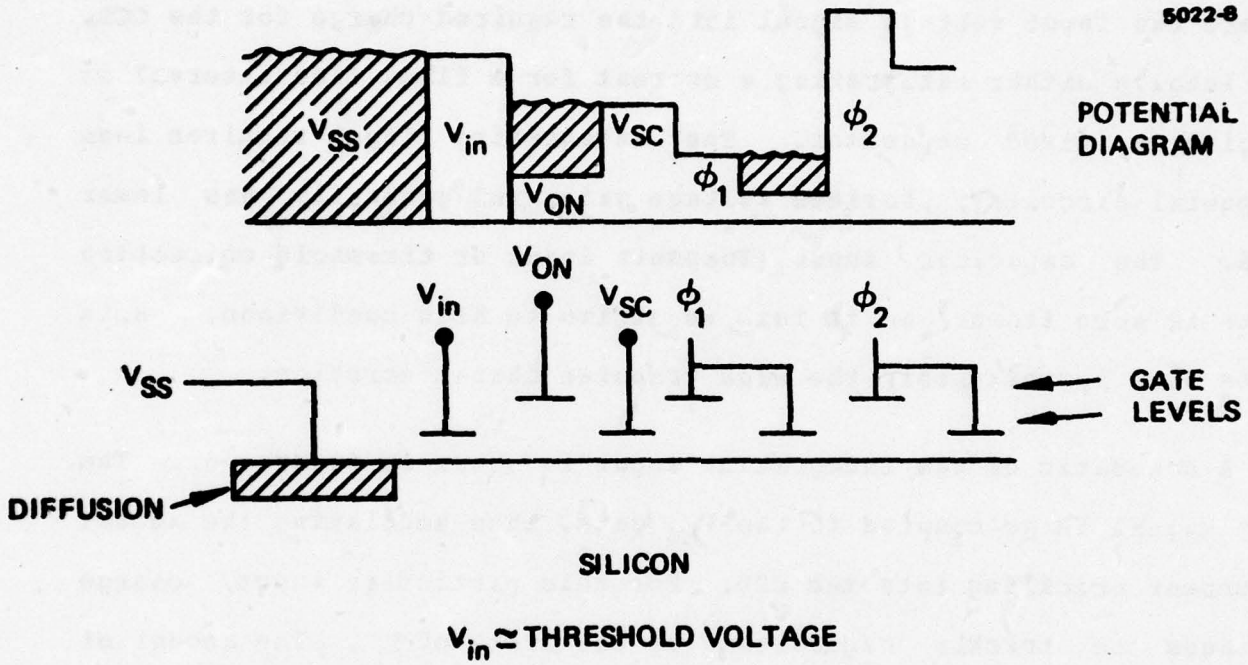


Figure 4.1-6. Schematic of integrating input.

the problems at a cost of gain and some noise. The input diffusion is driven by ϕ_{in} to a level where the V_{IN} gate is overfloved. The diffusion then recedes and the extra charge falls back with it leaving an amount of charge proportional to the difference between V_{in} and V_{dc}. When ϕ_1 fails to receive the input charge, it picks up an amount of charge proportional to the minimum of the input signal during the input aperture time.

The threshold correcting input makes use of the fact that although the threshold voltage may vary across the chip and from wafer to wafer, the variation between two adjacent (or almost adjacent) gates can be made very small thus ensuring that when the V_{in} gate and the V_{dc} gate are at the same potential, no charge will be entering the CCD. This differential aspect of the input scheme, combined with the large input full scale signal, make the input biasing very noncritical. This voltage type of input is also independent of clock rate unlike the previous integrating input. The speed of input devices of this type is presently limited to about 100 MHz but high resolution bipolar circuitry presently being developed may increase this to several hundred MHz.

Output Circuitry: At the output of the CCD the charge must be converted into a voltage. This can be achieved by discharging the signal through a precision controlled resistor and measuring the resulting voltage. Alternatively the charge can be dumped into a capacitor and the voltage measured.

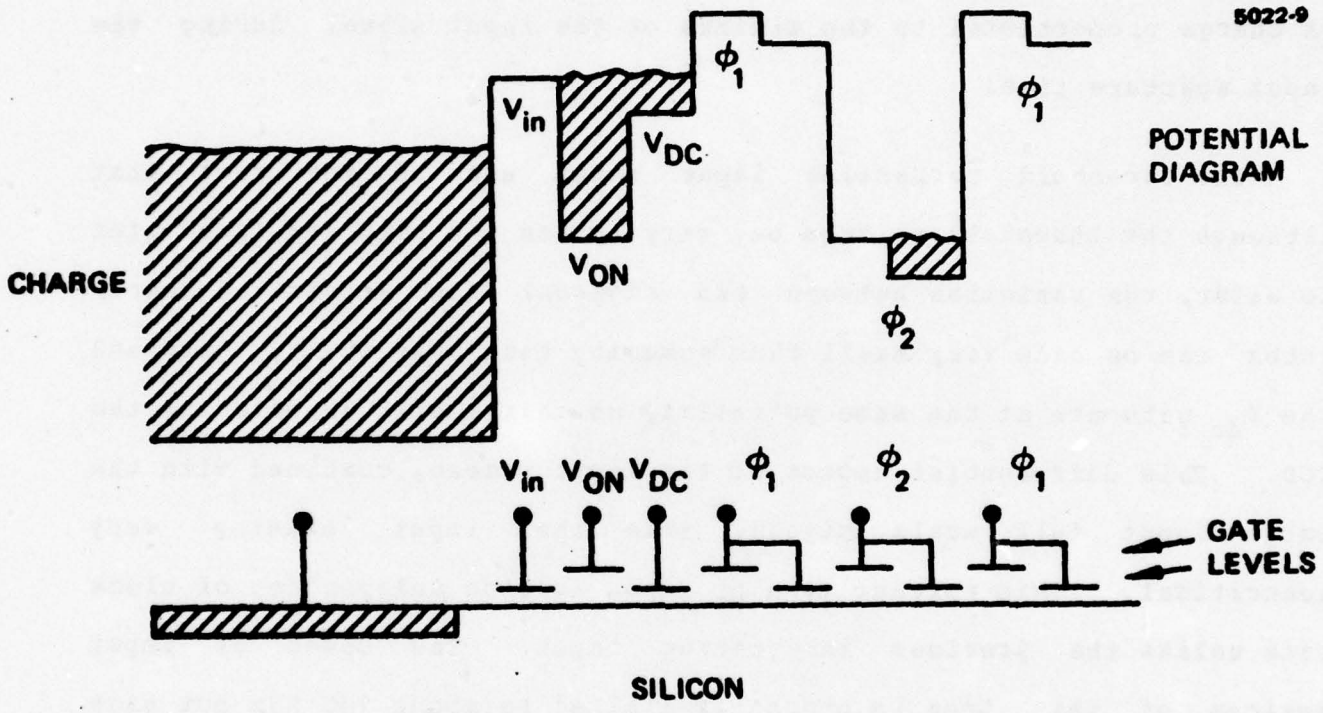


Figure 4.1-7. Threshold correcting input.

The former technique is illustrated in figure 8. It is basically simpler to implement and operates at high speeds. In general, when R is large enough to match the resistor noise current to the CCD noise current, the signal bandwidth is below one-half the clock frequency. To increase the bandwidth, a transimpedance amplifier is used instead of the resistor. The transimpedance amplifier is generally an op amp with a feedback resistor R. The output diffusion is tied to the summing junction of the op amp, and the output voltage bias is connected to the positive input of the amplifier. Such a circuit is shown in figure 9. The advantage of this low input impedance transimpedance amplifier is that the effect of the parasitic capacitance on the output node is divided by the open loop gain of the amplifier. Discrete circuits capable of handling 10 pF on the output node have been built with a transimpedance of 50 K ohms at signal frequencies up to 10 MHz. The bandwidth can be extended further if lower transimpedances are acceptable.

The second output circuit (normally called a reset output) is more common than the direct current mode type. The two kinds of reset output circuits are called floating diffusion and floating gate. A schematic of the floating diffusion output is shown in figure 10. The method of operation for this output is (1) while ØRST is high, Ø 2 rises to dump charge over the screen gate; (2) this charge increases the surface potential around the floating diffusion thus increasing the diffusion potential; and (3) the source follower rises as the floating diffusion rises. The output is reset when ØRST lowers and allows the signal charge to flow into the V_{DD} supply. The output

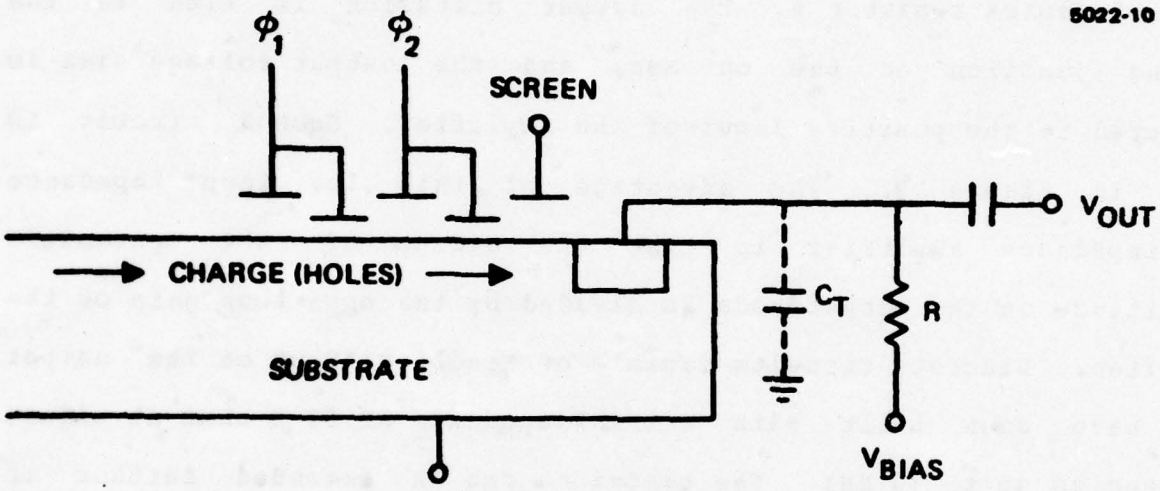


Figure 4.1-8. Output circuit.

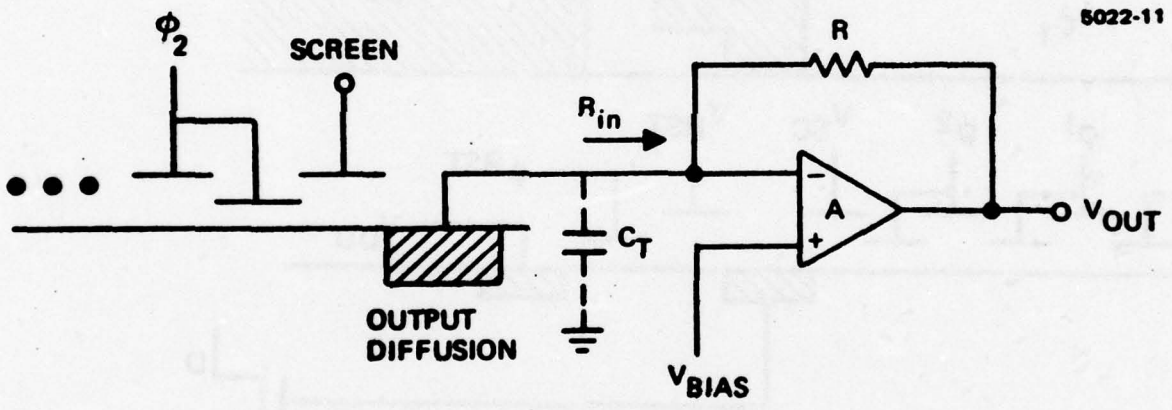


Figure 4.1-9. DC output using transimpedance amplifier.

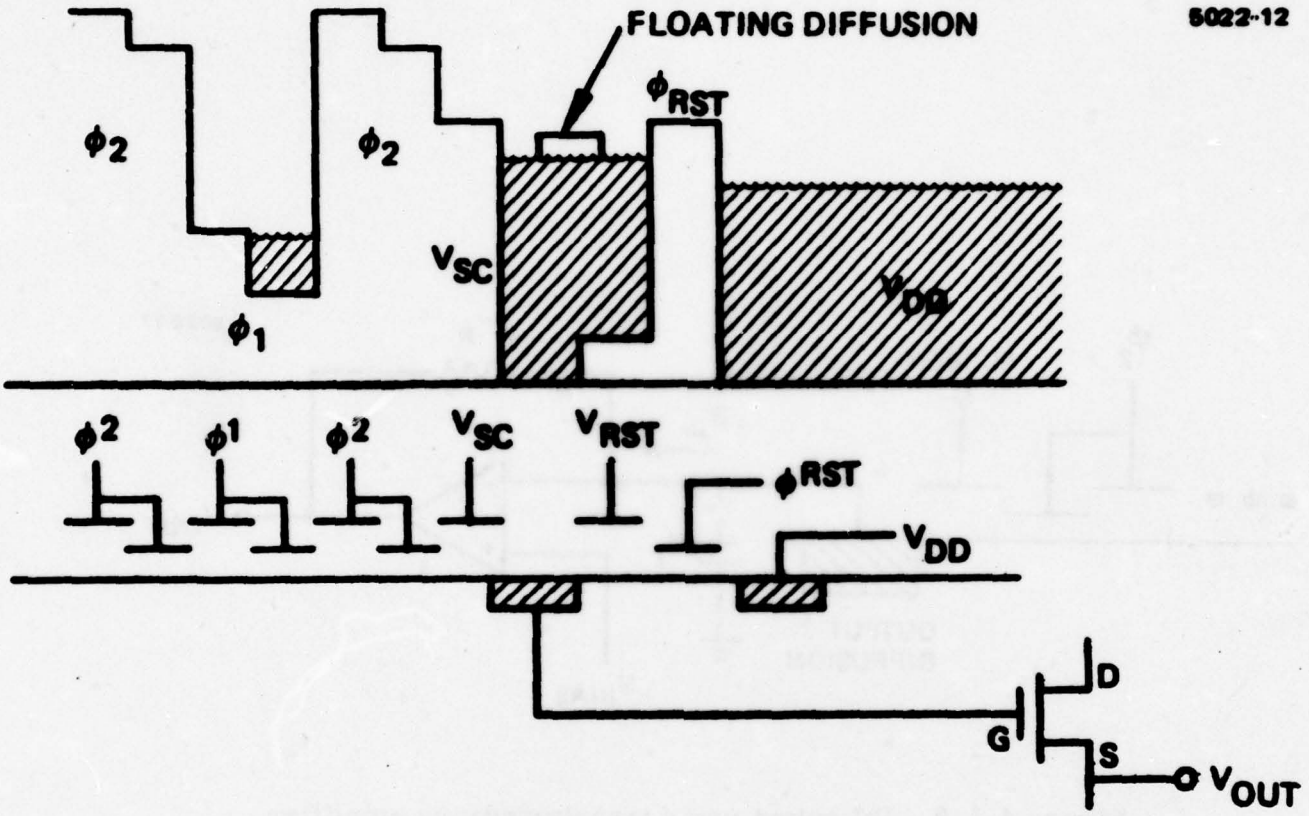


Figure 4.1-10. Floating diffusion output.

voltage from this circuit is constant with clock frequency if the duty cycle of the output is held constant. The magnitude of the output is determined by the amount of full scale charge and the capacitance on the output node. With a 50% duty cycle, the output magnitude can be as high as 0.5 V rms.

Using standard photolithography output circuits of this type have been operated routinely at 20 MHz. Presently these bipolar devices are being developed using electron beam exposure to operate at speeds in excess of 100 MHz.

Charge Transfer Efficiency (CTE): The charge transfer process, even for a single transfer, is never complete. A certain fraction of charge, ϵ, is lost. The principal causes of the incomplete transfer are the finite velocity of the carriers and the trapping states. The CTE is usually lower for SCCD devices than for BCCD because (a) the density of trapping states are greatest at the Si-SiO_2 interface and (b) the transfer is usually incomplete (at all but the very slowest clock rates) since field aided drift is absent. In addition the charges trapped in the interface states empty slowly and hence contribute to trailing data. In fat zero operation a constant charge level, of 10 to 20% of the full signal level, is used to keep all the states full and then reduce the charge loss. The technique is primarily used for digital applications and can result in $\epsilon \simeq 10^{-4}$.

The density of traps deep in the silicon is much less and hence the need for fat zero in BCCD is reduced. A principle effect of the

finite charge inefficiency, apart from resulting attenuation and limiting the dynamic range, is to limit the total number of transfers possible. Since the charge attenuation is cumulative, the overall characteristic for n transfers is given by

$$\frac{Q_{out}}{Q_{in}} = (1 - \epsilon)^n \approx 1 - n\epsilon \qquad (4.1\text{-}7)$$

To avoid excessive degradation in signal to noise ratio usually requires less than $<0.1/\epsilon$.

The effect of charge spilling into trailing data slots is to create dispersion. This can be seen from the Z-transform characteristic of a single stage, which can be written as

$$T(Z) = \left(\frac{1 - \epsilon}{1 - \epsilon Z^{-1}}\right) Z^{-1} \qquad (4.1\text{-}8)$$

where

$$Z = \exp\left[\frac{j2\pi f}{f_c}\right] \qquad (4.1\text{-}9)$$

and f_c is the applied clock rate.

For a transversal filter with ideal impulse response

$$T(Z) = \sum_{k=1}^{n} h_k Z^{-k} \qquad (4.1\text{-}10)$$

where h_k are the weighting coefficients, the actual transform can be written as

$$T(Z) = \sum_{k=1}^{n} h_k \left(\frac{1 - \epsilon}{1 - \epsilon Z^{-1}} \right)^k Z^{-k} \qquad (4.1\text{-}11)$$

It has been shown that this is equivalent to the transfer function of an ideal device evaluated at frequency

$$f^1 = f + \frac{\epsilon f_c}{2\pi} \left[\sin\left(\frac{2\pi f}{f_c}\right) - j\left\{1 - \cos\left(\frac{2\pi f}{f_c}\right)\right\} \right] \qquad (4.1\text{-}12)$$

which represents a frequency shift, or dispersion, given by

$$\Delta f = \frac{\epsilon f_c}{2\pi} \sin\left(\frac{2\pi f}{f_c}\right) \qquad (4.1\text{-}13)$$

and an envelope modulation given by

$$A(f) = 1 - \frac{n}{2}\left[1 - \cos\left(\frac{2\pi f}{f_c}\right)\right] \qquad (4.1\text{-}14)$$

The effect of this on the modulation transfer function is shown in figure 11. For large fractional bandwidth the product, $n\epsilon$, must be maintained as small as possible. Thus for a large time-bandwidth product, the signal, ϵ, must be reduced as much as possible. Further it can be shown that the charge transfer inefficiency does not alter the signal-to-noise ratio to first-order.

Accuracy of Weighting Technique: Typically CCD devices are weighted by the split electrode technique illustrated in figure 12. As shown, both ends of each storage electrode are connected to a differential current amplifier which senses the charge directly beneath it. The location of a small gap determines the tap weighting.

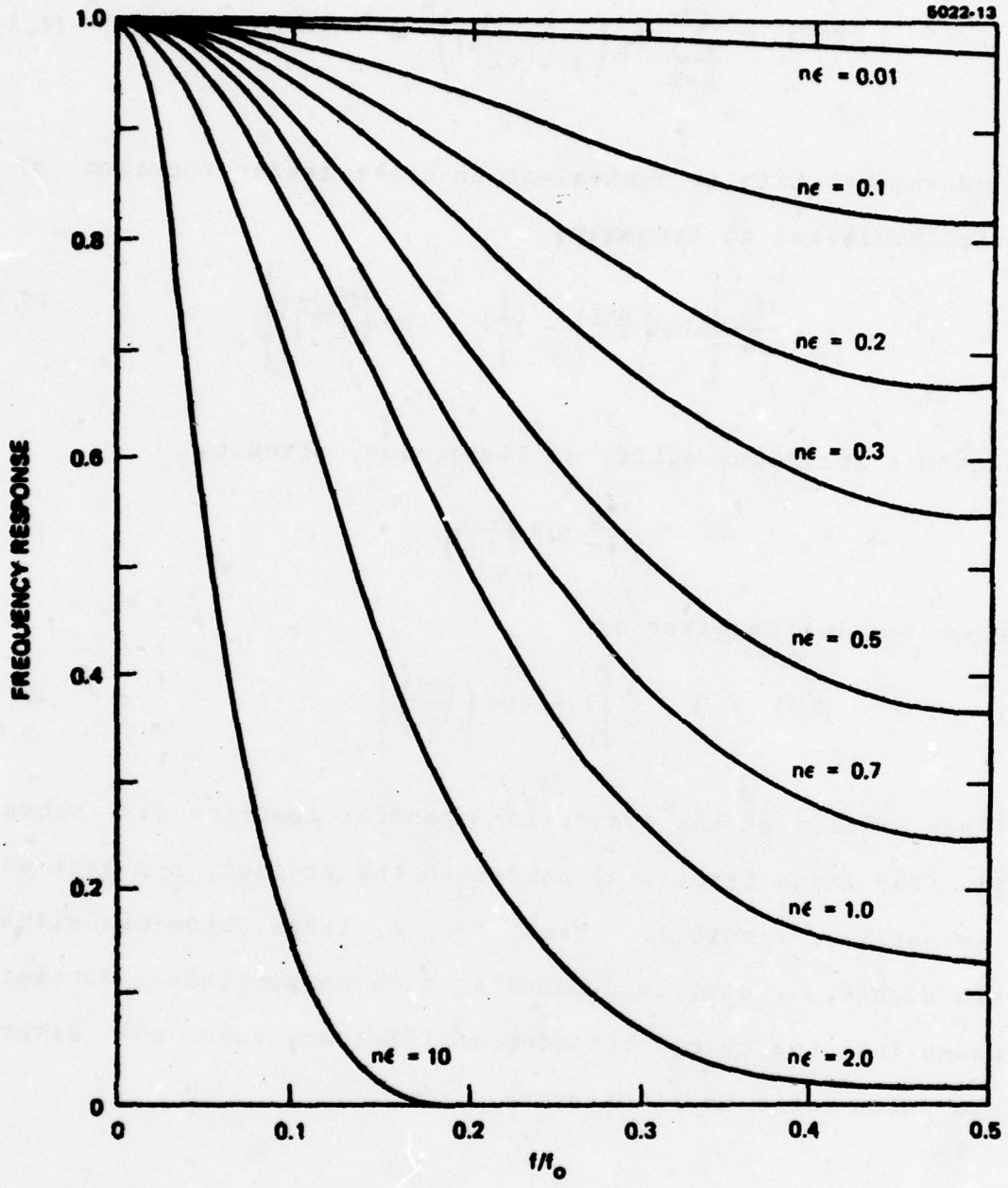


Figure 4.1-11. Modulation transfer function as a junction of charge transfer inefficiency, ϵ .

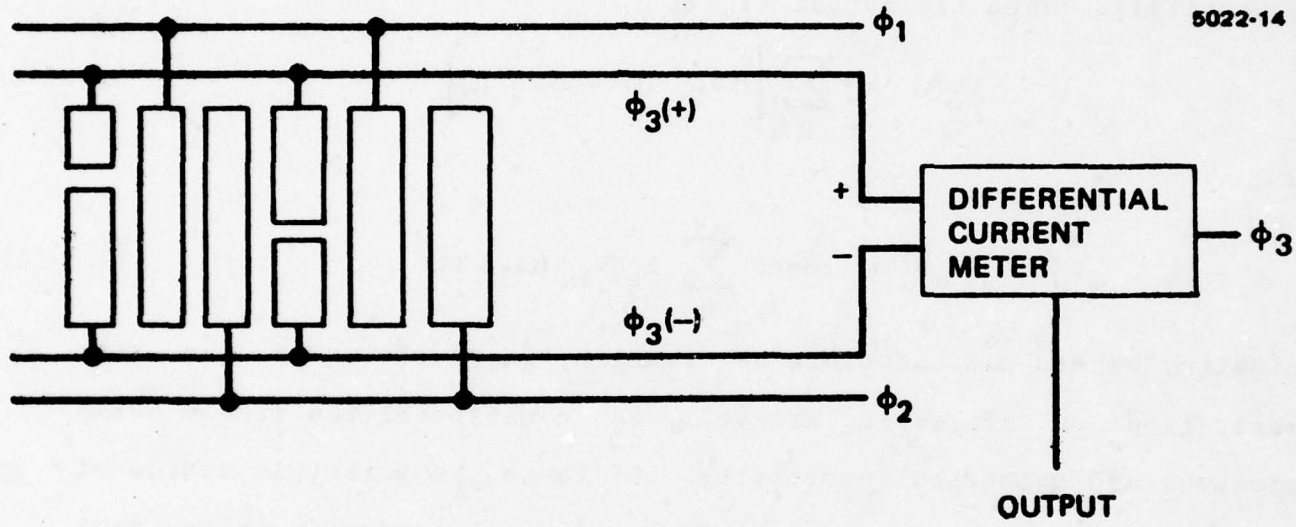


Figure 4.1-12. Concept of split electrode weighting.

For example, if the gates have length L and the distance between the gap and the center of the gate is h_k the fraction of charge coupled to the positive and negative clock line is given by

$$Q_k^+ = \frac{L}{2}(1 + h_k)$$

and

$$Q_k^- = \frac{L}{2}(1 - h_k) \qquad (4.1\text{-}15)$$

respectively. Then the output signal is

$$V_o(t) = \sum_k \left[\Delta V_k^+(t) - \Delta V_k^-(t)\right]$$

or

$$V_o(t) = \text{const} \sum_k h_k V_{in}(t - kT) \qquad (4.1\text{-}16)$$

Weighting errors can therefore be caused by inaccuracies in the mask fabrication or alignment as well as nonuniformities in the oxide thickness and substrate sensitivity. Of these the principle source of coherent error is the quantization error introduced in the mask generator. The actual resolution attainable after exposure and etching is usually considerably less than that of the mask. For example routine photolithographic techniques are limited in resolution to about 5 μm. For a gate length of 200 μm this represents an error of about ±1%. The situation is much improved if electron beam microfabrication techniques are employed. For example the facilities developed at Hughes Research Laboratories use a 15 bit D/A converter to provide resolution of 1 part in 32,000, and can provide final line accuracies of about 0.2 μm in an overall scan of 1 mm. Hence, for

typical gate lengths of 200 μm an error of approximately 0.05% is attainable.

The effect of weighting errors, σ_k, on the overall MTF of a transversal filter can be written in the terms of eq. (10) as

$$\tilde{T}(f) = \sum_{k=1}^{n} (h_k + \sigma_k) \exp\{\frac{-j2\pi kf}{f_c}\} \qquad (4.1\text{-}17)$$

where $\tilde{T}(f)$ is the distorted transfer function. The error term

$$\Delta T(f) = \tilde{T}(f) - T(f) \qquad (4.1\text{-}18)$$

then can be expressed as

$$\Delta T(f) = \sum_{k=1}^{n} \sigma_k \exp\{\frac{-j2\pi kf}{f_c}\} \quad . \qquad (4.1\text{-}19)$$

In general the exact evaluation of $\Delta T(f)$ depends on the desired filter response $T(f)$ and the characteristic of the weighting errors. However, if the weighting errors are assumed to be random with variance σ_{rms} then the fractional error can be written as

$$\frac{\Delta T(f)}{T(f)} = \frac{2\sigma_{rms}}{\sqrt{n}\, W} \qquad (4.1\text{-}20)$$

where W is the channel width.

For a device fabricated using photolithography and with 100 stages, $W = 200\mu m$ and $\sigma_{rms} = 2\mu m$

$$\frac{\Delta T(f)}{T(f)} = 0.2\% \qquad (4.1\text{-}21)$$

With electron beam fabrication techniques this error can be further reduced by about an order of magnitude.

Factors Affecting CCD Noise: Noise is an important consideration in all CCD applications and novel circuit techniques have been developed to minimize its effects. Three basic noise sources can be identified as (1) thermal noise in the input and output circuitry, (2) noise due to trapping states, and (3) dark current noise. A number of low-noise circuits have been developed to minimize the noise in the input and output circuitry. Typical of these are the correlated double sampling output circuit and the distributed gate amplifier. For specific details of these circuits the reader is referred to references <4> and <5>.

Since the CTE is limited, a fraction of the charge is left behind in each transfer and some additional charge is gathered from the preceding data. In general the signal fluctuation depends on the data format but an effective noise signal can be evaluated assuming independent fluctuations as

$$\Delta Q = 2n \overline{\Delta Q_{tr}^2}$$

where ΔQ_{tr}^2 is the mean square fluctuation in each transfer. The noise spectral density $S_{tr}(f)$ has been calculated as

$$S_{tr}(f) = \frac{4nf_c}{q} \overline{\Delta Q_{tr}^2} \left[1 - \cos\left(\frac{2\pi f}{f_c}\right)\right]$$

for low transfer inefficiency, i.e., $\epsilon << 1/n$ showing the noise at high frequencies to be enhanced.

A further source of noise is associated with the statistical emission of carriers. Thornber <6> has shown that this varies between

$$\overline{\Delta Q_r^2} = \frac{2}{3} KTC_s$$

and

$$\overline{\Delta Q_e^2} = \frac{1}{2} KTCs$$

for different parts of the clock cycle.

The trapping states within the SCCD contribute as another source of noise, since charge is absorbed and emitted independently of the clock cycle. The charge fluctuation from these sources can be evaluated as

$$\overline{\Delta Q_{ss}^2} = q^2 kT A_s N_{ss} \ln (2)$$

and hence an effective signal-to-noise ratio for a given charge packet can be calculated as

$$\frac{S}{N} = \frac{A_s C_{ox}^2 V_s^2}{2n kT N_{ss} \ln (2)}$$

where N_{ss} is the interface trap density (typically $\sim 10^9$ $cm^{-2}eV^{-1}$). Similar noise contributions are made in BCCD due to the bulk traps but in this case the noise is frequently dependant, reaching a maximum when the transfer time is approximately equal to the emission time constant. Effective models for the noise from bulk traps have not been developed. For most BCCD devices the equivalent noise phenomenon is perhaps an order of magnitude less in BCCD than SCCD, and is usually dominated by input and output noise sources. A full analysis

of the inherent noise source in CCDs, is made in references <6> and <7>.

4.1.2 Basic Signal Processing Functions With CCDs

The capabilities of CCD described in this section make them of great interest for processing large time bandwidth signals at frequencies up to ~100 MHz. The full range of basic operations such as addition, multiplication, etc. have been suggested using CCDs for analog processing, and in addition all digital operations can be performed. In this section the basic function blocks required to accomplished both analog and digital processing are discussed.

Addition: The basic serial operation of a CCD shift register makes it a natural technology for analog sampled data operations. Linearity of the order of 1% can be achieved over time bandwidth products of about 10^3; the data being stored in the form of finite amounts of charge. Addition can most easily be accomplished by simply merging the charge from two or more potential wells into a cell bucket. The technique is illustrated in figure 13 where the contents of two parallel analog delay lines are dumped into a single line. In this case the voltage on a screen separating the two channels is pulsed to allow charge flow. The technique is not limited to any number of separate lines and can be used to form a multiplexer. Further, by altering the well size (by varying the area of the receiving gate) the resultant output can be scaled by a constant factor.

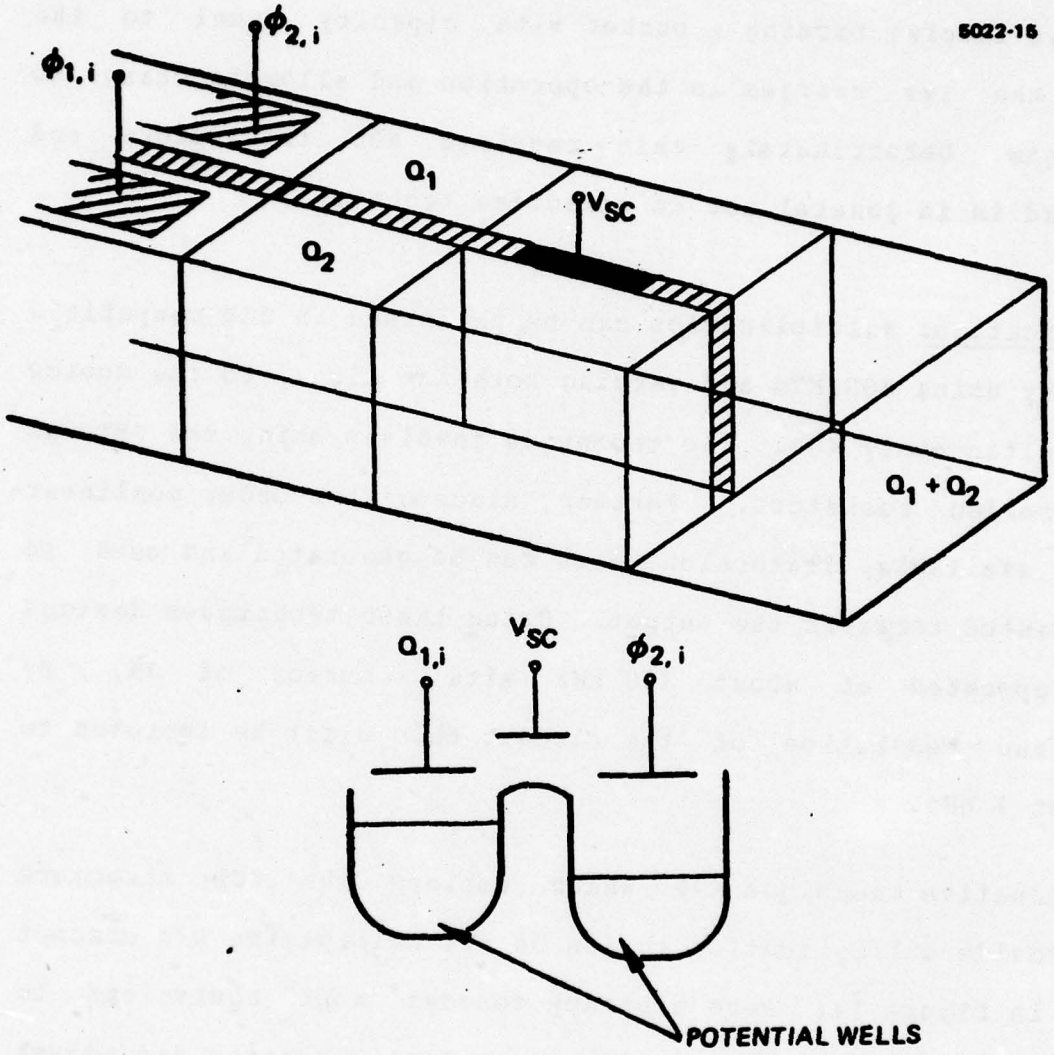


Figure 4.1-13. Schematic of technique for addition.

Generalized subtraction is a somewhat more complex operation and is not naturally suited to CCD technology. One technique, for example, might involve forming a bucket with capacity equal to the smaller of the two charges in the operation and allowing charge to drain into this. Unfortunately this requires MOS transistors and capacitors and is in general not an effective technique.

Multiplication: Multiplication can be performed in CCD compatible technology by using MOSFETs and varying both the signal to the source and gate simultaneously <8>. The technique involves using the FETs as voltage-controlled resistors. Further, since higher-order nonlinear products are available, distortion terms can be generated and used to null out unwanted terms in the output. Using these techniques devices have been operated at about 100 kHz with accuracy of 1%. By increasing the resolution of the circuit this might be improved to about 0.5% at 1 MHz.

An alternative technique <9> which employs the CCD structure itself to provide multiplication relies on the multiplying D/A concept illustrated in figure 14. Here a binary encoded word equivalent to one input is used to switch in and out parallel CCD gates and adjust the devices capacitance. In this way the capacitance is made proportional to one input and the driving voltage applied across the array proportional to the second. The output charge is hence given by

$$Q = \sum_{i=1}^{n} \frac{C}{2^i} V_{in}$$

and is proportional to the product of the two signals. At present

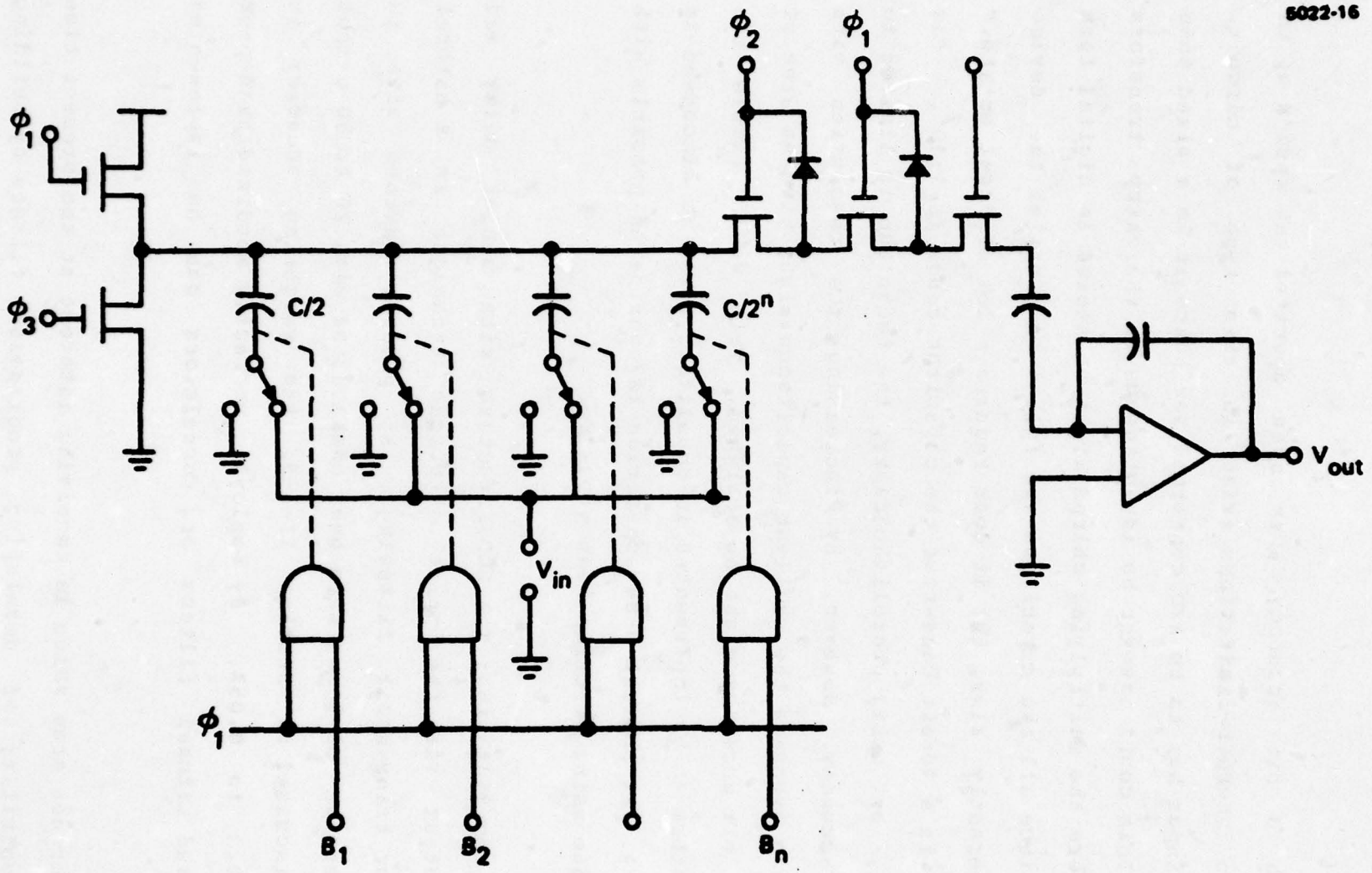


Figure 4.1-14. Concept of CCD multiplying D to A.

devices with 8 bit accuracy have been operated at speeds up to 100 kHz. Two inherent limitations exist in this type of circuit. First, one input has to be A/D converted and hence it is a mixed mode operation. This could however be advantageous in the chirp transform technique where the multiplying chirps might be stored in digital form in a ROM. Since all the capacitors are filled in parallel the device is not inherently slow, but it does require a lot of "real estate." For example, in a 10-bit converter the capacitor range is $1:10^3$. For this reason, by using photolithography, the technique is limited to about 8 bit accuracy. However, by increasing the resolution with electron beam exposure the smallest capacitance might have an area of $1\mu m^2$ and 11 bit accuracy might be achieved. This would enable the chirp transform to be implemented in monolithic form, as discussed in Section III-A. At present 5 MHz operation is considered possible with this technique using electron beam exposure.

Transverse Filtering: The CCD structure with serial delay and parallel output via the split electrode technique is a natural component for transversal filtering. At present devices with an impulse response of 1 sec have been constructed with 20 to 50 output taps. As discussed in Section II-A-5, the weighting accuracy is currently 0.1 to 0.05%. By employing correctly apodized band-pass filters, fixed matched filters or correlators can be implemented directly.

An important area which is receiving interest at the present time is the possibility of developing programmable filters by building

variable nondestructive tapping circuits. The concept of the floating clock electrode sensor (FCES) is presently being developed. By driving a source follower the charge can be read nondestructively and provide a low output impedance source. The circuitry discussed in reference <10> for this is shown in figure 15. At present the capacitance load on the sensing MOSFET limits the dynamic range, but with careful design and using a short gate bipolar output this problem might be overcome. Further, by using voltage-controlled FET conductance the technique can be fully electronically programmable. At present the accuracy is limited to about 3% at 100 kHz.

An electronically programmable correlator (or filter) can be implemented by using two counter-running analog CCD delay lines and feeding the parallel outputs into a series of MOSFET multipliers (as discussed in Section III-B-2). This technique requires synchronization of the signal carrying the correlation (or convolution) function with the input data and is limited to at most 50% duty factor for all but sliding convolution. However, this may have application where either the data is intermittent or the phase information is not required.

A/D Conversion: Analog to digital conversion can be implemented directly with CCDs by forming potential wells equivalent in capacity to each quantized bit, and moving charge through the array. The circuit required is similar to that shown in figure 14. A comparitor circuit must be included at each site. Again the technique is limited to about 8 to 10 bits and is quite bulky in this implementation.

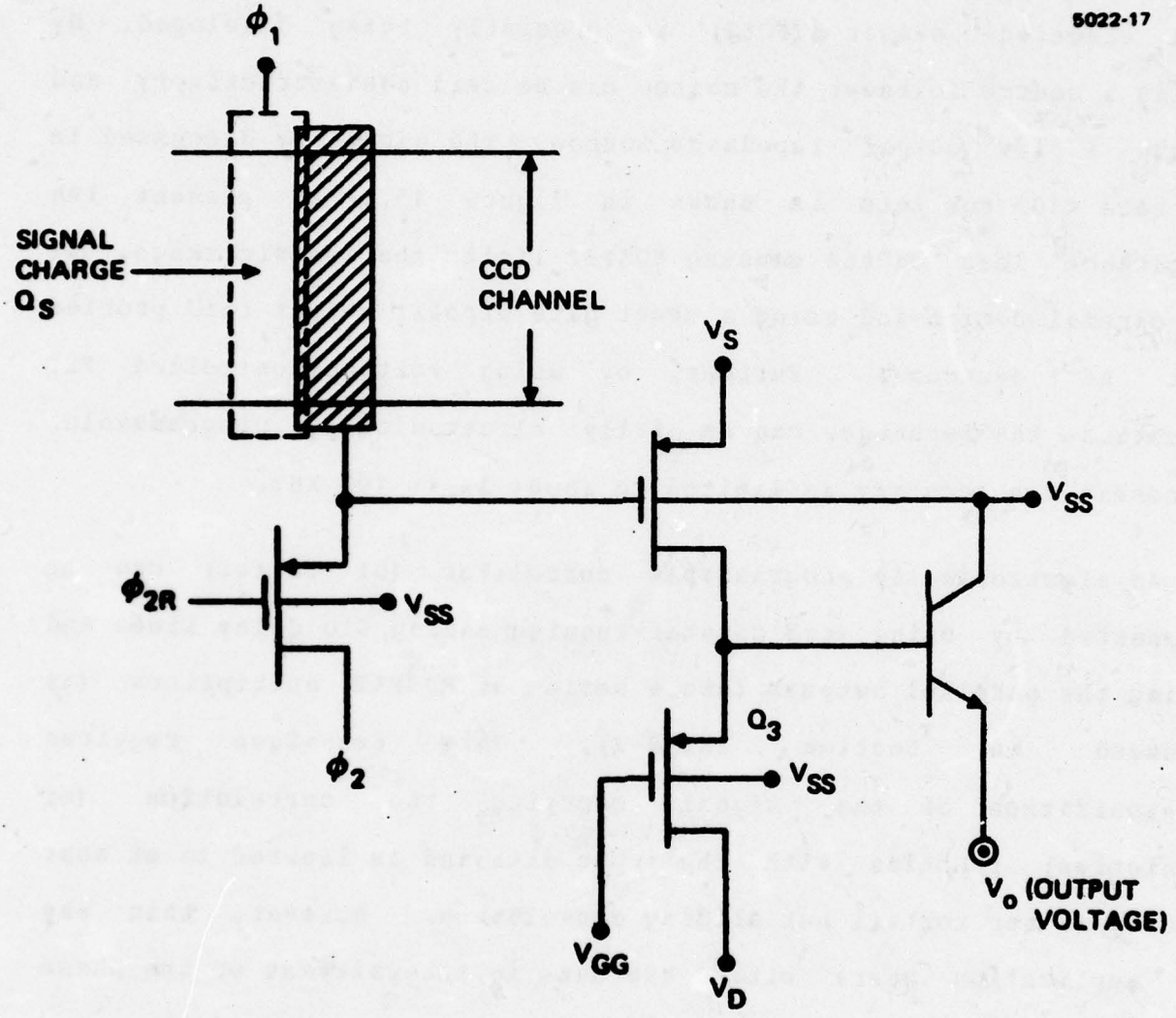


Figure 4.1-15. Floating clock electrode sensor circuit for nondestructive signal charge sensing.

Techniques to reduce the area of the silicon required and hence improve the reliability etc. have been suggested involving successive approximations and comparator operations. In general the CCD-implemented techniques are presently limited both in accuracy (number of bits) and speed.

Digital Logic: A full range of logic functions involving only CCD elements has been developed <11> and much of the interest in CCD device applications is centering on digital implementation, because of its inherent accuracy and the ease of regeneration. As an example of this type of processing two functions, the OR/AND gate and the EXCLUSIVE OR, are described.

A schematic of the layout for the OR/AND gate is shown in figure 16a. The two input signals are fed into parallel gates A and B with equal capacity. When a clock pulse is applied to the transfer gate the contents of both channels flow into the bucket under gate C. Since the area of this gate is identical to both A or B, a single input (or one) is sufficient to fill the bucket and be read as one. An implanted barrier separates this bucket from the succeeding one. In the event of both inputs A and B being "ones" the excess charge spills over to C and is read as a one, representing the AND output.

The accuracy of the technique is, of course, extremely high since binary decisions only are required and frequent refresh operations can be performed. The single control gate is the only active element and is not required to be linear. With buried channel devices the speed

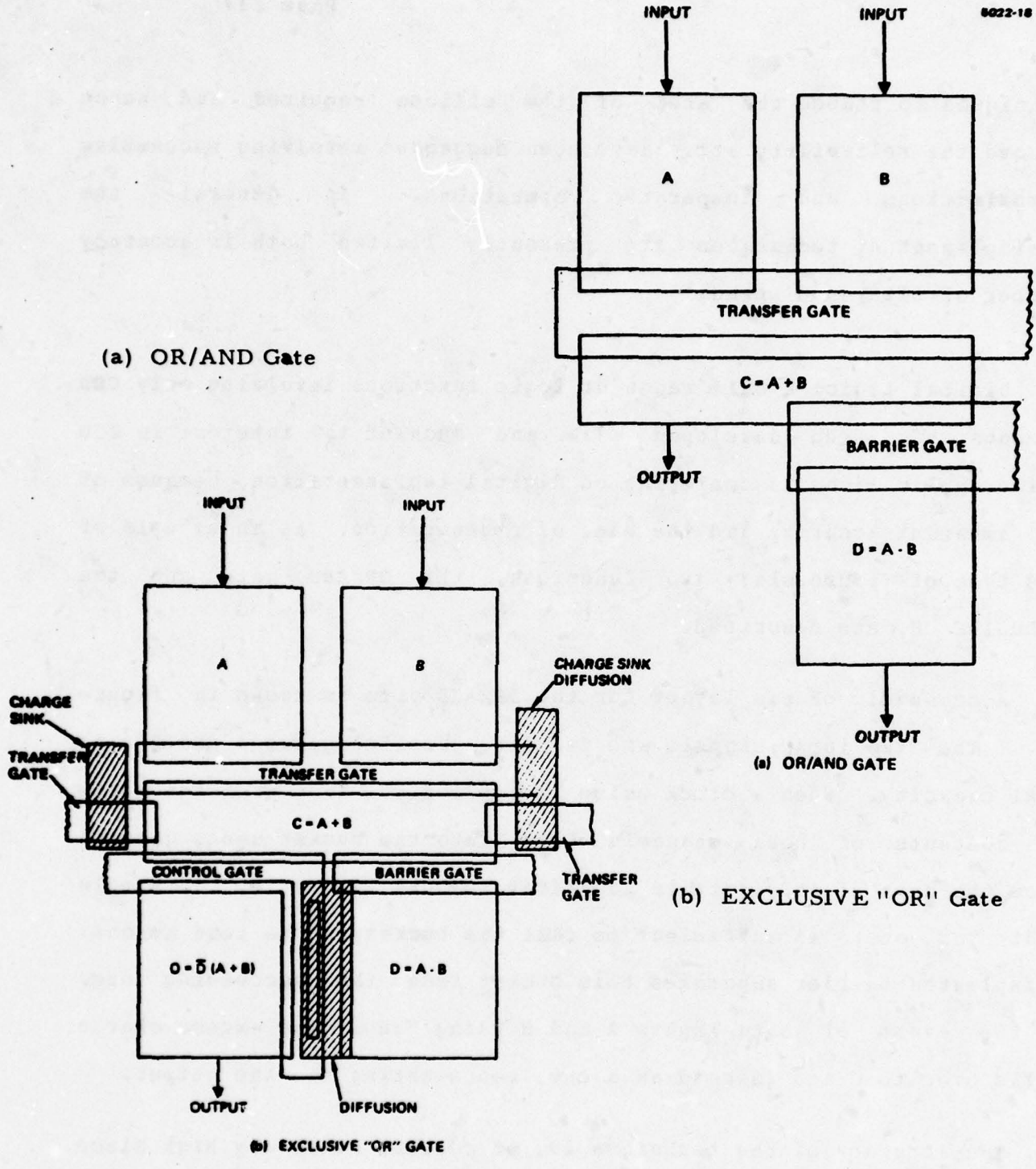


Figure 4.1-16. Digital logic gates implemented with CCDs.

is limited by the final input and output sensors.

The schematic of the EXCLUSIVE OR is shown in figure 16b. Again the two inputs are fed in via gates A and B, and the operation through gate C is as described above. However, the charge under gate C is sensed and connected to the control gate separating C and E. In the advent both inputs are "ones" the control gate maintains a voltage barrier over which the charge from gate C is prevented from flowing. The absence of charge then reads as a "zero." If only one input is present the control gate will not maintain the voltage barrier and the charge from C will flow into the final gate registering as a "one." The performance of this circuit relies on the effective operation of the refresh circuit as described in reference <11>. This requires that the floating diffusion accurately following the charge level under gate D. This requires effective decoupling from the surrounding fixed gate potentials. In practice the capacitive coupling to the adjacent gates and substrates is the limiting factor and high speed operation depends on reducing the overlap capacitance.

4.2 Chirp Transform Implementation Using CCDs

The chirp-Z transform (CZT) algorithm provides a technique for performing Fourier transformation, convolution and correlation of arbitrary signals in real time. It basically relies on restructuring the Fourier integral into two multiplications, and a linear filtering operation which can be performed with a transversal filter. It is an important processing technique because of its simplicity in terms of

hardware and because of its speed of operation. The most widely used technique for Fourier transformation at present is the fast Fourier transform (FFT). It requires $N\log_2(N)$ complex multiplications to form an N point transform as contrasted to N^2 operations for the direct technique. The chirp transform however can essentially operate in real time. The delay involved being only that of the filtering operation.

The advent of CCD transversal filters makes it possible to form transforms of 10^3 and 10^4 points with speeds as high as 100 MHz. With buried channel operation the basic transversal filter can operate at frequencies as high as 100 MHz but the bipolar sensing devices required to interface the basic CCD structure with the external circuitry presently limit their bandwidth to 20 MHz or so. It should be emphasized that CCD filters are operated as essentially base-banded devices and hence the maximum operating frequency is the device bandwidth, and the time bandwidth product is the number of stages.

The basic CZT algorithm derives directly from the discrete form of the Fourier transform

$$F_K = \sum_{n=0}^{N-1} f_n \exp\{\frac{-j2\pi nK}{N}\} \qquad (4.2\text{-}1)$$

where the data to be transformed f_n $(0 < n < N-1)$ can be considered to be uniform time samples of an input signal f(t). The transformed data F_K then represents a uniformly sampled version of the signal spectrum F(ω) where

$$F(\omega) = \int f(t)\, e^{-j\omega t}\, dt \qquad (4.2\text{-}2)$$

Bluestein et.al. <12> recognized that by expressing the complex exponent of eq.(1) in the form

$$2nk = n^2 + k^2 - (n-k)^2 \qquad (4.2\text{-}3)$$

the integration can be replaced by a multiplication and convolution. This can be seen by writing

$$F_K = e^{\frac{-j\pi k^2}{N}} \sum_{n=0}^{N-1} f_n\, e^{\frac{-j\pi n^2}{N}}\, e^{\frac{j\pi (k-n)^2}{N}} \qquad (4.2\text{-}4)$$

which is the form of the CZT. Signals of the form

$$\exp\{\frac{\pm j\pi k^2}{N}\}$$

and

$$\exp\{\frac{j\pi n^2}{N}\}$$

are known as chirp signals from pulse compression radar <13> and are basically signals whose frequency is linearly related to time. The hardware to perform the algorithm of eq.(4) is shown in figure 1. The symbol * represents convolution, in this figure, with a chirp

$$\exp\{\frac{j\pi n^2}{N}\}$$

such that

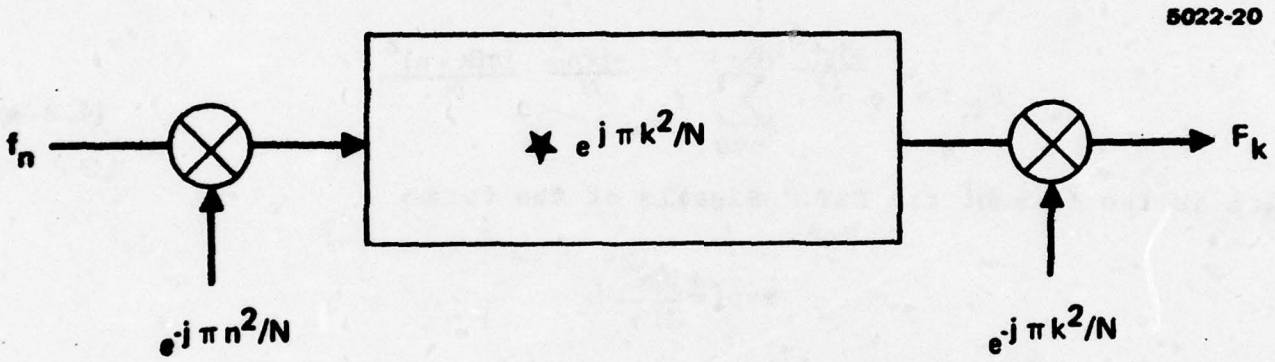


Figure 4.2-1. Hardware implementation of chirp transform.

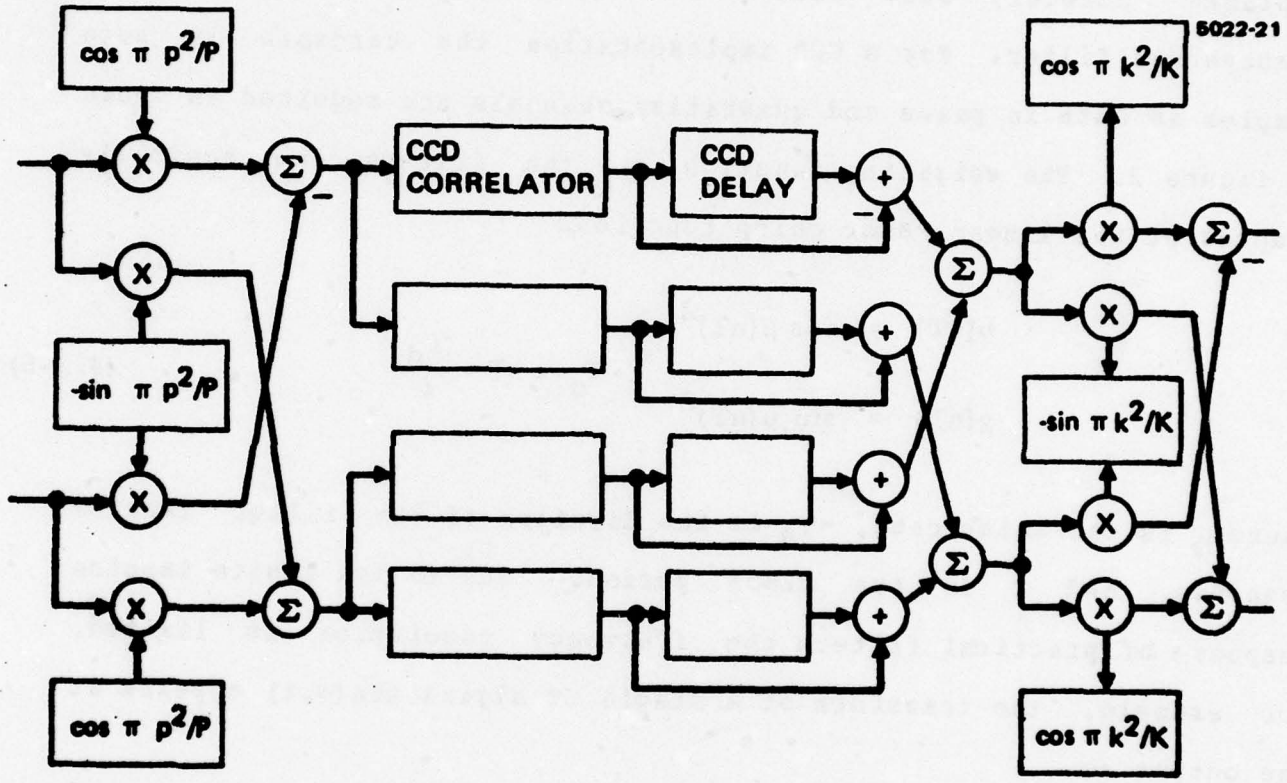


Figure 4.2-2. CCD chirp z-transform.

$$S_o = \sum_{n=1}^{N-1} \sin \exp\{\frac{-j\pi(k-n)^2}{N}\} \qquad (4.2\text{-}5)$$

This is the basic operation of a chirp filter and hence can be implanted directly with either a surface acoustic wave or CCD transversal filter. For a CCD implementation the hardware is more complex as both in phase and quadrative channels are required as shown in figure 2. The weighting function on the filters are basically samples of the linear FM or chirp functions

$$h(nT) = \cos \mu(nT)^2$$
$$g(nT) = \sin \mu(nT)^2 \qquad -T_d < nT < \frac{T_d}{2} \qquad (4.2\text{-}6)$$

where μ is the chirp rate, $-T_d$ is the duration of the filter impulse response, and T is the clock period. Due to the finite impulse response of practical filters the frequency resolution is limited. For example, the transform of a single CW signal $\sin(\omega,t)$ appears at the output as

$$\delta(\omega - \omega_1) \bullet \text{sinc}\,(\mu T_d t)$$

Presently the thermal generation of carriers within the CCD cells limits the maximum storage time T_d to about 1 sec, and the frequency resolution to about 1 Hz. The time bandwidth product of a CCD transversal filter is simply equal to the number of stages N, i.e.

$$B \times T_d = f_c \times \frac{N}{f_c}$$

where f_c is the clock frequency. To implement the conventional CZT,

the filters are weighted, using the split electrode technique (discussed in Section 4.1), to form a chirp from $-f_c$ to $+f_c$ and the premultiplying waveform has a duration N/f_c which chirps from dc to $-f_c$. For a single frequency input at frequency f_1 the output of the first mixer sweep from $-f_c$ to $+f_1$. This waveform is shifted through the filter and forms a correlation peak at time

$$t_1 = \frac{N}{f_c^2} f_1 \quad . \qquad (4.2\text{-}7)$$

Thus the frequencies components appear time ordered at the processor output, the proportionality constant between frequencies and time being N/f_c^2, the filter chirp rate. The final post multiplication simply corrects the phase of the spectral components and is often omitted when the phase is not required.

The accuracy of the transformation is limited primarily by the accuracy of the CCD filter and the multipliers. Factors affecting the filter response include linearity, accuracy of weighting technique and thermal noise generation.

CCD chirp transform processors have been implemented at 100 kHz with frequency resolution of 200 Hz. The technique has an advantage over surface wave acoustic implementations in that the timing can be established precisely by stable clocking circuits. Further, the very low insertion loss associated with the CCD process allows a dynamic range of approximately 80 dB to be obtained. The principle source of noise is the differential current integrator at the output. Present

state of the art indicates that the chirp transform process might be operated at 20 MHz with 10^3 resolution elements. The inaccuracies included in the mask pattern used to generate the weighting limits the amplitude resolution to about 0.1%. The nonuniform etching rate and oxide thickness together with mask misalignments contribute to an overall error of 0.5%. The largest CCD chirp transform processor reported to date has 500 stages <14> and a transfer efficiency of 0.9999. At present the chirp signals are generated off chip in a 8 bit ROM and the multiplier is a discrete multiplying D/A. The accuracy is limited by the dynamic range in the input and output circuitry and nonlinearities in the multiplying circuits. Further, the dispersion caused by the finite transfer inefficiency of 1 part in 10^{-4} cause a distortion term of amplitude $(NP_\epsilon/2)^2$ times the desired response.

In concept all the functions involved in the process can be accomplished by CCDs. For example the multiplication can be accomplished by the CCD multiplier D/A concept discussed in Section 4.2 and the multiplying chirps might be stored in CCD memory. By using electron beam exposure the accuracy of the multiplying D/A might be made equivalent to 10 bits plus sign and integrated on the chirp. However, a capable MOS ROM would be a more effective technique since the CCD memory is basically a dynamic concept and subject to degradation. The speed of the process is basically limited by the capacitive area and hence an increase in resolution to say 1 μm from the present 5 μm might lead to a system operating at about TV rates.

4.3 Real Time Nonlinear Optical Data Processing

During January 1975, we began a study of real time nonlinear optical data processing. We are using a hybrid field effect liquid crystal light valve <15> as an incoherent to coherent imaging interface. This device is used as a real time substitute for hard-clipping film to accomplish pulse width encoding of continuous tone picture information <16>.

The light valve is used in the coherent optical data processing system shown in figure 1. The incoherent input light derived from a mercury arc lamp filtered at a center wavelength of 525 nm. This light is modulated by a continuous tone transparency in contact with a grating designed for each nonlinear operation. The experiments to date have been concerned with the nonlinear operation of image level slicing. The appropriate grating for this operation is a Ronchi ruling <16>. The Ronchi grating used has a fundamental spatial frequency of 12 cy/mm and optical density levels of 0.21 and 0.68. The grating modulated scene is imaged onto the light-sensitive CdS layer of the light valve. The resultant modulation of the CdS conductivity produces a modulation of the coherent light amplitude reflected from the light valve. (For a detailed description of the light valve structure and function, the reader should see reference <15>). The coherent light source is a 50 mW He-Ne laser.

The set-up of figure 1 approximated the level-slicing experiment of Dashiell and Sawchuk (see reference <16>). The desired

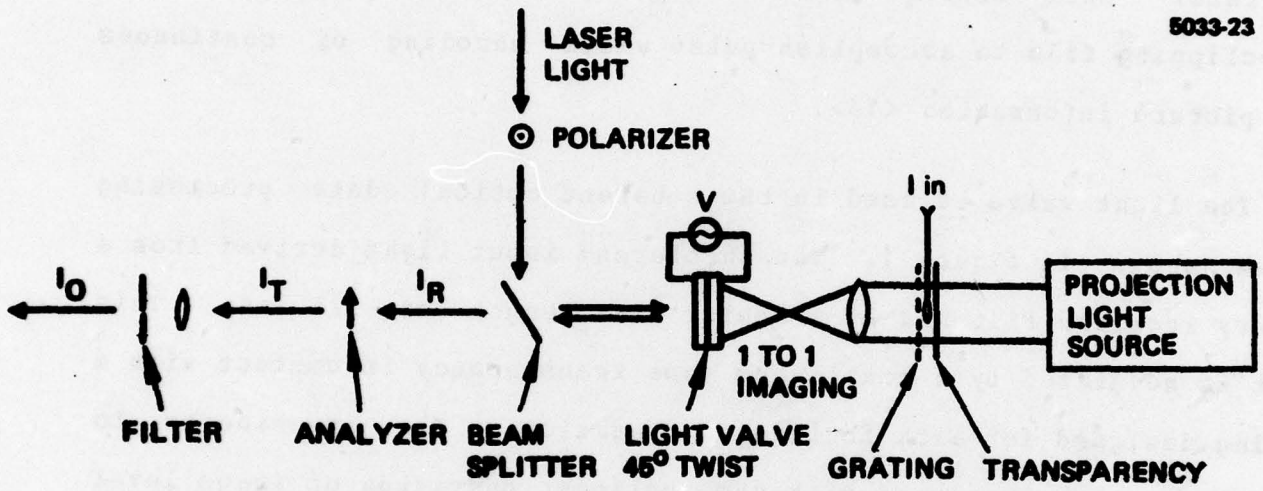


Figure 4.3-1. Coherent optical data processor.

input/output intensity relationship for the light valve has a high gamma. This high gamma hard-clips the grating modulated image. The resultant coherent light output is then pulse width modulated into three distinct states: no modulation (dark), square wave modulation at the Ronchi ruling spatial frequency and no modulation (bright). Band-pass spatial filtering of the spatial frequencies around the t first diffraction order of the grating produces an output image of the input transparency. This output image shows only a narrow slice of the total density range of the input. The density value for the selected level is determined by the intensity of the noncoherent input illumination to the light valve.

Figure 2 shows the actual input/output intensity relationship achieved with the light valve used in the experiments to date. These data were recorded with a (typical) light valve drive of 4.0 volts rms at 2.0 kHz. Figure 2 shows that this valve has a maximum gamma of 2.1.

Figure 3 shows the transfer function achieved in the level slicing experiment. The gamma of the light valve is too low to achieve an ideal level slice. However, a broad level slice is observed. The filtered first order intensity increases to a peak value corresponding to maximum modulation of the reflected coherent light. Beyond that peak, the first order output intensity decreases. This decrease is caused by saturation of the light valve at high input intensities.

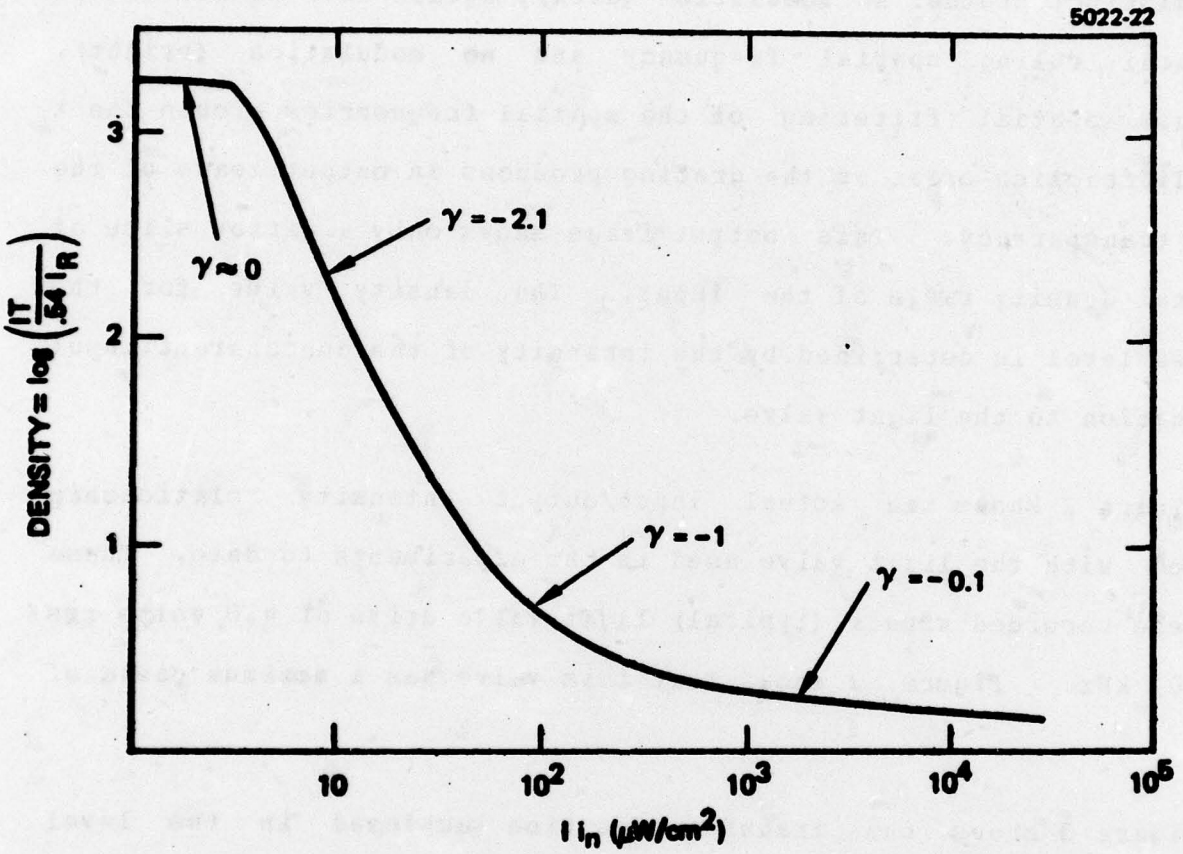


Figure 4.3-2. Light valve transfer function.

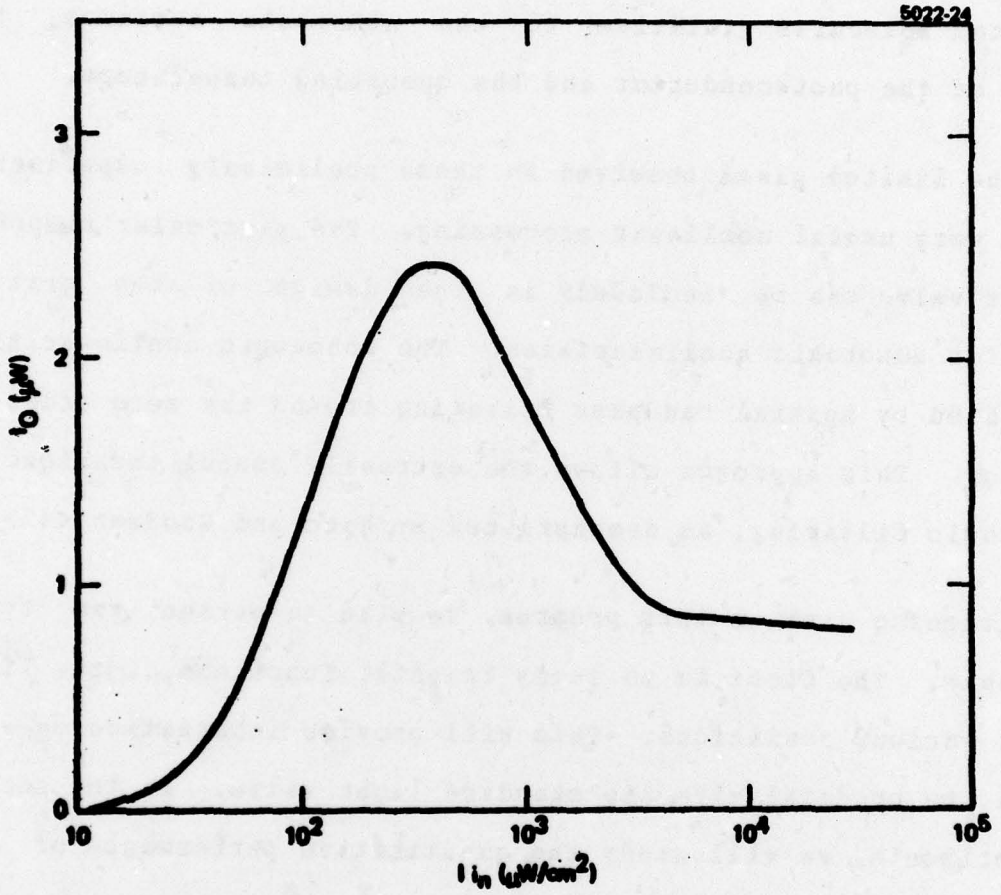


Figure 4.3-3. Light valve transfer function for level slicing.

The sensitometry of the light valve depends on the drive voltage and frequency. This drive can be varied to maximize the gamma of the light valve. Other possible techniques for increasing the gamma include modification of the liquid crystal type, the alignment of the liquid crystal molecules relative to the substrate surfaces, the preparation of the photoconductor and the operating temperature.

Even the limited gamma observed in these preliminary experiments can provide very useful nonlinear processing. The particular response of the light valve can be included in the design of the grating modulator, for monotonic nonlinearities. The monotonic nonlinearities are implemented by spatial bandpass filtering around the zero order of the grating. This approach allows the extremely useful technique of the logarithmic filtering, as demonstrated by Kato and Goodman <17>.

In continuing work on this program, we plan to pursue two types of experiments. The first is to study transfer functions of the light valve under various conditions. This will provide information on what gammas can be produced with the standard light valve. In the second set of experiments, we will study the quantitative performance of the light valve in producing monotonic nonlinear response functions.

References

1. W.S. Boyle, G.E. Smith, "Change Coupled Semiconductor Devices," Bell System Technical Journal, Vol. 49, 1970, pp. 587-593.

2. M.F. Tompsett, "Charge Transfer Devices," Journal Vac. Sci.

Technology, Vol. 9, 1972, pp. 1166-1181.

3. L.J.M. Esser, "The Peristaltic Charge-Coupled Device for High Speed Charge Transfer," Proceedings 1974 IEEE Solid State Conference, pp. 28-29.

4. C.H. Sequin and M.F. Tompsett, Charge Transfer Devices, Academic Press, New York, 1975.

5. D.F. Barbe, Imaging Devices Using the Charge Coupled Concept, Proceedings IEEE, Vol. 63, No. 1, January 1975.

6. K.K. Thornber, "Theory of Noise in Charge Transfer Devices," Bell System Technical Journal, Vol. 53, pp. 1211-1261.

7. J.E. Carnes and W.F. Kosonocky, "Noise Sources in Charge Coupled Devices," RCA Rev. Vol. 33, June 1972, pp. 327-343.

8. Patrick Bosshart, An Integrated Analog Correlator using Charge-Coupled Devices, 1976 IEEE International Solid State Circuits Conference, Philadelphia, February 1976.

9. J.F. Albaran, D.A. Hodges, A Charge Transfer Multiplying D to A Converter, 1976 IEEE International Solid State Conference, Philadelphia, February 1976.

10. Marvin H. White, Ingham A. Mack, Francis J. Kub, Donald R. Lampe and John L. Fagan, An Analog CCD Transversal Filter with Floating Clock Electrode Sensor and Variable Tap, IEEE International Solid State Conference, Philadelphia, February 1976.

11. W.F. Kosonocky and L.E Carnes, "Charge-Coupled Digital Circuits," IEEE Journal Solid State Circuits, SC-6, 1971, pp. 314-322.

12. L.I. Bluestein, "A Linear Filtering Approach to the Computation of Discrete Fourier Transform," NEREM Record No. 10, November 1968, published by Boston Section IEEE, pp. 218-219.

13. C.E. Cook, M. Bernfeld, "Radar Signals," Academic Press, New York, 1967.

14. R. Brodersen, C.R. Hewes, D.D. Buss, A 500 Stage CCD Transversal Filter for Spectral Analysis, IEEE Journal SSC, Vol. SC-11, No. 1, February 1976.

15. J. Grinberg, et.al., Optical Engineering, Vol. 14, 1975, pp. 217-225.

16. S.R. Dashiell and A.A. Sawchuk, Optics Communication, Vol. 15, 1975, pp. 66-69.

17. H. Kato and J.W. Goodman, Applied Optics, Vol. 14, 1975, pp. 1813-1824.

5. Publications

The following is a list of papers, articles, and reports of research staff members of the USC Image Processing Institute which have been published or accepted for publication during the past seven months, and which have resulted from ARPA sponsored research.

1. H.C. Andrews, "Monochrome Digital Image Enhancement," Applied Optics, Vol. 15, No. 2, February 1976, pp. 495-503.

2. H.C. Andrews and B.R. Hunt, Digital Image Restoration, Prentice Hall, Englewood Cliffs, New Jersey, 1977.

3. H.C. Andrews and C.L. Patterson, "Digital Interpolation of Discrete Images," IEEE Transactions on Computers, Vol. C-25, No. 2, February 1976, pp. 196-202.

4. H.C. Andrews and C.L. Patterson, "Outer Product Expansions and their Uses in Digital Image Processing," IEEE Transactions on Computers, Vol. C-25, No. 2, February 1976, pp. 140-148.

5. H.C. Andrews and C.L. Patterson, "Singular Value Decompositions and Digital Image Processing," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-24, No. 1, February 1976, pp. 26-53.

6. H.C. Andrews and C.L. Patterson, "Singular Value Decomposition (SVD) Image Coding," IEEE Transactions on Communications, Vol. COM-24, No. 4, pp. 425-432.

7. S.R. Dashiell and A.A. Sawchuk, "Analysis and Synthesis of Nonlinear Optical Processing," Proceedings Electro-Optics/International Laser Conference 1975, Anaheim, California, November 1975, (invited paper).

8. S.R. Dashiell and A.A. Sawchuk, "Optical Synthesis of Nonlinear Nonmonotonic Functions," Optics Communications, Vol. 15, September 1975, pp. 66-70.

9. S.R. Dashiell and A.A. Sawchuk, "Optical Synthesis of Nonlinear Nonmonotonic Functions," 1975 Annual Meeting, Optical Society of America, Boston, October 1975, Journal of the Optical Society of America, Vol. 65, October 1975, p. 1177.

10. S.R. Dashiell and A.A. Sawchuk, "Synthesis of Nonlinear Nonmonotonic Functions in Optical Image Processing," Proceedings of the Optical Society of America Topical Meeting on Image Processing, Asilomar, California, February 1976, Journal of the Optical Society of America, Vol. 66, February 1976, p. 170.

11. W. Frei and G.S. Robinson, "Color Image Coding with a Visual Fidelity Criterion," Picture Coding Symposium, January 1976.

12. W. Frei and G.S. Robinson, "Final Research Report on Computer Processing of ERTS Images," USCEE Report 640, September 1975.

13. H.S. Hou and H.C. Andrews, "Fundamental Limits and Degrees of Freedom of Imaging Systems," Proceedings of the OSA Topical Meeting on Image Processing, Asilomar, California, February 1976.

14. F. Naderi and A.A. Sawchuk, "Nonlinear Digital Processing of Images Corrupted by Film-Grain Noise," 1975 Annual Meeting, Optical Society of America, Boston, October 1975, Journal of the Optical Society of America, Vol. 65, October 1975, p. 1202.

15. F. Naderi and A.A. Sawchuk, "Nonlinear Detection and Estimation of Images Degraded by Film-Grain Noise," Proceedings of the Optical Society of America Topical Meeting on Image Processing, Asilomar, California, February 1976, Journal of the Optical Society of America, Vol. 66, February 1976, p. 167.

16. R. Nevatia, "Image Segmentation," presented at the Workshop on Data Structures and Pattern Recognition, Albuquerque, New Mexico, February 11-13, 1976.

17. R. Nevatia, "Locating Object Boundaries in Textural Environments," accepted for publication in IEEE Transactions on Computers.

18. M.J. Peyrovian and A.A. Sawchuk, "Image Processing by Smoothing Spline Functions," Proceedings of the Optical Society of America Topical Meeting on Image Processing, Asilomar, California, February 1976, Journal of the Optical Society of America, Vol. 66, February 1976, p. 167.

19. A.A. Sawchuk, "Real-Time Correction of Intensity Nonlinearities in Imaging Systems," accepted for publication in IEEE Transactions on Computers.

20. W.K. Pratt, "Digital Image Enhancement and Restoration Display Techniques," Society for Information Display, 1976 SID International Symposium, Los Angeles, California, May 1976.

21. W.K. Pratt, "Scene Analysis Image Coding Concepts," Picture Coding Symposium, Pacific Grove, California, January 1976.

22. W.K. Pratt, "Survey and Analysis of Image Coding Techniques," Invited Paper, Optical Society of America, Topical Meeting on Image Processing, Pacific Grove, California, February 1976.

23. W.K. Pratt and C.E. Mancill, "Spectral Estimation Techniques for the Spectral Calibration of a Color Image Scanner, Applied Optics, November 1975.