USCEE REPORT #444

Semiannual Technical Report
Covering Research Activity During the Period
1 September 1972 to 28 February 1973

by

William K. Pratt

February 1973

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

ABSTRACT

This technical report summarizes the image processing research activities performed by the University of Southern California during the period of 1 September 1972 to 28 February 1973 under Contract No. F08606-72-C-0008 with the Advanced Research Projects Agency, Information Processing Techniques Office.

The research program, entitled, "Image Processing Research," has as its primary purpose the analysis and development of techniques and systems for efficiently generating, processing, transmitting, and displaying visual images and two dimensional data arrays. Research is oriented toward digital processing and transmission systems. Five task areas are reported on: (1) Image Coding Projects, the investigation of digital bandwidth reduction coding methods; (2) Image Enhancement and Restoration Projects: the improvement of image fidelity and presentation format; (3) Image Data Extraction Projects: the recognition of objects within pictures and quantitative measurement of image features; (4) Image Analysis Projects, the development of quantitative measures of image quality and analytic representation; (5) Image Processing Support Projects, development of image processing hardware and software support systems.

PROJECT PARTICIPANTS

Project Director

William K. Pratt

Research Faculty

Harry C. Andrews
Lee D. Davisson
Ali Habibi
Ronald S. Hershel
Anil K. Jain
Richard P. Kruger
Nasser E. Nahi
Alexander A. Sawchuk
Lloyd R. Welch

Support Staff

Angus B. Cossey
Carolyn Matthews
James M. Pepin
K. A. Pratt
Michael Reichik
Mark A. Sanders
John Tahl
Linda M. Webster

Students

Touraj Assefi
Wen-Hsiung Chen
Faramarz Davarian
Norman Diamond
Roy M. Glantz
Michael Huhns
Mohammad Jahanashahi
Joseph Jimerson
Ralph Larson
Robert Liles
Eduardo Lopez
Clanton Mancill
Nelson Mascarenhas
Firouz Naderi
Sumeet Pasricha
Michael Patton
Clifford Reader
Stuart Robinson
Dennis Smith
Wai Sun Szeto
Andrew Tescher
Fred Tydeman
Robert Wallis
Carl Wedberg
Pamela Welch

TABLE OF CONTENTS

1. Research Project Overview

This report describes the progress and results of the University of Southern California image processing research study for the period of 1 September 1972 to 28 February 1973.

The image processing research study has been subdivided into five projects:

Image Coding Projects
Image Restoration and Enhancement Projects
Image Data Extraction
Image Analyses Projects
Image Processing Support Projects

In image coding the orientation of the research is toward the development of digital image coding systems that represent monochrome and color images with a minimal number of code bits. Image restoration is the task of improving the fidelity of an image in the sense of compensating for image degradations. In image enhancement, picture manipulation processes are performed to provide a more subjectively pleasing image or to convert the image to a form more amenable to human or machine analysis. The objectives of the image data extraction projects are the registration of images, detection of objects within pictures and measurements of image features. The image analysis project comprises the background research effort into the basic structure of images in order to develop meaningful quantitative characterizations of an image. Finally, the image support projects include research on image processing computer languages and the development of experimental equipment for the sensing, processing, and display of images.

The next section of this report summarizes some of the research project activities during the past six months. Sections 3 to 7 describe the research effort on the projects listed above during the reporting period. Section 8 contains a short description of new projects that are

being initiated, and are not yet to the reporting stage. Section 9 is a list of publications by project members.

2. Research Project Activities

The following sections describe some of the significant project activities of the past six months:

Image Processing Institute. The USC group involved in image processing research has been designated as an Institute within the School of Engineering. Other Institutes of the School include: the Information Sciences Institute, the Biomedical Engineering Institute, and the Transportation Institute.

USC Engineer Special Issue. The December 1972 issue of the USC Engineer--a quarterly magazine published by USC Engineering students--has been devoted to image processing research at USC. The magazine contains several survey articles at the "Scientific American" level describing image processing topics. Copies may be obtained from the Institute.

Image Coding Symposium. The University of Southern California hosted the Fourth Picture Coding Symposium on 22-24 January 1973. Over eighty research workers in the field attended, with a third of the attendance from foreign countries. Topics of the conference included: properties of the human observer; facsimile coding; intraframe picture coding; interframe television coding; color image coding; and multispectral image data coding. An image coding contest was held at the conference to determine the best image coding algorithms and systems for monochrome and color images. The Slant transform coding system, developed at USC, received three of the seven achievement awards presented at the conference.

3. Image Coding Projects

The research effort in image coding has been directed toward a wide variety of applications. Coding systems are under investigation for: monochrome and color imagery; slow scan and real time television; and information preserving and controlled fidelity operation. The results of this research study during the past six months are summarized here and presented in greater detail in subsequent sections.

A study of the application of slant transform coding techniques to natural color images is the subject of the first report. In the system developed, the red, green, and blue sensor signals are converted to the conventional YIQ luminance and chrominance signals, which in turn are individually transformed in 16 x 16 pixel blocks. Efficient quantization algorithms have been developed which enable coding with as few as 2.0 to 3.0 bits/pixel. Color reproductions of the coded images are included in the report.

The next report is concerned with the development of adaptive quantization techniques for transform domain coefficients. Techniques have been found which permit a reduction of about 0.2 to 0.4 bits/pixel above that possible with conventional non-adaptive quantization methods.

A study has been performed on optimal means of quantizing Fourier transform coefficients for image coding. It has been found, that for coarse quantization, it is more efficient to quantize the magnitude/phase representation than the real/imaginary representation. The results of this work have been applied to Fourier transform coded frame differences for a real time television system.

Transform coders provide a relatively high degree of compression-picture quality performance, but their implementation complexity is somewhat greater than other types of coders. A research effort has been initiated to study a system in which pixels are transform coded along scan lines, and coded between scan lines by a differential pulse code modulation

(DPCM) technique. The process offers the potential of nearly as efficient coding as a two dimensional transform coder, but with much less complexity.

In the following report consideration has been given to a new technique of DPCM image coding in which time adjacent frames are differenced, and the difference signal is coded by contour tracing. Preliminary results indicate that the performance of the system will surpass that of conventional frame-to-frame coders which simply quantize the frame differences.

Contour tracing is also of interest in facsimile coding of binary (black or white) images. A theoretical study has been performed to establish performance bounds for systems which code the contours of facsimile images. The results indicate that contour coding is superior to conventional run length coding for imagery such as script and weather maps, but does not perform as well for typewritten documents containing many contours.

The final report involves one of the first practical applications of universal coding to video data. A computer simulation has been performed to evaluate a particular class of information preserving universal codes. It was found that coding can be performed at a rate below the average entropy of pictures, by taking advantage of the local statistical structure of the images.

3.1 Slant Transform Color Image Coding

Wen-Hsiung Chen and William K. Pratt

The slant transform has been applied quite successfully to obtain a bandwidth reduction and tolerance to channel errors for monochrome images [1]. Studies indicate that the spatial redundancy of color images and the limitations of human color vision can be exploited by slant transform coding to achieve a bandwidth reduction for color image transmission [2].

Slant Transformation A separable, unitary transformation of an image array f(j,k) into an array of transform coefficients F(u,v) may be defined by the matrix equation

$$[F(u,v)] = [A][f(j,k)][A]^T$$

where [A] is a matrix whose rows are basis vectors of the transformation. A desireable property for image coding is that the transform compact the image energy into as few of the transform domain samples as possible. The energy compaction of a unitary transformation will be "high" if the basis vector "resemble" typical lines of an image. In most natural images a large number of lines or line segments are of nearly constant brightness. Also, many lines and line segments either increase or decrease in brightness over their length in a nearly linear fashion. With this rationale, a family of basis vectors containing a constant and a sawtooth basis vector has been developed for image coding. The remaining basis vectors have been chosen so that the sequency (number of zero crossings) of each basis vector is equal to its row number minus one. Also, by design, the resulting set of basis vectors possesses a fast computational algorithm.

Figure 1 contains a block diagram of the slant transform color image coding system. In the system the color image is represented by three source tristimulus signals R(j,k), G(j,k), B(j,k) that specify the red, green, and blue content of a pixel at coordinate (j,k), according to the National Television System Commission (NTSC) receiver phosphor primary system. The source tristimulus signals are then converted to a new three dimensional space Y(j,k), I(j,k), Q(j,k) which specify the luminance and the chrominance information of the image pixel according to the NTSC television transmission primary system. The conversion is defined by

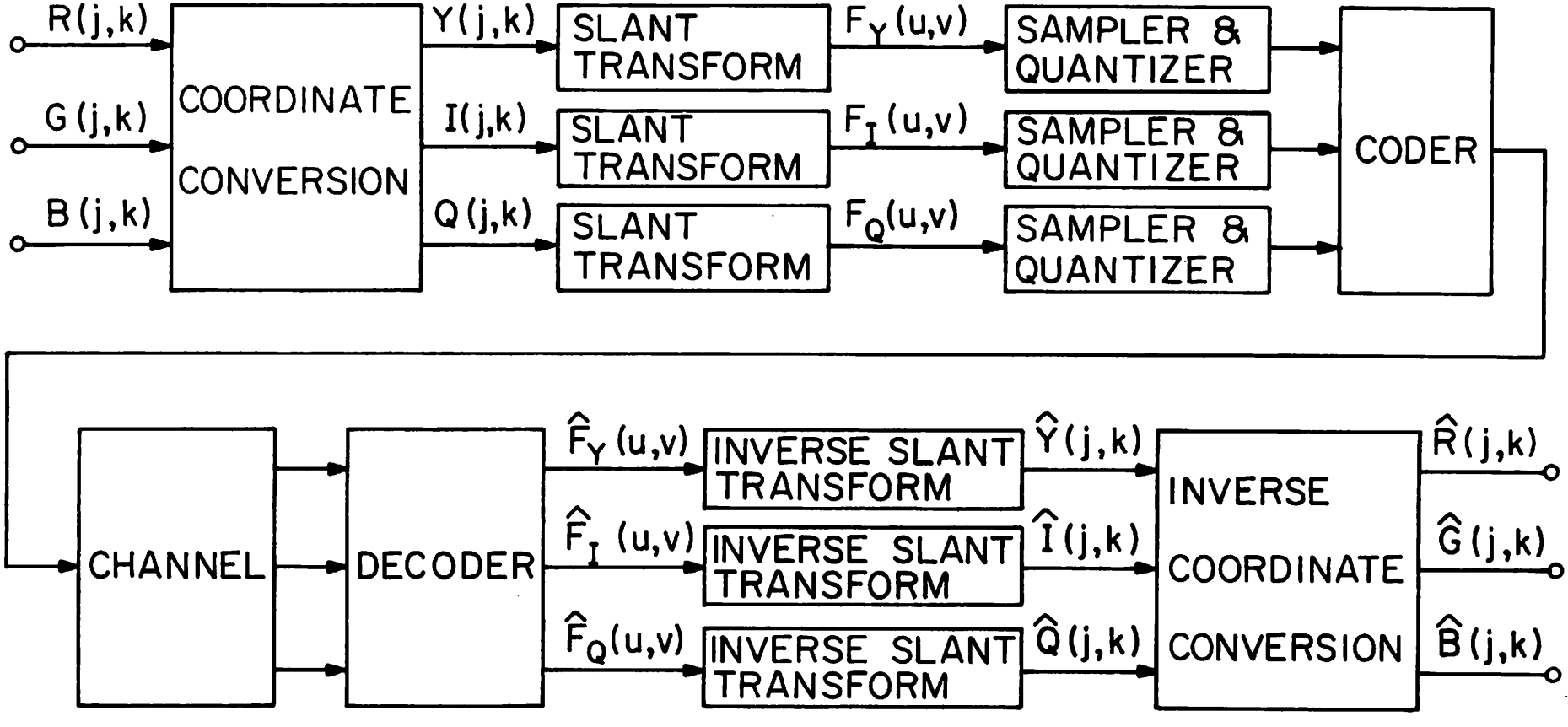


Figure 3.1-1. Block diagram of Slant transform color image coding system.

$$\begin{bmatrix} Y(j,k) \\ I(j,k) \\ Q(j,k) \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.253 & 0.312 \end{bmatrix} \begin{bmatrix} R(j,k) \\ G(j,k) \\ B(j,k) \end{bmatrix}$$

The reason for transform coding the YIQ signals rather than the RGB signals is that YIQ signals are reasonably well uncorrelated and most of the color image energy is compacted into the Y plane.

The converted signals then individually undergo a two dimensional slant transform over the entire image, or repeatedly over subsections of the image, called blocks. This results in three transform domain planes $F_Y(u,v)$, $F_I(u,v)$, $F_Q(u,v)$ obtained from

$$[F_Y] = [S][Y][S]^T$$

$$[F_I] = [S][I][S]^T$$

$$[F_Q] = [S][Q][S]^T$$

where [S] is the slant transform matrix. Next, the transform samples are quantized with the number of quantum levels made proportional to the expected variance of each pixel, and with the quantization level spacing allowed to be variable to minimize the mean square quantization error. The quantized samples $\hat{F}_Y(j,k)$, $\hat{F}_I(j,k)$, and $\hat{F}_Q(j,k)$ are then coded and transmitted over a possibly noisy channel. At the receiver the channel output is decoded, and inverse slant transforms are taken to obtain

$$[\hat{Y}] = [S]^T [F_Y][S]$$

$$[\hat{I}] = [S]^T [F_I][S]$$

$$[\hat{Q}] = [S]^T [F_Q][S]$$

Finally, an inverse coordinate conversion results in the reconstructed tristimulus signals

$$\begin{bmatrix} \hat{R}(j,k) \\ \hat{G}(j,k) \\ \hat{B}(j,k) \end{bmatrix} \begin{bmatrix} 1.000 & 0.956 & 0.621 \\ 1.000 & -0.272 & -0.647 \\ 1.000 & -1.106 & 1.703 \end{bmatrix} \begin{bmatrix} \hat{Y}(j,k) \\ \hat{I}(j,k) \\ \hat{Q}(j,k) \end{bmatrix}$$

Performance A bandwidth reduction is achieved with the slant transform color image coding system by restricting the number of code bits assigned to the quantized transform coefficients. Efficient quantization strategies has been developed to minimize the mean square quantization error for a given bit assignment [3].

A computer simulation has been performed to subjectively evaluate the performance of the slant transform color image coding system. Figure 2 contains monochrome photographs of the red, green, and blue components of an original image of 256 by 256 pixels. Each component of the original is quantized to 255 levels. It should be noted that visually, the R G B components are highly correlated. The corresponding Y I Q components in Figure 2 appear much less correlated. Figure 3 contains photographs of the logarithm of the magnitude of each slant transform plane of the color image for transformation in 16 by 16 pixel blocks to illustrate the spatial energy compaction. In one of the simulation experiments the transform coefficients, F_Y, F_I, F_Q were assigned code bits such that $\hat{Y}, \hat{I}, \hat{Q}$ were coded with an average of 1.2, 0.54 and 0.26 bits/pixel, respectively. The corresponding reproductions of $\hat{Y}, \hat{I}, \hat{Q}$ and $\hat{R}, \hat{G}, \hat{B}$ are shown in Figure 3. In this experiment the coding has been reduced from 24 bits/pixel to 2.0 bits/pixel. The R G B reconstructions exhibit some degradation as a result of the coding process, but the visual effect of the degradation is much less visible in the color reconstruction because of the spatial frequency limitations of the human visual system. Figure 4 contains color reproductions of two original color images and coded versions of the images coded with an average of 2.0 and 3.0 bits/pixel.

R G B

Y I Q

Figure 3.1-2. RGB and YIQ components of a color image.

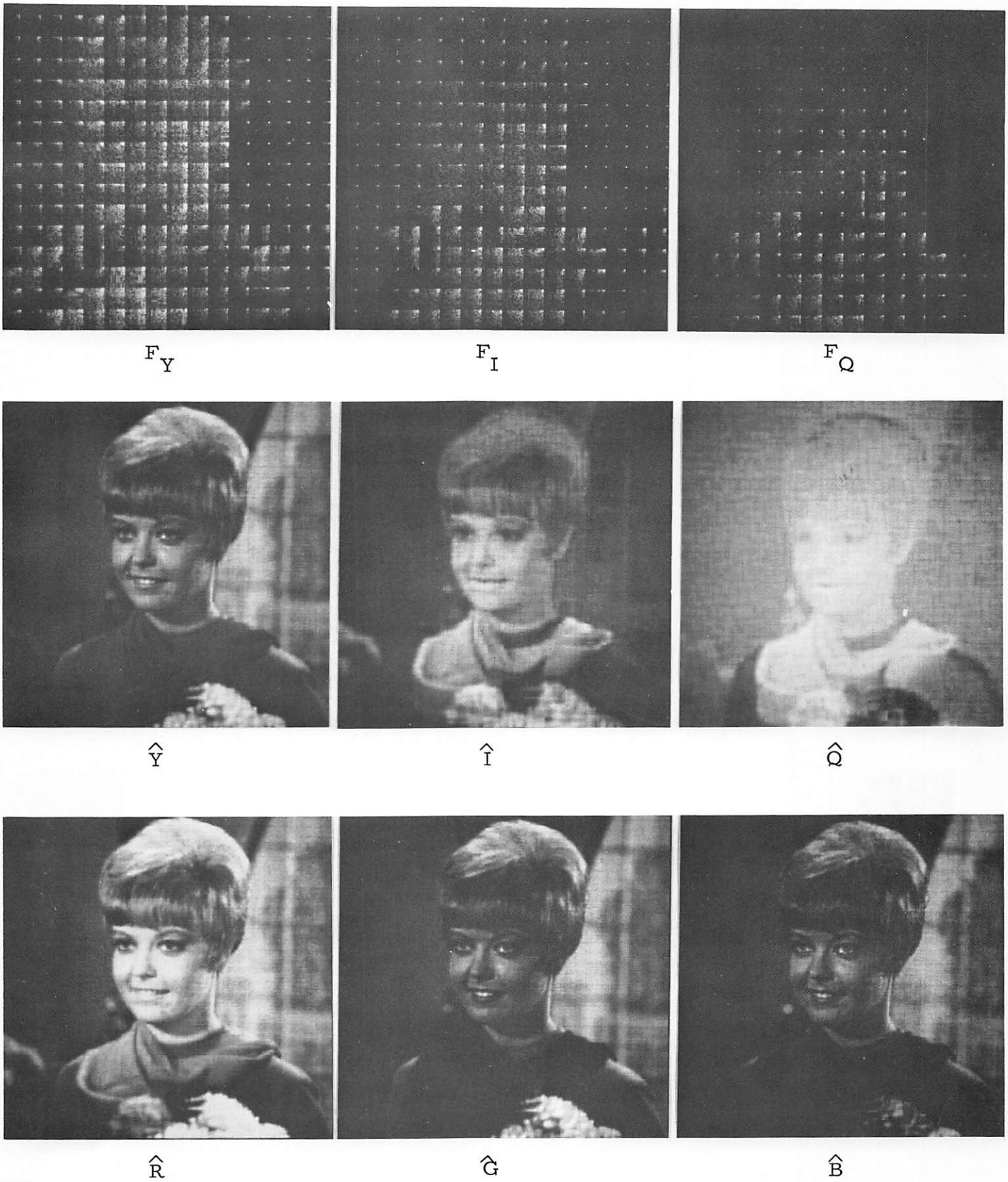


Figure 3.1-3. Slant transform domains of Y, I, Q; reduced bandwidth representations of Y, I, Q; and R, G, B components of a color image. Color image coded with 2.0 bits/pixel.

original

coded, 3 bits/pixel

coded, 2 bits/pixel

Figure 3.1-4. Original color image and coded reconstructions for Slant transform color image coding.

References

1. W. K. Pratt, L. R. Welch, and W. Chen, "Slant Transforms for Image Coding," Proceedings of the Symposium on Application of Walsh Functions, March, 1972.

2. W. K. Pratt, "Spatial Transform Coding of Color Images," IEEE Transactions on Communication Technology, Vol. COM-19, No. 6, December, 1971, pp. 980-992.

3. W. Chen, "Slant Transform Image Coding," University of Southern California, Image Processing Institute, USCEE Report 441, May, 1973.

3.2 A Study of Transform Domain Quantization

Clifford Reader

The decorrelating effect of unitary transformations may be exploited for the transmission of pictures if an effective scheme may be found for quantizing the samples in the transform domain of the picture. A study has been made of several techniques, all but one of them adaptive in nature. To implement the quantization schemes the data was first modelled statistically. Much of the work performed has been concerned with fitting the data to the model and analysis of data which is not modelled well by the assumed statistics. Results are presented to show the effectiveness of the schemes examined.

The Quantization Process The quantization process is the following sequence of operations: The determination of a scheme to allocate a number of quantization levels (i.e., a number of bits) to each transform domain sample, the number being in accordance with the expected importance of that sample; the determination of quantization and reconstruction levels having the same statistical distribution as that of the data; the scaling of the data according to its expected distribution in two-dimensional transform domain space, such that the range of values of each samples falls within that of the quantization levels; and finally, the determination of the

correct reconstruction level for each sample.

The expected distribution of the data within the transform domain is obtained from modelling the data as a Markov process. The implication is that energy within the domain is compacted at the lower sequences. Figure 1 shows samples of bit allocation schemes with an average of 1.5 bits/pixel. It is to be noted that many samples are allocated zero bits and are not quantized at all. The total error resulting from the quantization process may thus be considered to have two components: the quantization error from the samples within the quantization zone and the error from discarding those samples outside the zone.

The distribution of the quantization and reconstruction levels is modelled as Gaussian or Exponential. The levels are evaluated according to the optimum scheme of Max [1].

Quantization Process Parameters

Inter-element Correlations The horizontal and vertical inter-element picture correlations are used to estimate the transform domain sample variances. The variances are used twice in the quantization process. First, the number of quantization levels that are allocated to a sample is made proportional to its estimated variance. Second, the scaling of a sample prior to quantization is made proportional to its estimated variance.

The number of bits allocated to a sample ranges from two to eight. The correlations may be used to adjust the spread of bits within the transform domain and also to distribute the bits to allow for any difference in horizontal and vertical correlation within the picture. Figure 1a shows the bit allocation scheme for horizontal and vertical correlations of 0.96, 0.96. Figure 1b shows the effect of changing the vertical correlation to 0.92.

The scaling of the data before quantization is a compression of the dynamic range of the data designed to fit it to the quantization levels.

8	8	8	8	8	8	8	8	7	7	7	7	5	5	5	5
8	8	8	5	5	5	3	3	3	3	3	3	2	2	2	2
8	8	5	3	3	3	2	2	2	2	2	2	0	0	0	0
8	5	3	2	2	2	0	0	0	0	0	0	0	0	0	0
8	5	3	2	2	2	0	0	0	0	0	0	0	0	0	0
8	5	3	2	2	2	0	0	0	0	0	0	0	0	0	0
8	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
8	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
7	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
7	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
7	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
7	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(a) Correlations 0.96, 0.96

8	8	8	8	8	8	8	8	7	7	7	7	6	6	6	6
8	8	8	8	5	5	4	4	3	3	3	3	3	3	3	3
8	7	5	4	3	3	2	2	2	2	2	2	2	2	2	2
8	5	4	2	2	2	0	0	0	0	0	0	0	0	0	0
8	5	3	2	2	2	0	0	0	0	0	0	0	0	0	0
8	5	3	2	2	2	0	0	0	0	0	0	0	0	0	0
6	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
6	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(b) Correlations 0.96, 0.92

FIGURE 3.2-1 Bit Allocation Schemes
1.5 Bits/Pixel

The high sequency samples are boosted in amplitude and the low sequency samples may be attentuated by division by a constant multiplied by the sample standard deviation. The constant is referred to as the amplitude parameter.

Amplitude Parameter This parameter is critical in the adaptation of data to the quantization process. Within the transform domain the sample standard deviations are relied upon to compress or expand the range of the samples to the range of the quantization levels. The amplitude parameter must ensure that the two ranges coincide and has been the subject of intensive study.

Experimental Results The performance of the scheme was examined with a range of values of the amplitude parameter to determine which parameter would yield the result with the minimum mean squared error and also which result was subjectively the best. The picture was processed in blocks of 16 x 16 picture elements using the slant transform exponential quantization level distribution and for two sets of horizontal and vertical correlation: 0.95; 0.93 and 0.86, 0.86. The subjectively best results are shown in Figure 2. The amplitude parameters for these results are 2.4 and 1.2. Tests were also made using different transforms and different pictures. For each case, the optimum amplitude parameter could be found only by repeated experimentation. It was felt that the scheme was of no practical use in this form and that it should be made to adapt itself for differing data and transforms.

First the quantization process was examined in greater detail. Ten of the 16 x 16 element blocks were selected from the picture such that all areas of the picture were represented. These blocks were coded to determine which amplitude parameter yielded the minimum mean squared error for each block. This was found to vary widely over the range two to fifty. In an attempt to reduce this variation, the amplitude parameter

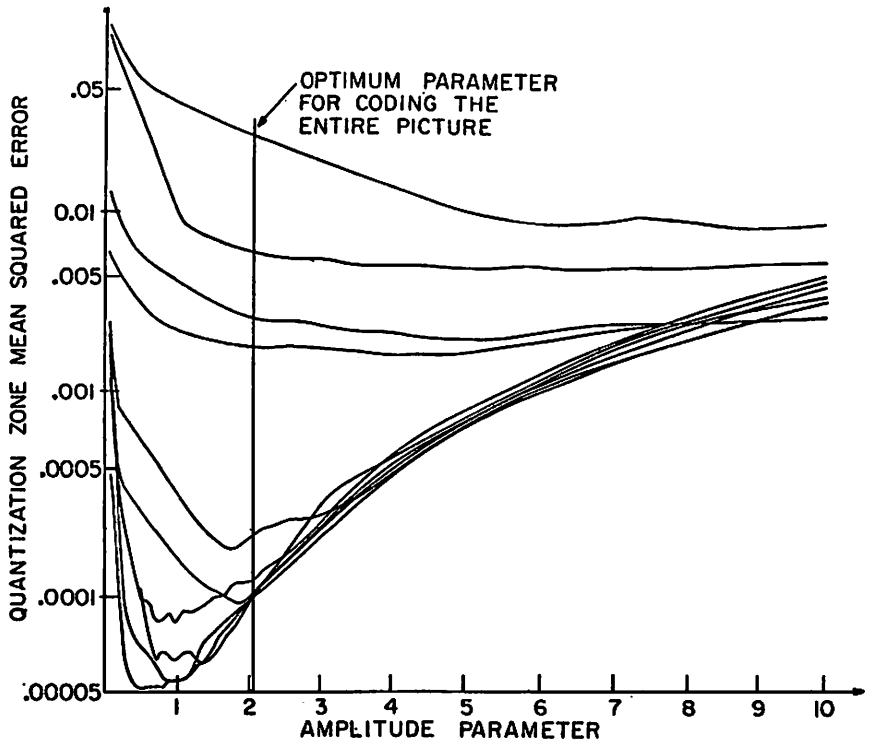
(a) correlations 0.95, 0.93
amplitude parameter 2.4

(b) correlations 0.86, 0.86
amplitude parameter 1.2

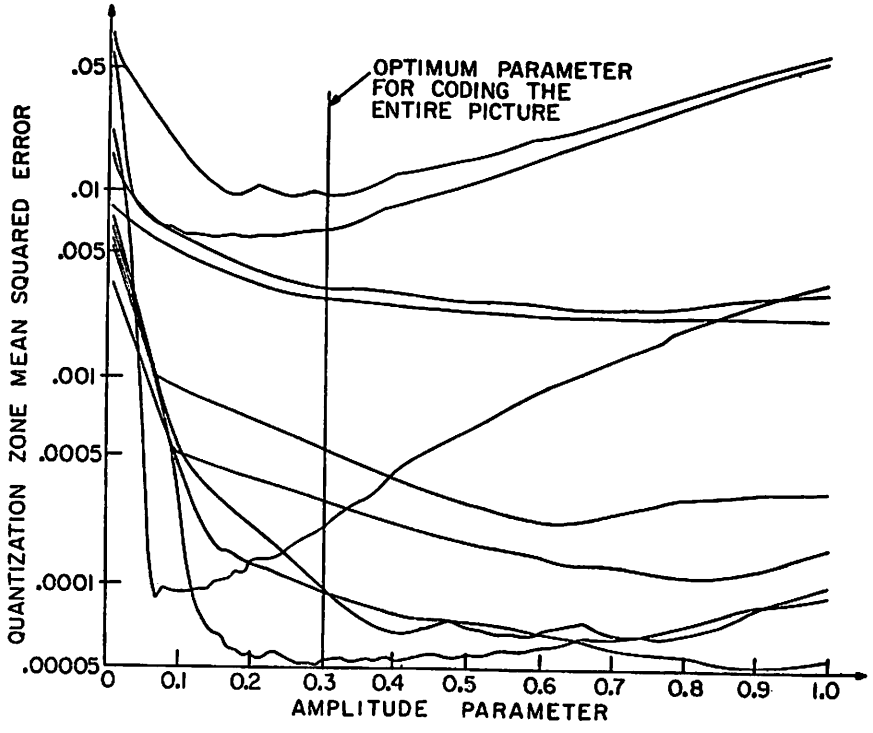
Figure 3.2-2. Non-adaptive coding, 1.5 bits/pixel.

was made dependent upon the magnitude of the largest a.c. sample in the transform domain of the block being quantized (i.e., excluding the zero sequency or d.c. sample). This did not reduce the variation significantly. The amplitude parameter was also made dependent upon the d.c. term, this also was unsuccessful. Finally, the parameter was made dependent upon the a.c. energy in the domain (i.e., the sum of the squares of the values of the a.c. samples). This reduced the range of variation to one to nine. Curves were plotted of the performance of the original schemes and the a.c. energy adaptive scheme (simply called hereafter the a.c. scheme) for the ten blocks examined. These results are shown in Figure 3. Particularly significant is the relative flatness of the a.c. scheme curves--the curves do not reach a minimum at the same point, but do not deviate widely from the minimum over a relatively large range of amplitude parameters. Also to be noted is that those curves in the non-adaptive scheme which do not reach a minimum for low values of the amplitude parameter are those with high mean squares error--the mean squared error scale is logarithmic and a small increment in one of the upper curves represents a large increase in error. Those blocks with large mean squared error are of course those corresponding to high picture detail so it is important that they be quantized optimally. An experiment was performed to determine the optimum amplitude parameter for the a.c. scheme with all other parameters the same as for the non-adaptive scheme experiment. Results are shown in Figure 4.* The most noticeable defect is that occurring on diagonal edges the appearance of the rectangular transform basis vector waveforms. Several blocks were selected containing diagonal edges and the process examined for those blocks in comparison with blocks containing high and low detail information. As expected the transform domains of the high detail blocks contain many high energy high sequency samples while the reverse is true for low detail blocks. For the diagonal edge blocks, there are only a few high

*The amplitude parameters for these results are 0.3 and 0.2.



(a) non-adaptive scheme



(b) adaptive scheme

Figure 3.2-3. Mean squared error curves for ten picture blocks

(a) correlations 0.95, 0.93
amplitude parameter 0.3

(b) correlations 0.86, 0.86
amplitude parameter 0.2.

Figure 3.2-4. Adaptive coding, 1.5 bits/pixel.

energy samples concentrated mainly along the axes leading away from the zero sequency term. The rest of the domain contains a relatively uniform spread of low energy samples. In some cases only the two first sequency terms have high energy and in other cases one axis predominates over the other. Thus the construction of a diagonal edge is seen to be from a few samples detailing the direction of the edge plus many samples defining the precise edge. This distribution is not modelled well by the assumed statistics and furthermore the a.c. scheme tends to err as it detects a quite high a.c. energy within the domain and thus quantizes only the low sequency samples well, assuming the high sequency samples to be of relatively high energy. In fact, the high sequency samples must be treated in the same fashion as those in low detail blocks. In order to understand more about the optimal quantization of the three classes of image blocks--high detail, low detail and diagonal edge--the effect of the bit allocation scheme and the sample standard deviation were studied in greater detail.

For experiments performed up to this point, the estimated correlations, through the transform domain variances, determined both the bit allocation scheme and (together with the amplitude parameter) the scaling constants. This model is logical for data which fit the assumed distributions but may be inappropriate for diagonal edge blocks and some high detail blocks. An experiment was performed in which a picture was coded for fifteen different pairs of correlations. The fifteen resultant bit allocation schemes were tested separately with each of the fifteen resultant data scaling schemes. Analysis so far performed confirms that all low detail image blocks and many high detail image blocks are optimally coded using bit allocations and scaling constants derived from the same correlations and furthermore that (for the picture examined) the horizontal and vertical correlations were the same. Diagonal edge blocks and other high detail blocks required very different correlations. For example, many blocks were optimally coded using scaling constants determined by

horizontal and vertical correlations of 0.96 and 0.92 with bit allocations determined by correlations of 0.84 and 0.80. Figure 5a shows the result of coding a picture with the optimum bit allocations and scaling constants. The a.c. energy scheme was used with an amplitude parameter of 0.3, all other parameters being the same as for previous results.

Finally, experiments were performed to examine the effect of threshold sampling significant samples in the transform domain which were not included in the quantization zone. The a.c. energy scheme was used for quantization within the zone using an amplitude parameter of 0.3, correlations of 0.95, 0.93 and the slant transform. Figure 5b shows the result when a threshold of 0.05 was set. A total of 2140 samples exceed this threshold and were allocated four bits each. Including a minimum of four bits to describe the position of threshold samples, the bit rate for the picture is 1.72 bits per pixel.

Conclusion The effectiveness of various quantization schemes has been examined. Results have indicated that for the purposes of coding, a picture may be considered to be composed of three classes of area. Optimal codings for those areas have been found. Work is continuing to determine a satisfactory implementation. Future work will consist of a comparative study of the threshold sampling scheme and a scheme to match the number of bits allocated to an image block with the detail contained within that block.

Reference

1. Max, J., "Quantizing for Minimum Distortion," IRE Transactions on Information Theory, March 1960, pp. 7-12.

(a) optimized adaptive coding scheme, 1.5 bits/pixel

(b) adaptive coding scheme with thresholding, 1.72 bits/pixel

Figure 3.2-5. Adaptive coding, amplitude parameter 0.3.

3.3 Adaptive Intra- and Inter-Frame Image Coding Utilizing Fourier Techniques

Andrew G. Tescher

In this section preliminary results are given for a research program involving the following tasks associated with the development of Fourier transform image coding techniques: 1) the relative importance of amplitude and phase quantization; 2) modeling of frame-to-frame coding in the transform domain; 3) demonstration of a coding example for the intra-frame case. In many applications, the observation is often made that amplitude errors in the frequency domain are less significant than phase distortions [1-3]. In order to provide an analytic basis for this observation, let the image $g(x,y)$ and its Fourier transform $G(u,v)$ be defined as

$$G(u,v) = \mathcal{F}\{g(x,y)\} = \int_{-\infty}^{\infty}\int g(x,y)\exp\{-2\pi j(ux+vy)\}\,dx\,dy$$

Generally $G(u,v)$ is complex-valued and can be expressed in terms of its real and imaginary parts as well as a complex phasor

$$G(u,v) = G_R(u,v) + jG_I(u,v) = |G(u,v)|\exp\{j\phi(u,v)\}$$

where

$$\phi = \tan^{-1}(G_I/G_R)$$

In the following, the quantization of $|G|$ and ϕ are investigated.

Phase Quantization The phase is assumed to be uniformly distributed in $[-\pi,\pi]$. The optimum quantization strategy will accordingly utilize a uniform quantizer for the range $[-\pi,\pi]$. Then the relative mean square quantization error is found to be

$$Q = \mathcal{E} \left| e^{i\phi} - e^{i\frac{\pi}{n}} \right|^2$$

or

$$Q = 2\left(1 - \frac{n}{\pi} \sin \frac{\pi}{n}\right)$$

where the symbol $\mathcal{E}$ indicates ensemble averaging.

Amplitude Quantization It has been observed, and can be argued on the basis of the Central Limit Theorem, that G_R and G_I should have Gaussian distributions. As a consequence, $|G|$ will be Rayleigh distributed.

In order to minimize the quantization errors, the following procedure has been adopted. The amplitude is mapped into a new random variable such that the transformed values are uniformly distributed, and can, therefore, be quantized uniformly. The quantized values are then mapped back into the original domain utilizing the inverse of the original mapping.

It is desirable to obtain quantitative values for the m.s.e. dependent on the number of quantization levels. Unfortunately, this problem cannot be solved in closed form and, therefore, numerical techniques must be employed. The appropriate m.s.e. values for both amplitude and phase are given in Table 1 for $n = 2^j$ levels where $j = 0, 1, 2, 3, 4, 5$. Note that in all cases the amplitude is less sensitive to quantization errors than the phase. The contrast between the two types quantization errors is most noticeable for coarse quantization. In fact, simply replacing the amplitude by its average value results only in 21.5% error. However, a four level quantizer is required for the phase to maintain the same appropriate level of error.

Frame-to-Frame Coding in the Frequency Domain In this section, consideration is given to the effects in the frequency domain of frame-to-frame image changes. Consider the following case of frame-to-frame change. Let $g_3(x, y)$ represent a sub-block in frame A which is shifted a distance, a, in the horizontal direction during the time that frame B is

TABLE 1

The relative mean-squared error caused by phase and amplitude quantization

n	PHASE	AMPLITUDE
1	2	.215
2	.73	.042
4	.2	.025
8	.05	.011
16	.013	.0048
32	.0031	.0020

generated. Let $g_1(x,y)$ represent the unchanging background. The altered parts of the background are represented by $g_2(x,y)$ and $g_4(x,y)$ as shown in Figure 1. In frame A, $g_2(x,y)$ is part of the frame while $g_4(x,y)$ is covered by $g_3(x,y)$. The roles of $g_2(x,y)$ and $g_4(x,y)$ are interchanged in frame B. Equivalently, one may write

$$g_A = g_1 + g_2 + g_3 \tag{1}$$

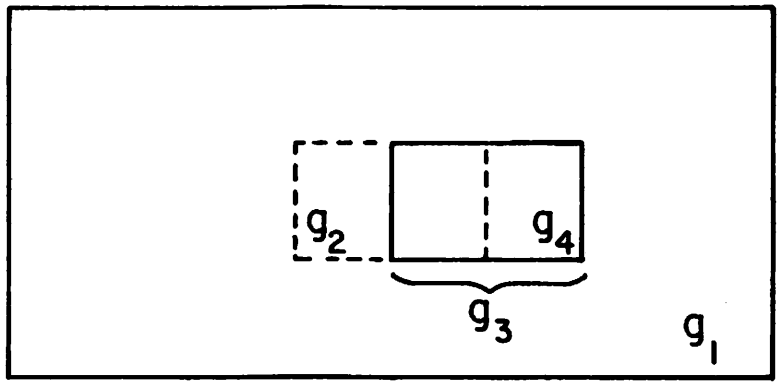
$$g_B = g_1 + g_3(x+a,y) + g_4 \tag{2}$$

Here, g_A and g_B represent frames A and B. Note that the argument (x,y) has been dropped to simplify the notation. Although eqs. (1) and (2) represent a rather simple type of inter-frame image variation, the analysis covers many realistic situations.

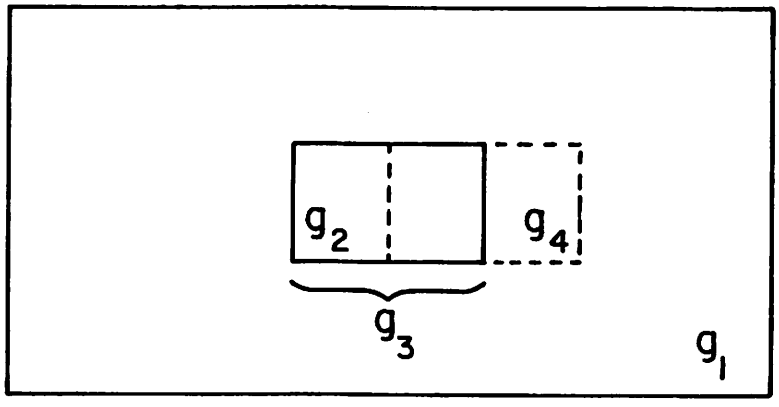
In terms of the previously developed notation, the frame-to-frame change is given in the following simpler form. First the following additional definitions are made.

$$g_a = g_2 + g_3 \tag{3}$$

$$g_b = g_3(x+a) + g_4 \tag{4}$$



(a) frame A



(b) frame B

Figure 3.3-1. Geometry of frame movements.

Equations (5) and (6) clearly indicate that for the model under discussion each frame can be decomposed into a varying part and one that remains unaltered between consecutive frames.

The following conventional definition is used for the Fourier transform

$$G_S(u,v) = G_S = \int_{-\infty}^{\infty} g_S(x,y) \exp\{-2\pi j(ux + vy)\}\,dxdy \tag{7}$$

$$G_S = |G_S| \exp\{j\phi_S\} = G_{SR} + jG_{SI} \tag{8}$$

$$\phi_S = \tan^{-1}[G_{SR}/G_{SI}] \tag{9}$$

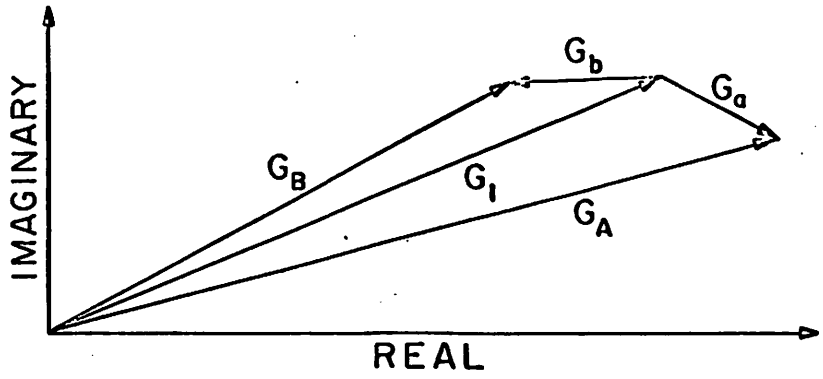
Equations (5) and (6) can directly be transformed yielding equations (10) and (11)

$$G_A = G_1 + G_a \tag{10}$$

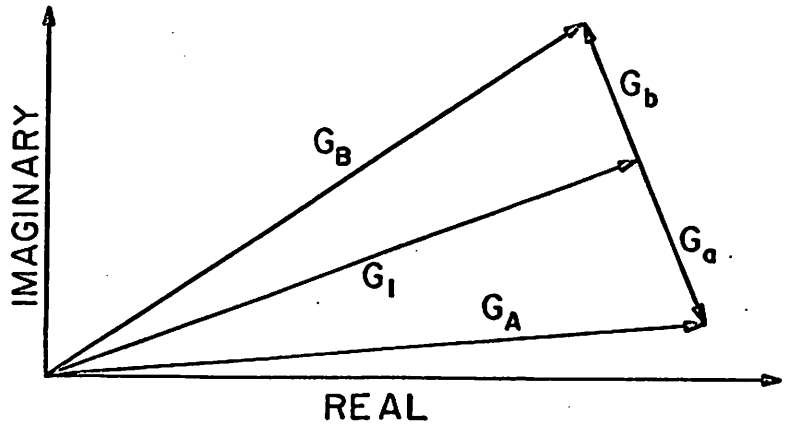
$$G_B = G_1 + G_b \tag{11}$$

In order to show how frame-to-frame variations appear in the frequency domain simultaneous graphical representations of equations (10) and (11) are given in Figure 3. The following assumptions have been made: 1) g_1, g_a and g_b have "similar" Fourier decompositions, and 2) the region over which g_1 is defined is larger than the similarly specified regions for g_2 and g_b. These assumptions imply that g_1, g_a and g_b have approximately the same power spectral density except for different scale factors.

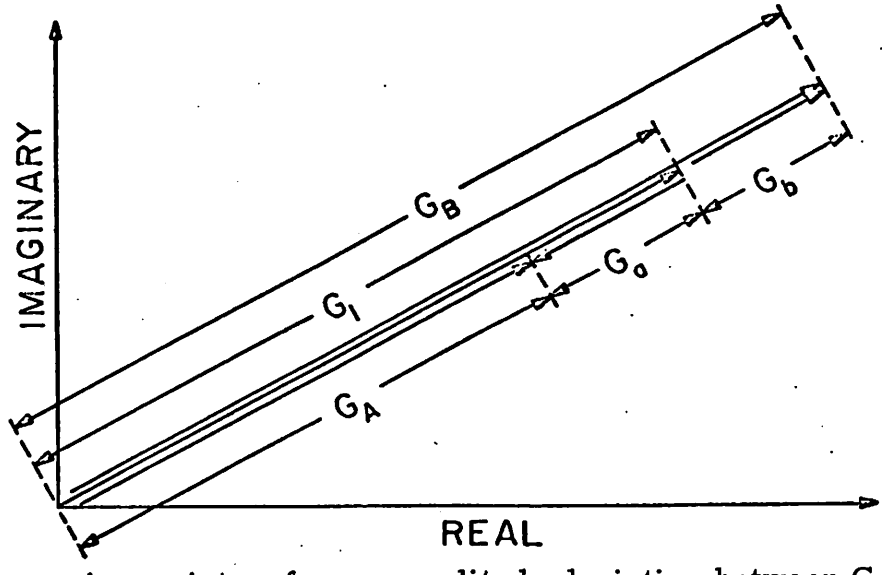
Clearly, small amplitude and phase changes are implied between frames A and B in Figure 2a. Using the graphical representation, the maxima of phase and amplitude can similarly be demonstrated as shown in Figures 2b and 2c. The following inequalities can be obtained



(a) phasor representation of inter-frame changes



(b) maximum inter-frame phase deviation between G_A and G_B



(c) maximum inter-frame amplitude deviation between G_A and G_B

Figure 3.3-2. Vector representation of inter-frame changes.

$$|\Delta\phi| = |\phi_A - \phi_B| \geq \left| \tan^{-1} \frac{G_B}{G_1} \right| + \left| \tan^{-1} \frac{G_a}{G_1} \right| \quad (12)$$

$$|\Delta G| = \Big| |G_A| - |G_B| \Big| \geq \left[|G_a| + |G_b| \right] \quad (13)$$

It is anticipated that the inter-frame coder will update the appropriate amplitude and phase values. The optimization of the updating procedure is greatly helped by the inequalities of equations (12) and (13).

One may substitute reasonable values into equations (12) and (13) to obtain quantitative bounds on the update values. In particular, let $|G_1| = |G_a| = |G_b|$. Allowing for similar power spectral densities for changing and unchanging image segments, this is the case where only one half of the image is unaltered. Yet, the maximum phase change cannot exceed the $\pi/2$ which is still a four-fold reduction on the phase range as compared to single images.

If only a 10% image area change is allowed, then under the same assumptions as before, the maximum phase range may not exceed $\pi/15$ which is a thirty-fold reduction on the phase values. Unlike the uniform distribution of the phase in single images, the phase changes should have a peaked distribution, which in fact should allow additional bandwidth reduction. Utilizing equation (13) similar analysis can be performed for amplitude changes as well.

A Coding Example for the Intra-Frame Case In the previous discussion, the importance of phase over amplitude in the frequency domain has been demonstrated. If the amplitude can be estimated, based on the particular image to be coded, various different adaptive schemes can be developed to minimize the required bandwidth for the image transmission. Here, a particularly simple, yet rather successful, scheme is presented.

The original 256 x 256, 8 bit per picture element "girl," Figure 3a,

(a) original, 8 bits/pixel

(b) coded, 0.75 bits/pixel

Figure 3.3-3. Adaptive coding experiment.

was Fourier transformed. The amplitude and phase values were individually quantized. The number of the quantum levels were fixed in advance. This number was not, however, a constant. Its value was highest in the low-spatial frequency region and decreased for increasing spatial frequencies. In all cases, the phase was quantized to two bits more than the amplitude.

The phase was uniformly quantized and the amplitude quantization utilized the procedure described earlier. The parameter τ in the Rayleigh distribution was determined adaptively utilizing the three previously quantized amplitude values. The receiver, if no channel errors are introduced, can also calculate τ and reconstruct the frequency domain. The final image is obtained by the application of the inverse Fourier transform. The schematic representation of the quantizer is shown in Figure 4. Note this quantizer can also be considered as one with memory. The result of the coding experiment is shown in Figure 3b. The average number of bits per picture element is approximately 0.75, which is better than a ten-fold bandwidth reduction. Additional improvements can be made by making the number as well as the location of the quantum levels adaptive. This technique is currently being implemented for both intra- and inter-frame coders.

References

1. B. L. McGlamery, "Image Restoration Techniques Applied to Astronomical Photography," in Astronomical Use of Television-Type Image Sensors, NASA Report SP-256, pp. 167-192, Princeton University, May 1970.

2. D. Kermisch, "Image Reconstruction from Phase Information Only," Journal of the Optical Society of America, Vol. 60, pp. 15-17.

3. H. C. Andrews, A. G. Tescher, R. P. Kruger, "Image Processing by Digital Computer," IEEE Spectrum, Vol. 9, pp. 20-32, 1972.

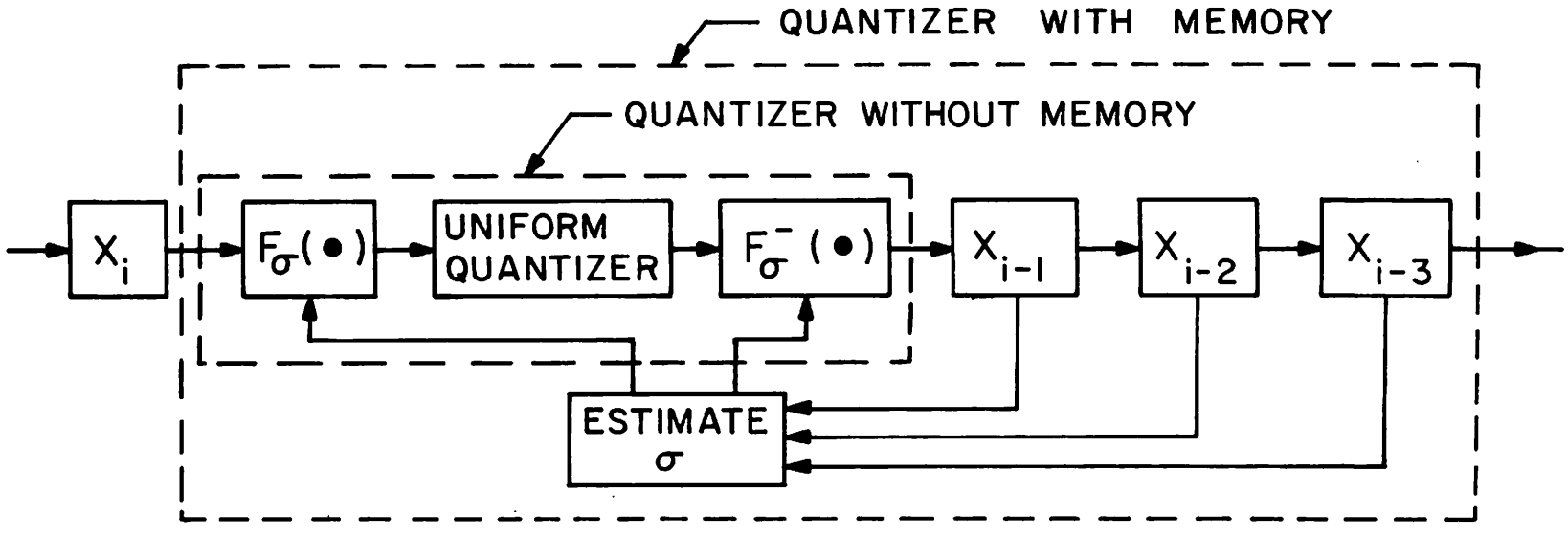


Figure 3.3-4. Schematic representation of amplitude quantization.

3.4 A Cascade of Unitary Transformations and DPCM Systems for Coding Pictorial Data

Ali Habibi

The search for efficient techniques of transmitting pictorial data over digital communication channels has led various researchers to a common approach to the problem. Briefly, this approach entails processing the correlated data (images) to generate a set of uncorrelated, or as nearly uncorrelated as possible, set of signals which in turn are quantized using a memoryless quantizer. The quantized signal is then encoded using either fixed or variable length code words, and is transmitted over a digital channel. This is the general approach taken in designing differential pulse code modulators (DPCM) and the systems that use unitary transformation and block quantization, as well as many other systems developed in recent literature. Both DPCM and transform coding techniques have been used with remarkable success in coding pictorial data. A study of both these systems has indicated that each technique has some attractive characteristics and some limitations. The transform coding systems achieve superior coding performance at lower bit rates; they distribute the coding degradation in a manner less objectionable to a human viewer, show less sensitivity to data statistics (picture-to-picture variation), and are less vulnerable to channel noise. On the other hand, DPCM systems, when designed to take advantage of spatial correlations of the data, achieve a better coding performance at a higher bit rate. The equipment complexity and the delay due to the coding operation is minimal. Perhaps the most desirable characteristic of this system is the ease of design and the speed of the operation that has made it possible for DPCM systems to be used in coding television signals in real time. The limitations of this system are the sensitivity of well-designed DPCM systems to picture statistics and the propagation of the channel error on the transmitted picture.

A hybrid coding system that combines the attractive features of

both transform coding and the DPCM systems is being studied. This system exploits the correlation of the data in the horizontal direction by taking a one dimensional transform of each line of the picture, then it operates on each column of the transformed data using a one-element predictor DPCM system. The unitary transformation involved is a one-dimensional transformation of individual lines of the pictorial data. Thus the equipment complexity and the number of computational operations is considerably less than what is involved in a two-dimensional transformation. Simulated results indicate good coding capabilities of the system The system is particularly attractive in the sense that the principle can be expanded to utilize interframe coding of television signals. Such a coding system would start by taking a two-dimensional transformation of each frame of the television signal then it would code the transformed signal in the temporal direction by a number of parallel DPCM encoders thus exploiting the correlation of data in spatial as well as temporal directions.

DPCM Coding of Transformed Data In the system proposed here the pictorial data is scanned to form N lines then each line is sampled at a Nyquist rate. This sampled image is then divided into arrays of M by N picture elements $u(x, y)$ where x and y index the rows and the columns in each individual array such that the number of samples in a line of image is an integer multiple of M. One dimensional unitary transformation of the data and its inverse are modeled by the set of equations

$$u_i(y) = \sum_{x=1}^{M} u(x, y)\, \varphi_i(x) \qquad \begin{matrix} i=1,2,\ldots,M \\ y=1,2,\ldots,N \end{matrix} \qquad (1)$$

$$u(x, y) = \sum_{i=1}^{M} u_i(y)\, \varphi_i(x) \qquad (2)$$

where $\varphi_i(x)$ is a set of M orthonormal basis vectors. The correlation of the transformed samples $u_i(y)$ and $u_i(y+\tau)$ is given by

$$C_i(\tau) = \sum_{x=1}^{M}\sum_{\hat{x}=1}^{M} R(x,\hat{x},y,y+\tau)\,\varphi_i(x)\,\varphi_j(\hat{x}) \qquad (3)$$

where $R(x,\hat{x},y,\hat{y})$ is the spatial autocovariance of the data.

Note that this equation indicates that the correlation of samples in each column of the transformed array is directly proportional to the correlation of sampled image in vertical direction, and also that the correlation of samples in various columns of the transformed array is different. Thus, a number of different DPCM systems should be used to encode each column of the transformed data. The block diagram of the proposed system is shown in Fig. 1. A replica of the original image $u^*(x,y)$ is formed by inverse transforming the coded samples, i.e.,

$$u^*(x,y) = \sum_{i=1}^{n} v_i(y)\,\varphi_i(x) \qquad n \leq M \qquad (4)$$

The mean square value of coding error, assuming that the quantization noise in the ith DPCM systems is uncorrelated with $u_i(y)$, is given by [1]

$$\epsilon^2 = \frac{1}{N}\sum_{x=1}^{n} K(m_i)e_i^2 + R(0,0,0,0) - \sum_{i=1}^{n} C_i(0) \qquad (5)$$

where e_i^2 is the variance of the differential signal in ith DPCM system and $K(m_i)$ is the quantization error of a variate with a unity variance. From published results [2], [3]

$$e_i^2 = C_i(0) - C_i^2(1) \qquad (6)$$

and $K(m_i)$ is approximated by an exponential function

$$K(m_i) = b\exp\{-am_i\} \qquad (7)$$

for constant values for a, b. Substituting Eq. (7) into Eq. (5) and

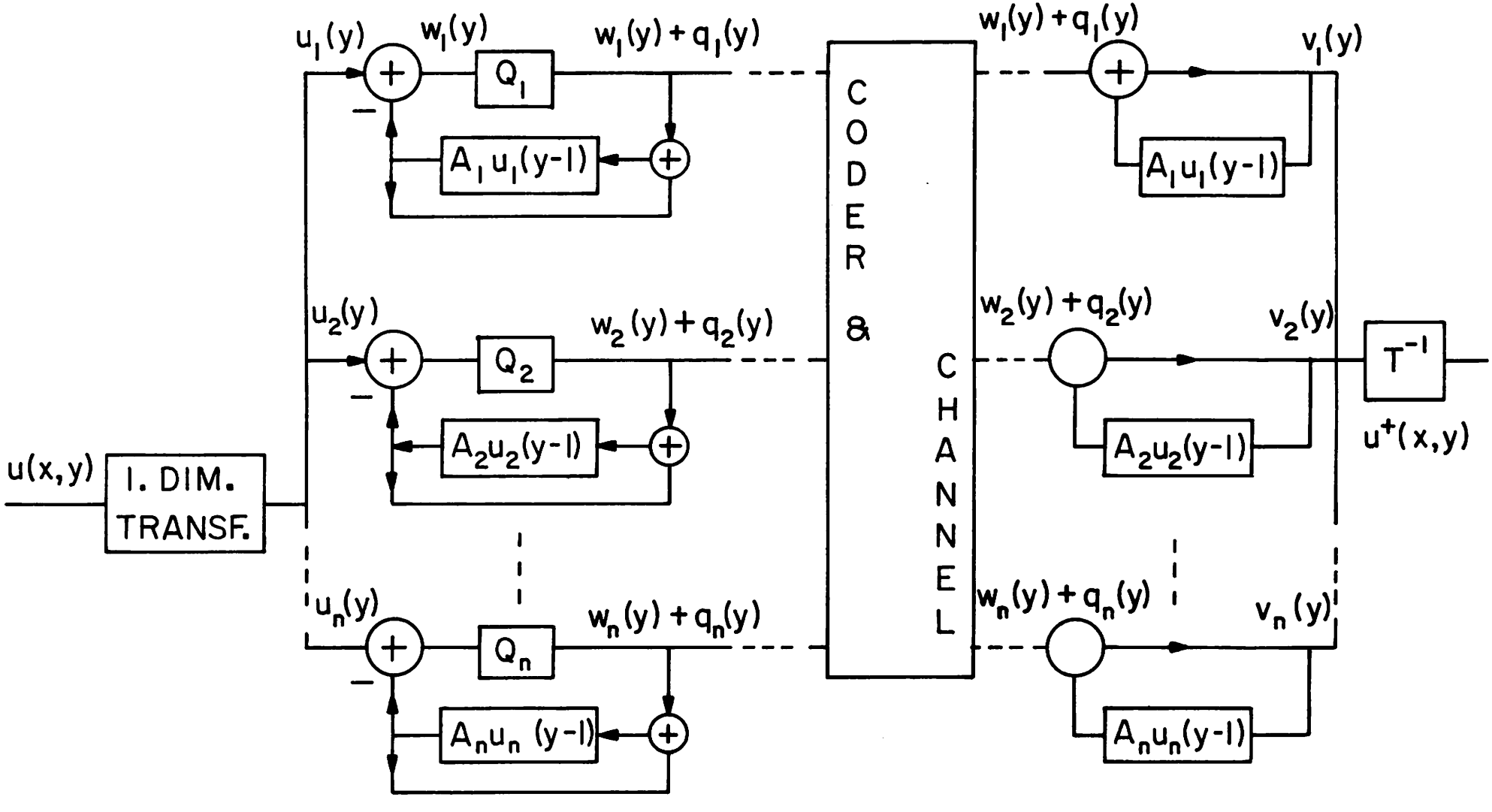


Figure 3.4-1. Block diagram of the coding system.

minimizing ϵ^2 with a constraint $M_b = \sum_{i=1}^{n} m_i$ the bit assignment is given by

$$m_i = \frac{M_b}{n} + \frac{1}{a}\left[\ln e_i^2 - \frac{1}{n}\sum_{i=1}^{n} \ln e_i^2\right] \qquad (8)$$

Figure 2 shows the coding error versus the bit rate for a third order Markov field with

$$R(x, x, y, y) = \exp\{-\alpha\,|x-\hat{x}\,| - \beta\,|y-\hat{y}\,|\} \qquad (9)$$

using Karhunen-Loeve, Hadamard, and Fourier transformations and DPCM systems with a one-element predictor.

The system proposed here has been simulated on a digital computer with a block size N = 256 and M = 16. The original and coded pictures are shown on Figure 3 using a Hadamard transformation. Preliminary results show that the system achieves good coding capabilities and is worthy of further studies.

References

1. A. Habibi, "A Cascade of Unitary Transformations and DPCM Systems for Coding Pictorial Data," Proceedings of Applications of Walsh Functions, April, 1973.

2. A. Habibi, "Quantization Error and Entropy of a Single Random Variate," USC Semiannual Technical Report, USCEE #425, 1 March 1972 - 31 August 1972, pp. 27-38.

3. A. Habibi, "Comparison of nth Order DPCM Encoder with Linear Transformations and Block Quantization Techniques," IEEE Trans. on Comm. Tech.; Vol. COM-19, No. 6, December, 1971, pp. 948-956.

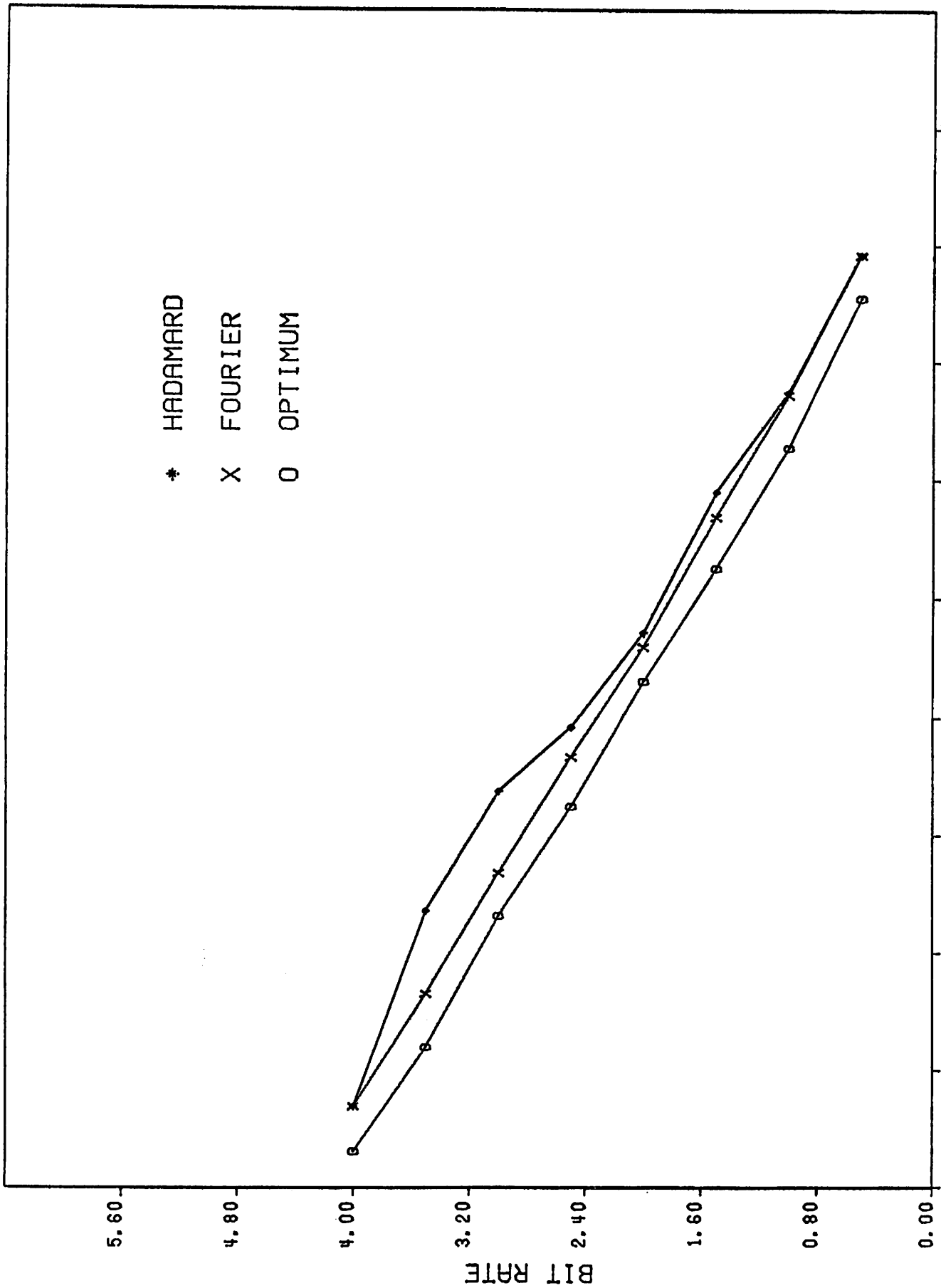


Figure 3.4-2. Performance of the cascaded system for a Markov field.

(a) original (b) 1 bit/pixel

(c) 2 bits/pixel (d) 3 bits/pixel

Figure 3.4-3. Original and encoded pictures using 1 dimensional Hadamard transformation and DPCM systems.

3.5 Coding Television by Contour Tracing the Interframe Differential Signals

Ali Habibi

The search for efficient ways of transmitting television pictures over digital communication circuits has been along two general approaches. One approach is to use the spatial correlation in a single frame of a picture to reduce the number of binary digits essential to transmit its information content within some level of degradation. This approach has led to techniques that manage to remove all or most of the redundancy in the spatial domain. The other approach is to use the correlation of picture elements in the successive frames of television signals and use this factor in reducing the bit rate essential to transmit the new information in each frame. This method for moving pictures performs better than the other techniques, but it has the disadvantage of ignoring the intraframe correlation.

A system for taking advantage of both interframe and intraframe correlations, thus resulting in a further reduction in the bit rate than is possible using either one of the schemes, entails application of a contour tracing algorithm to the differential picture created by taking the frame-to-frame differences. Besides a further reduction in the bit rate, this technique will produce better encoded pictures than either the contour tracing or the conditional frame replenishment techniques since it eliminates the granular noise on the background without effecting the quality of the encoded picture significantly.

Description of the Encoding System The concept of tracing contours of constant gray levels in a single frame is attactive since in a given digital picture many of the points in one region have the same gray level. Thus if a contour could be traced around these points, which are all the same gray level, then only the addressing information that would enable the receiver to trace a similar contour along with the common level of elements in the contour is needed in the receiver. Naturally the scheme

is more useful where the possible number of gray levels is small or there is a high probability of having a large number of points at the same gray level. Wilkins and Wintz introduced a two-dimensional contour tracing algorithm that locates and traces contours enclosing a maximum number of points of the same gray level [1,2]. The algorithm consists of two subalgorithms, one for locating the initial point of a new contour (IP algorithm) and the other for tracing the contours after they are located (T algorithm). The T algorithm traces the outer boundary of the largest connected set of elements having the same value as the initial point, and always terminates back at the initial point. The direction of travel on the T contour can be limited to 4 or 8 spatial directions to limit the directional information to 2 or 3 bits respectively. All elements enclosed by the contour and having the same value as the contour are neglected, but can be reconstructed at the receiver. The authors also developed an algorithm for reconstructing the original data from the system output, and coded a number of still pictures using the contour tracing algorithm. If some contours, say contours that consist of only single points, are deleted and are replaced by the gray level of the neighboring contours, the algorithm will result in some degradation of the encoded data. But at the same time this reduces the number of binary digits essential to reconstruct the data at the receiver. In transmitting the addressing information various types of coding could be used for a further reduction in the bit rate.

In a television signal only a small percentage of picture points change in successive frames. Thus a typical frame differential picture will consist of a large gray area (after a shift in gray scale to eliminate the negative components), some bright and dark spots along the moving edges, and also a number of scattered points in the background that are caused by granular noise. A typical interframe differential picture is shown in Figures 1a and 1b. The bright and the dark spots in an interframe differential picture are concentrated around the moving edges, thus the contour tracing algorithm is ideal for encoding the interframe

differential picture. The granular noise in the background is eliminated simply by ignoring very short contours. This reduces the bit rate without degrading the encoded pictures significantly. Further reduction in the bit rate can be realized by using some coding schemes such as the Huffman code to encode the addressing information.

Implementation of the System and Experimental Results The block diagram implementation of the proposed scheme is shown in Figure 2. The input signal is a television signal that is sampled at the Nyquist rate. This signal is delayed one frame and is subtracted from the signal that corresponds to the succeeding frame. The quantizer Q digitizes the interframe difference signal before it is processed by the contour tracer. The quantizer operates on a differential signal, in analogy to a DPCM system. A 4-bit nonuniform quantizer will produce 8-bit PCM signal quality. This is also the optimum number of possible levels for the contour tracer. The interframe differential picture is reconstructed in the receiver by the reconstructing algorithm. The receiver also employs an analog frame memory to reconstruct the original picture. In the system proposed here the differential signal is obtained by simply subtracting the incoming signal from the signal that corresponds to the preceding frame In analogy to the DPCM system the frame memory could be replaced by a predictor which would make a prediction of the present frame using one or more previous frames.

The system shown in Figure 2 has been simulated on a digital computer and two frames of typical television signals separated by 1/30th of a second have been processed. The original of the second frame along with its reconstructed forms are shown in Figure 1c and 1d. Table 1 summarizes the performance of the system proposed here. The bit rate is obtained without using any coding technique on the transmitted values. A 50% reduction in the bit rate when the transmitted values are encoded optimally is expected. This subject is under investigation. Figure 1d shows one frame of the encoded signal. To make a meaningful evaluation

(a) a typical interframe differential picture

(b) differential picture quantized to 8 levels

(c) original of frame #2

(d) encoded picture using an 8-level quantized differential picture

Figure 3.5-1. Original, interframe and encoded signals.

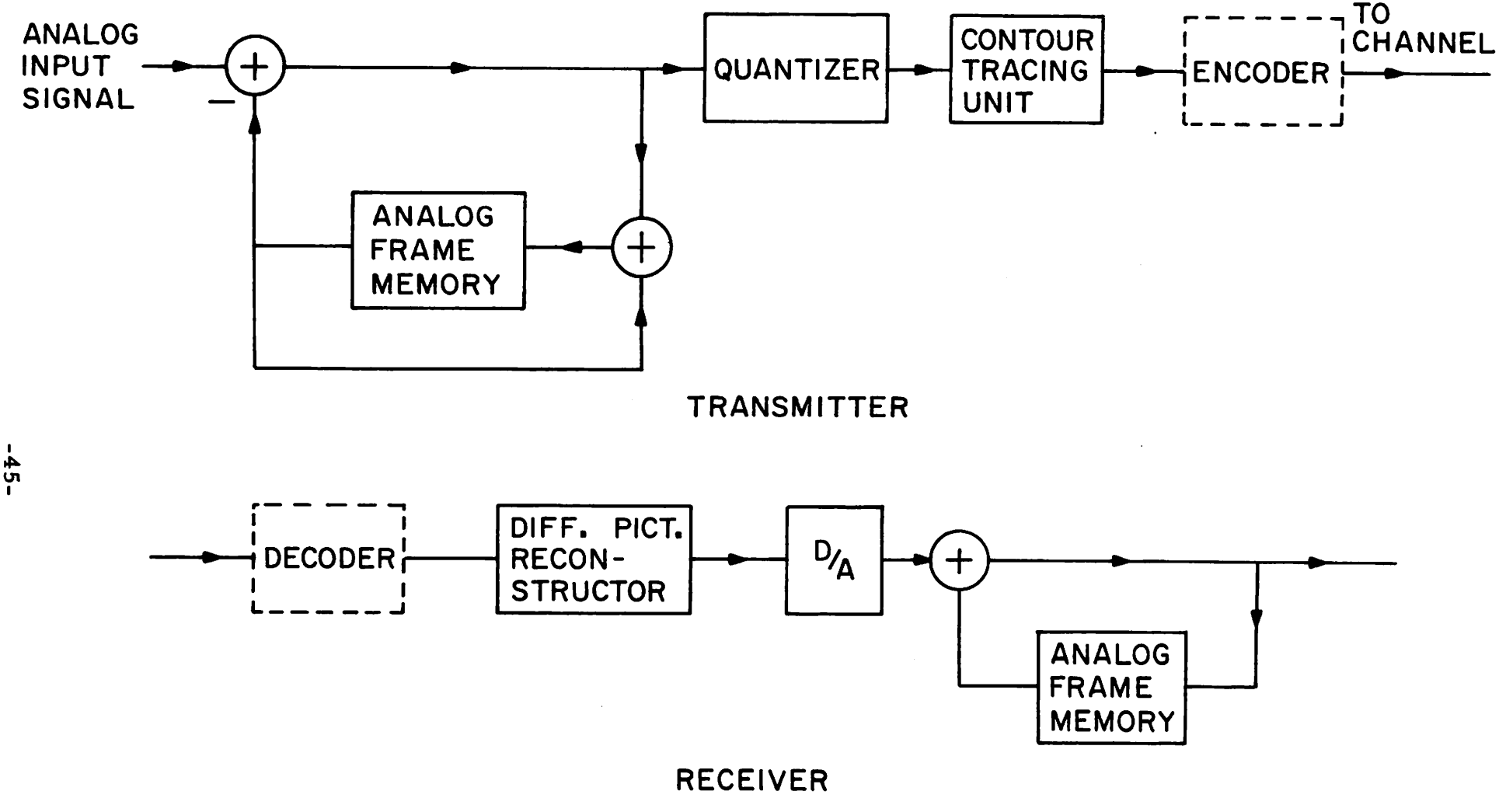


Figure 3.5-2. Block diagram of the transmitter and the receiver.

TABLE 1

Bit rate, number of contours and the CPU time on 360/155 for encoding interframe pictures for various numbers of quantization levels

Number of Quantization Levels	Number of Contours	Number of Bits/Pixel	CPU Time for Tracing All Contours	CPU Time for Reconstruction of Picture
256	18403	5.99	57 sec.	34 sec.
8	5883	1.97	40 sec.	21 sec.
4	1309	0.71	22 sec.	11 sec.

of the system the coding effects should be observed on a number of sequential frames that would indicate a typical and complete motion.

References

1. Wilkins, L. C. and Wintz, P. A., "A Contour Tracing Algorithm for Data Compression for Two-Dimensional Data," TR-EE69-3, School of Electrical Engineering, Purdue University, Lafayette, Indiana. Also International Symposium on Information Theory, Ellenville, N.Y., January 1969.

2. Wilkins, L. C. and Wintz, P. A., "Studies on Data Compression," TR-EE70-17, School of Electrical Engineering, Purdue University, Lafayette, Indiana.

3.6 Contour Tracing in Facsimile Processing

Lloyd R. Welch

Pictures with a two level gray scale (black and white) can be specified by stating the level at each point. However, such a method requires a number of bits of information equal to the number of picture elements. An alternative is to specify pairs of adjacent points with

opposite levels. Run length coding is one such technique which codes the distance between successive horizontal changes. Another alternative is contour tracing described below. For some classes of pictures this promises to require a considerably smaller amount of information than run length coding.

Contour Tracing For the purposes of this discussion a picture is a function $f(x, y)$ which takes on the values 0 and 1 in the rectangle $0 \leq x \leq N$, $0 \leq y \leq M$. The black region, B, is the set of (x, y) with $f(x, y) = 1$. The discrete picture corresponding to f is an array of colored unit squares, the square $S(i, j)$ is the set of points (x, y) with $i \leq x \leq x+1$, $y \leq y \leq j+1$ and the color is black if $B \cap S(i, j)$ has positive area. The directed edges between squares are indicated by the self explanatory notation (i, j, Z) where Z = Up, Down, Left, Right. Note that N+1 will be a permissible value for i and j, and it will be necessary to assume the exterior of the picture is white. A boundary edge is an edge between a black square and a white square. A direction is attached to such an edge so that the black square is on the left of the directed edge.

The set of directed edges along with their end points form a directed graph. A simple analysis shows that at each vertex the number of edges directed toward the vertex equals the number of edges directed away from the vertex. A component of such a graph has a closed, complete directed path, that is, all the edges of the component can be arranged in a sequence with the terminal point of one coinciding with the initial of the next and the terminal point of the last coinciding with the initial point of the first. Such a sequence will be called a contour. The simpler contours trace the boundary of black regions counter clockwise and the boundary of white regions clockwise.

Enumeration of the consecutive edges of a contour requires at most $\ln 3$ bits per edge. For example, consider the edge (i, j, U). The next edge must be of the form $(i, j+1, Z)$. Furthermore, Z cannot be D.

Therefore, there are only three possibilities. Similar analysis is valid for boundary edges of the form (i,j,D), (i,j,R) or (i,j,L). Thus one can envision the following method for describing boundary edges: For each contour describe an initial edge (info = ln (4NM)), then describe the length, L, of the contour (info = entropy of length distribution) and finally describe the succession of edges (info = L ln 3). If we can assume there are relatively few contours, then the dominant part of the total information is ln 3 times the total number, E, of the boundary edges. This can be compared with run length coding as follows: Let H be the entropy at the distribution of distance between consecutive vertical boundaries (horizontal transitions), and assume that $\frac{1}{2}$ E of the boundary edges are vertical. Then run length coding requires $\frac{1}{2}$ HE bits of information while contour tracing requires Cln (4NM) + Eln 3 bits. If $C << E$ and $H < 2 \ln 3$ the contour tracing method is superior.

Potential Uses The contour tracing method is useful whenever the number of edges per component is large. If the picture is typewritten text, this condition will not be met. However, in handwritten material the connection of consecutive letters reduces the number of components and contour tracing may be of use. In addition, handwriting offers further possibilities for information reduction. The black regions are narrow-width curves and the edge tracings on the two sides are considerably similar except for direction along the path. This may allow nearly a factor of two reduction.

Further Research The potential use of contour tracing of hand-written material suggests further investigation. There are several aspects not treated in this report. First, there is the discretizing process for handwritten material. How small must the pixel squares be relative to the line width for reasonable representation? What should the decision criterion be for labeling a square black or white? Second, there is the area of component and edge statistics. Statistical regularities may further reduce the ln 3 estimate. Third, there is the possibility of using

the similarity of the two sides of a line for information reduction. Finally, there is the problem of designing simple mechanisms for encoding and decoding contour tracing information to produce black and white pictures.

3.7 An Application of Universal Coding to Video Data

Lee D. Davisson

In the last report a formulation of the universal coding problem was presented. If a time and state discrete source is characterized probablistically up to an unknown parameter, θ, or arbitrary dimensionality, a universal code is one which results in the minimum possible coding redundancy (in an appropriate sense) over all possible values of θ, with the redundancy approaching zero for large enough code block sizes. This report contains a discussion of: (1) a simple model for video data; (2) universal coding considerations for this model; and (3) some empirical results on some sample video data.

A reasonably good approximation to the probability mass function of the line sample-to-sample quantized difference, $x = 0, \pm 1, \pm 2, \ldots$ is

$$p(x|\theta) = \frac{1 - \theta}{1 + \theta} \theta^{|x|}$$

i.e., double-exponential with parameter θ, unknown and, in most cases, slowly varying depending upon picture activity. If one makes the further simplifying assumption that the differences are statistically independent, which has been found to be approximately so experimentally, then the probability mass function of a block of N differences $\underline{x}_N = (x_1, \ldots, x_N)$ is

$$p(\underline{x}_N|\theta) = \left[\frac{1 - \theta}{1 + \theta}\right]^N [\theta]^{\sum_{i=1}^{N} |x_i|}$$

If θ were known, an optimal code could be designed with arbitrarily small code redundancy. A block code has been found which is universal in the previously defined sense. Essentially, the idea is to transmit the sufficient statistic,

$$\sum_{i=1}^{N} |x_i|$$

for θ using a vanishingly small portion of the coded representation as $N \to \infty$. The additional information needed to reconstruct $\underline{x}_N$ can be coded optimally because the probability of $\underline{x}_N$ given the sufficient statistic is independent of θ.

Another method of universal coding results by redesigning several subcoders for representative values of θ. Call these values θ_i, $i=1,2,\ldots,M$. Given a block of length N, the subcoder with the shortest coded representation is used as output to the channel with an additional log M bits to distinguish the subcoder. If one permits $M \to \infty$ as $N \to \infty$ in such a way that $\log (M/N) \to 0$, so that a dense set of values θ_i on $[0,1]$ results, the redundancy goes to zero for any actual fixed θ.

Both universal coding techniques were applied to a 4096 pixel/line and 2400 line picture. For the second method M=5 codes were selected. One subcoder was a run length coder for values of θ near zero, another subcoder was a straight PCM coder for values of θ near one, and the three other subcoders were Huffman coders for intermediate values of θ. The results are summarized below.

Average Picture Entropy:	3.2 bits/pixel
First Universal Code at N = 64:	2.95 bits/pixel
Second Universal Code at N = 64:	2.98 bits/pixel

The value of N = 64 was the optimum value for both universal codes, but variations up and down by a factor of two made little difference. If N gets smaller than about 16, though, the "overhead" information reduces performance, and if N gets larger than about 256, the variations in θ

reduces performance. The most important point to notice is that universal coding produces a 10% reduction below the average entropy, the minimum achievable by any code of the usual type.

4. Image Enhancement and Restoration Projects

Image enhancement techniques have two major purposes: improvement in the visual quality of a picture to a human viewer; and manipulation of a picture for more efficient processing and data extraction by a machine. Image restoration techniques seek to reconstruct or recreate an image to the form it would have had if it had not been degraded by some physical imaging system. Both techniques are subjects of continuing study; results of this effort during the past six months are summarized below.

The first research task deals with practical methods for the correction of image sensor intensity nonlinearities for real time television systems. In many applications the brightness of an image will exhibit unwanted spatial variations resulting from electronic or optical deficiencies of the sensor. It is possible to perform a posteriori correction of images from such sensors to linearize their intensity response. Such processing not only improves the visual appearance of images, but assists in the detectability of objects, and permits quantitative radiometric measurements to be made from the images.

The next two reports discuss two approaches to image restoration in which a physical realizability constraint is placed upon the image reconstruction. Light reflected or emitted from physical objects is obviously a positive quantity, and it is limited in value to some maximum level. However, most restoration algorithms do not take advantage of this physical fact, and as a result, the reconstructions contain non-physical artifacts. These two reports suggest two techniques for positive image restoration. One method is based upon techniques of constrained mathematical programming, and the other involves a constrained deconvolution process. The research has led to algorithms which will theoretically provide an optimum restoration. Work is continuing to develop practical means of implementing the algorithms for large size images.

The three reports that follow are all directed toward a similar aspect of the image restoration problem: the development of recursive

image restoration algorithms to avoid the computational problems associated with the ordinary methods of implementation on digital computers. The first report is concerned with a multi-dimensional modelling approach to imagery, which, in many applications, permits the design of simple recursive restoration filters. In the next two sections a discussion is given on the problems of non-stationarity and nonlinearity on the design of recursive image processors.

In the next report a theoretical description is given of physical imaging systems which introduce image motion blur. From the physical models developed it is possible to design restoration processors that avoid many of the problems typically associated with spatially variant imaging systems. An example of restoration of photographs subjected to rotational blur is given in the last report.

4.1 Correction of Image Intensity Nonlinearities in Real Time

Alexander A. Sawchuk

In general, all imaging systems respond to incident illumination in a nonlinear fashion and are said to have photometric or radiometric distortion. Denoting a point (u_1, u_2) in a two-dimensional coordinate system by the vector $\underline{u}$, the imaging system has intensity nonlinearities at some particular point $\underline{u}_o$ if the output at $\underline{u}_o$ is not a linear function of the incident light at $\underline{u}_o$. To add another element of difficulty to the analysis, the relation between input and output may change its functional form with the position of the point $\underline{u}_o$. This overall intensity nonlinearity in imaging systems may be variations in light transmission across the image plane due to the properties of the lenses and optical system components preceding the sensor (vignetting, for example) or nonlinearities in the sensor itself due to the scanning electronics and variations in image tube response. The objective of the correction techniques outlined here is to make the overall system produce an output which is a linear function of the incident

light and does not vary over the field of view. The intent is to perform the correction in real time and to present a number of different possible correction concepts. Imaging systems with memoryless point nonlinearities and lack of geometrical distortion of blurring are considered.

Representing the input intensity by $f(u_1, u_2) = f(\underline{u})$ and the output intensity $g(u_1, u_2) = g(\underline{u})$ in the nonlinear system, the system description is

$$g(u_1, u_2) = R\{f(u_1, u_2), u_1, u_2\} = R_{u_1, u_2}(f(u_1, u_2)) \tag{1}$$

where the arguments of R indicate that it may be a nonlinear function of $f(\underline{u})$ and that the nonlinear function itself may change with position (u_1, u_2) over the field.

Perfect intensity correction is possible if the ideal image $f(u_1, u_2)$ can be obtained from knowledge of $g(u_1, u_2)$ at any point (u_1, u_2). This operation of producing the ideal image estimate $\hat{f}(u_1, u_2)$ is written

$$\hat{f}(u_1, u_2) = R^{-1}\{g(u_1, u_2), u_1, u_2\} = R^{-1}_{u_1, u_2}\{g(u_1, u_2\} \tag{2}$$

where R^{-1} is the inverse intensity distortion operation. Perfect correction is possible theoretically, but is a difficult problem if the correction must be done in real time.

For the intensity nonlinearity R_{u_1, u_2} at some particular point $\underline{u}$, R may be expanded as a Taylor series in $f(\underline{u})$ and like powers of $f(\underline{u})$ collected to obtain

$$g(\underline{u}) = a_o(\underline{u}) + a_1(\underline{u}) f(\underline{u}) + a_2(\underline{u}) f^2(\underline{u}) + \cdots \tag{3}$$

where the a_i, $i=1,2,\ldots\ldots\infty$, are coefficients which depend on position. One possible approximation is to simply truncate Eq. (3) after a finite number of terms. By a similar procedure, a Taylor series may also be

written for the inverse distortion of Eq. (2) in the form

$$\hat{f}(\underline{u}) = d_o(\underline{u}) + d_1(\underline{u})\, g(\underline{u}) + d_2(\underline{u})\, g^2(\underline{u}) + \cdots \tag{4}$$

where the d_i, $i=1,2,\ldots,\infty$, be directly obtained from the a_i.

Correction Systems A number of different possible intensity correction methods will be compared in terms of computational requirements, system complexity, ease of calibration, and other factors. It is found that there are tradeoffs between systems with large storage, greater computation, and relatively easy calibration. The correction system is usually placed between scanner and display, and the overall system may include other corrections for geometrical distortion, magnification, or frequency response. The final link in imaging systems is often the human observer and care must be taken not to overdesign the fixed nonlinearity correction system past the observers' ability to use the corrected information. The display or measurement input is a quantized level defined on a discrete matrix of points, and correction computations are assumed to be performed under digital computer control with additional digital or analog special-purpose hardware as may be required for the best implementation. A number of systems are proposed, although they represent only system concepts. The best configuration may be a combination of elements from many designs.

Table Lookup The most general correction method is that used for Ranger and Mariner correction--a table lookup on nonlinear intensity mappings for each pixel in the field of view [2], [3]. Denoting the number of samples in the u_1 coordinate by $\#\{u_1\} = n$, and assuming a square image so that $\#\{u_2\} = n$, the total number of pixels is n^2, or 2^{18} for $n = 512 = 2^9$. Assuming that the corrections for k input intensity levels are stored for each pixel and that each output level may take on one of 2^m values, n^2k words must be stored, requiring n^2km total bits of storage for this system. For a typical system with $k = 64$ input levels

and $2^6 = 64$ output levels, $m = 6$, with $2^{18} \cdot 2^6 = 2^{24}$ required words and $2^{18} \cdot 2^6 \cdot 6 = 6 \cdot 2^{24}$ total bits to be stored. The calibration procedure for this system is extremely tedious and unnecessary for a human observer. If very accurate calibrated measurements are to be made on a picture in real time this technique gives the best results. In a simplified version of table lookup in which the same nonlinear correction is applied to each pixel, only one word for each of the k levels is needed, and km total bits in the system are required.

Fixed Bias and Contrast If the intensity distortion of each point does not vary over the field and the coefficients a_i and d_i in Eqs. (3) and (4) are constants, a two term approximation of the sum in Eq. (3) is probably adequate. To the human observer, a variation in response over the field as described by non-constant a_i's and d_i's is much more noticeable than intensity nonlinearities at each point described by taking more terms in the series. Taking a two term approximation of Eq. (3) for the constant coefficient case and solving for $\underline{f}(\underline{u})$ gives

$$\hat{f}(\underline{u}) = \frac{g(\underline{u})}{a_1} - \frac{a_o}{a_1} \tag{5}$$

as the estimator of the corrected intensity $f(\underline{u})$. This correction can be accomplished in real time using either digital or analog hardware, and only two coefficients need be stored.

Varying Coefficients--Two Terms--Computed Assuming a two term approximation as in the previous section, but allowing the coefficients to vary with position $\underline{u}$, the correction method of Eq. (5) can be used except that $a_o(\underline{u})$ and $a_1(\underline{u})$ must be available as a function of $\underline{u}$ over the field of view. Figure 1 shows a possible implementation where two coefficients are computed analytically from the scanning signals in real time by a digital, analog, or hybrid system. This system is well suited to real time computation because of its low storage requirements and easy calibration [1]. The $a_o(\underline{u})$ and $a_1(\underline{u})$ may be approximated by constant

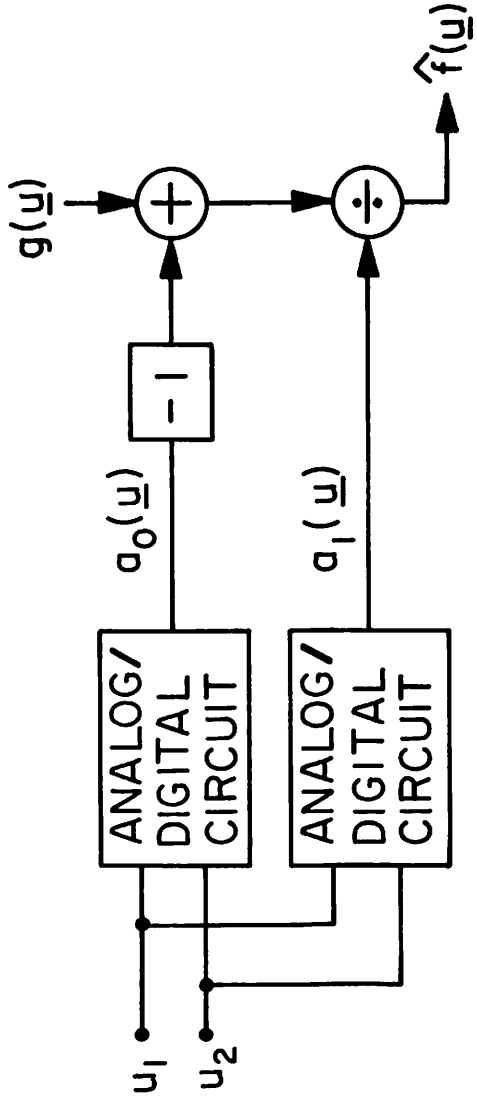


Figure 4.1-1. Two-term varying correction-computed.

coefficient polynomials in u_1 and u_2 of degree G.

For applications involving a human observer, a second degree (G = 2) polynomial approximation is sufficient, while for measurement purposes higher degree polynomials may be required. Only the polynomial coefficients need be stored and in many cases, even better approximations with a limited number of coefficients are possible by taking into account the usual peaked response in the center of the field [1].

Varying Coefficients--Two Terms--Stored As an alternative to computing the corrections described above, the corrections may be stored in a special purpose memory system and applied to each point in a two-term correction shown in Figure 2. The memory can be implemented by programmable read-only memory (PROM) hardware, which can store as many as 2048 bits per chip. For $\#\{u_1\} = \#\{u_2\} = n$, n^2 memory words must be stored for a $a_o(\underline{u})$ and $a_1(\underline{u})$, or 2n words in all. For 2^m different possible correction levels, there are $2mn^2$ required bits to be stored. For $n = 2^9$ and an upper limit of $m = 6$, storage is required for $3 \cdot 2^{20}$ bits. For presently available memories with $2048 = 2^{11}$ bits, a total of $3 \cdot 2^9 = 1536$ memories is needed for full storage. Further reduction to as few as 6 PROM's is possible when the two-variable coefficients can be approximated by a separable sum of polynomials [1].

Storing Polynomial Terms Instead of computing the polynomial terms, they may simply be stored as functions of u_1 and u_2 and read out to be summed, saving multiplication operations and simplifying the hardware. For an n = 512 system, most of the correction can be implemented with 12 PROM's, a considerable saving over some of the previous systems. Unfortunately, one of the polynomial terms requires half the total storage of the two-term varying coefficient system and is better left computed.

Computing Times Some idea of the required computing times can be obtained from the frame rate of the system. Assume that r frames per second are displayed, and n^2 points per frame as before. For the all-digital system in which every point in a two term correction is stored, at

least $2rn^2$ digital operations per second are required, where the factor of 2 arises as a minimum time for a digital addition and division/multiplication. If r = 30 frames per second and n = 512, $15 \cdot 2^{20}$ or $\approx 1.63 \cdot 10^7$ operations per second are needed--quite a few for even the fastest computers. The two term computed method reduces storage at the expense of even more computing. In this case, $2rn^2$ increases to $16\,rn^2$ required digital operations. The stored polynomial terms method only doubles $2rn^2$ operations to $4rn^2$ because some terms may be stored while others are computed. Significant improvements in these times are possible with parallel hardware and simultaneous correction of neighboring points, taking into account the characteristics of the human observer.

References

1. A. A. Sawchuk, "Real-Time Correction of Intensity Nonlinearities in Imaging Systems," submitted to IEEE Transactions on Aerospace and Electronic Systems.

2. F. C. Billingsley, "Applications of Digital Image Processing," Appl. Optics, Vol. 9, pp. 289-289, 1970.

3. W. B. Green and R. M. Ruiz, "Removal of Photometric Distortion from Mariner-9 Television Images," J. Opt. Soc. Amer., Vol. 62, pp. 1351-1352, 1962.

4.2 Positive Restoration by Mathematical Programming

Nelson D. A. Mascarenhas

The utilization of the fact that pixels are nonnegative quantities can lead to improvement in the methods for restoring images. This is the result of a better use of prior information in a statistical decision process.

It is known that two dimensional filtering can be formulated as a one dimensional vector operation [1]. A convenient model is

$$\underline{y} = H\underline{x} + \underline{z}$$

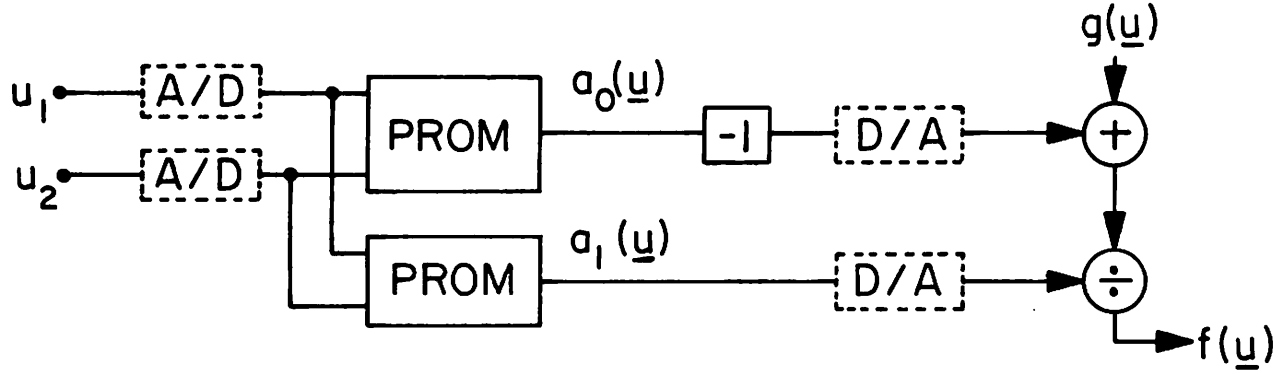


Figure 4.1-2. Two-term varying correction-stored.

where $\underline{y}$ is a $(m \times 1)$ vector of observed values

H is the $(m \times n)$ blur matrix

$\underline{x}$ is a $(n \times 1)$ vector of original pixel values

$\underline{z}$ is the random $(m \times 1)$ vector representing the noise process.

Several techniques for estimating $\underline{x}$ based upon mathematical programming methods are now considered.

Quadratic Programming In many situations, in addition to a lower bound on the pixel values, one also knows an upper bound. In this formulation of the restoration problem it is assumed that the vector $\underline{x}$ is uniformly distributed in the n-dimensional region defined by

$$\underline{\ell} \leq \underline{x} \leq \underline{u}$$

Also $\underline{z}$ is assumed to be a zero mean gaussian distributed vector with known covariance matrix V.

Under the maximum a posteriori (MAP) estimation criterion (or the maximum likelihood, since $p(\underline{x})$ is a constant in the interval) one looks for the vector $\hat{\underline{x}}$ such that $p(\hat{\underline{x}} \,|\, \underline{y}) \geq p(\underline{x} \,|\, \underline{y})$ for any $\underline{x}$. Using the fact that the logarithm is a monotonic increasing function, and also the gaussian assumption on the noise, it is equivalent to maximize the function

$$\log p(\underline{x} \,|\, \underline{y}) = \log p(\underline{x}) = \tfrac{1}{2}(\underline{y} - H\underline{x})^T V^{-1} (\underline{y} - H\underline{x}) - \log p(\underline{y})$$

Since $p(\underline{x})$ is constant on the interval $\underline{\ell} \leq \underline{x} \leq \underline{u}$, the problem is reduced to minimizing the quadratic form

$$\tfrac{1}{2}(\underline{y} - H\underline{x})^T V^{-1} (\underline{y} - H\underline{x})$$

under the linear constraints $\underline{\ell} \leq \underline{x} \leq \underline{u}$. It should be observed that this formulation gives the same result as a weighted least squares criterion under the same linear constraints. The original estimation problem, therefore, has been transformed into a gaussian programming problem

under linear constraints for which there are very efficient computational algorithms. One of these, the Wolfe's algorithm, is based on a modification of the Simplex method for linear programming [2].

Linear Programming Assume now that the noise components z_i are independent and identically distributed according to an exponential distribution:

$$p(z_i) = \tfrac{1}{2} \exp \{ - | z_i | \}$$

Following steps analogous to the previous derivations, it is easy to verify that in this case the function to be minimized is given by

$$Q(x) = \sum_{i=1}^{n} | Y_i - (H\underline{x})_i |$$

under the constraints $\underline{\ell} \leq \underline{x} \leq \underline{u}$. In this case the same result could have been obtained through the criterion of estimation of least sum of absolute deviations under linear constraints.

This problem can be formulated in terms of linear programming [3]. That is one seeks

$$\underset{\underline{x}}{\text{Min}} \sum_{i=1}^{n} (\epsilon_{i1} + \epsilon_{i2})$$

such that

$$(Hx)_i - y_i = (\epsilon_{i2} - \epsilon_{i1})$$

and

$$\epsilon_{i1} \leq 0$$

$$\epsilon_{i2} \leq 0$$

As another possible criterion, suppose that the objective is to minimize the maximum absolute deviation in the regression model

(Chebyshev criterion). Then one seeks

$$\underset{\underline{x}}{\text{Min}} \left(\underset{i}{\text{Max}} \left| y_i - (Hx)_i \right| \right)$$

Under this assumption the problem again can be reduced to a linear programming formulation [3] of determining

$$\underset{\underline{x}}{\text{Min}} (\epsilon)$$

such that

$$-\epsilon \leq y_i - (H\underline{x})\, i \leq \epsilon$$

Large Scale Mathematical Programming In the case of image processing a very basic problem is found in the solution of these mathematical programming schemes. This has to do with the extremely large dimensionality that can be involved when two dimensional data is converted into vector form. Therefore, research is underway to find feasible computational methods of large scale mathematical programming to deal with this problem.

Other Areas of Investigation Other possible areas for investigation include: a) the use of "soft" constraints [4] where the actual estimator is a linear convex combination of the constrained and unconstrained solutions; b) the existence of constraints on linear combinations of pixels $A\underline{x} \leq \underline{t}$ where A may not be full rank. In this case the solution would involve the use of a matrix pseudo-inverse. Research is needed on the most suitable computational methods to solve the problem.

References

1. W. K. Pratt, "Vector Space Formulation of Two Dimensional Signal Processing Operations," Internal Report, Department of Electrical Engineering, University of Southern California, February 1973.

2. P. Wolfe, "The Simplex Method for Quadratic Programming, Econometrica, Vol. 27, No. 3, July 1959.

3. H. M. Wagner, "Linear Programming Techniques for Regression Analysis," American Statistical Association Journal, March 1959.

4. A. Zellner, "Decision Rules for Economic Forecasting," Econometrica, Vol. 31, No. 1-2, January - April 1963.

4.3 Positive Restoration by Deconvolution

Ronald S. Hershel

Generally, linear methods are used to restore pictures degraded by diffraction and noise. These models generally assume a stationary and linear representation of the picture formation process as given by

$$i = s \otimes o + n \qquad (1)$$

$i \equiv$ sampled $N \times N$ image

$o \equiv$ sampled $N \times N$ object

$s \equiv$ sampled $N \times N$ point spread function

$n \equiv$ sampled $N \times N$ noise array

An optimum deconvolution filter d can be found such that the estimated object o^* where

$$o^* = d \otimes i \qquad n^* = i - s \otimes o^* \qquad (2)$$

is statistically most likely to have occurred. Two major problems arise when using the restoring formula of Eq. (2) for pictures which are highly degraded (i.e., s is broad compared to the Nyquist sampling interval). The first is the high sensitivity to image noise, thereby requiring a conservative deconvolution filter d. The other is the appearance of unphysical solutions for o due to the analytic representation. The severity of both effects can be lessened by applying more a priori information to the restoring algorithm.

First one should realize that the formula of Eq. (2) is based only on the second order statistics of o, which proves adequate when o is of

low contrast ($|o^* - 0|$ is small compared to $|o|$). But this condition is often violated in practice, and as a result in many cases one obtains the non-physical solution

$$o^* < 0 \tag{3}$$

To develop a more realistic statistical model for o which avoids negative intensities assume each object point to be the square of a Gaussian random variable. With

$$o_i = a_i^2$$

where the probability density of o_i becomes

$$p(o_i) \propto o_i^{-\frac{1}{2}} \exp(-o_i/2\alpha) \text{ with } \alpha = \langle o_i \rangle$$

It should be noted that $p(o_i)$ becomes complex for $o_i < 0$. The joint density for o with $\langle a_i a_j \rangle = 0$ $i \neq j$ then becomes

$$p(o) = \prod_i p(o_i) \tag{4}$$

and the solution to $s \otimes o = i$ which maximizes Eq. (4) is found to be

$$o^* = (s \otimes \lambda)^{-1} \tag{5}$$

such that $s \otimes \lambda > 0$ where λ is obtained by

$$\epsilon = |s \otimes o^* - i| \equiv \text{minimum} \tag{6}$$

Since there exists a unique mapping from λ to o in (5), a unique solution can always be found to (6) where in the presence of no noise the error $\epsilon = 0$.

Of particular importance is the pseudo inverse properties of (5) which optimally interpolates and extrapolates o^* to a higher dimensioned array M x M, M > N by defining s as an M x N array. Hence "super

resolution" can be obtained which requires a finer sampling than the Nyquist interval.

Other models for o are being examined which yield the following representations

$$o^* = \exp(s \otimes \lambda) \qquad \text{positive } o \leq 0$$

$$= [\exp(s \otimes \lambda) + 1]^{-1} \qquad \text{bounded } o \leq 0 \leq 1$$

Numerical Methods The nonlinear solution to (6) using formula (5) for large 2-D arrays requires sophisticated numerical techniques to keep calculation times reasonable. Newton-Raphson type schemes which require an explicit representation of the derivative matrix $a_{ij} = \partial^2 \epsilon / \partial \lambda_i \partial \lambda_j$ are out of the question since a_{ij} requires N^4 words! A conjugate gradient method using elementary operations on the solution obtained by equation (2) shows promise in providing rapid convergence with minimum computer time.

4.4 Multidimensional Modelling for Fast Real Time Image Enhancement

Anil K. Jain

A large percentage of real-world phenomena can be described adequately only by multidimensional processes described by partial difference equations. One of the fundamental problems that must be solved with such physical processes is the filtering of data that are known to be corrupted by noise due to both observation and modeling errors. Digital image processing, tumor detection, distribution of population, and weather prediction are just a few examples of multidimensional processes in which the filtering problem is of paramount importance. For one dimensional linear systems governed by ordinary difference or differential equations, application of the basic Kalman-Bucy filtering method is a routine matter. However, the extension of this method to

multidimensional problems introduces many computational difficulties. These difficulties include the high dimension of discrete approximations for numerical computation and the instability of many of the associated numerical algorithms [1,2]. Another approach to recursive filtering of two-dimensional images corrupted by white Gaussian noise will be described here. Special consideration is given to the nature of the two-dimensional data so that application of the standard results to large amounts of data is practically feasible. In particular, the unique features of the approach to this two-dimensional filtering technique are:

(i) A special model for two-dimensional filtering which gives an estimator very close to the optimal interpolator;

(ii) Matrix-vector filtering equations which are easily decomposed into recursive scalar equations. This amounts to a scalar filter for vector scanning of the image;

(iii) A model which allows fast implementation of the filter, thus making the method suitable for on-line image restoration;

(iv) An isotropic filter so that it is equally effective for all orientations of the image.

The filter model is based on the notion of nearest neighbor interaction of image pixels in the spatial domain. The filtering problem is formulated as a "dual control" problem (or a quadratic variational problem) via the maximum likelihood function. Although such problems (usually) lead to a two point boundary value problem, a stable initial value solution is readily obtained together with a decomposed scalar filtering algorithm eliminating all matrix operations.

The Model By assuming a stationary autocorrelation function of the image, i.e.,

$$E[x_{n,m}x_{n+i,m+j}] = R(|i|, |j|) \tag{1}$$

a nearest neighbor model can be written as a second order Markov process given by

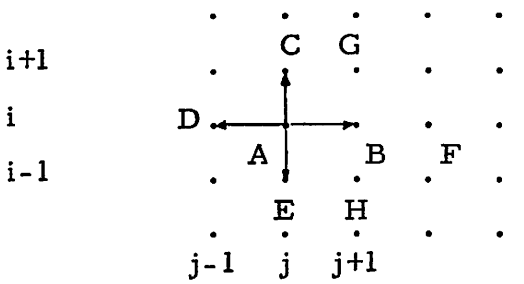
$$x_{i,j} = \alpha_1 x_{i,j+1} + \alpha_2 x_{i+1,j} + \alpha_3 x_{i,j-1} + \alpha_4 x_{i-1,j} + \beta u_{i,j} \qquad (2)$$

where α_1, α_2, α_3, α_4 are the regression coefficients, u_{ij} are white noise inputs and β^2 is the mean square error [1]. In Figure 1a, this model implies that the predicted value at A is directly correlated to the nearest four neighbors B, C, D and E. Although simpler models as shown in Figures 1b and 1c could be used, this model has the desired dimensionality reducing and isotropic features as mentioned above. Besides, for the case of non-stationary image restoration, this model easily leads to an adaptive on-line algorithm without incurring excessive computational cost. It is noteworthy that this model also represents several other physical phenomena such as the two-dimensional random walk, steady state diffusion, birth and death processes, etc. Finally, for $\alpha_i > 0$ and $i = 1,2,3,4$ eq. (2) belongs to a class of linear elliptic systems, so that there are several other similar multidimensional models [1] that can be used for image representation, and can be extended to interframe image processing [3].

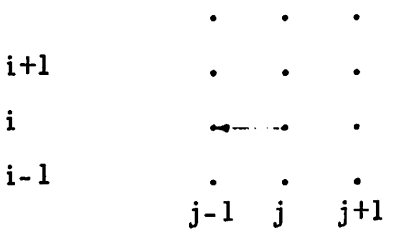
Two-Dimensional Filtering If eq. (2) is rearranged and written in vector form, and if x_i is a $(1 \times N)$ vector, then for a $N \times M$ image,

$$x_{i+1} = Qx_i - x_{i-1} + bu_i, \qquad i = 1, \ldots, M \qquad (3)$$

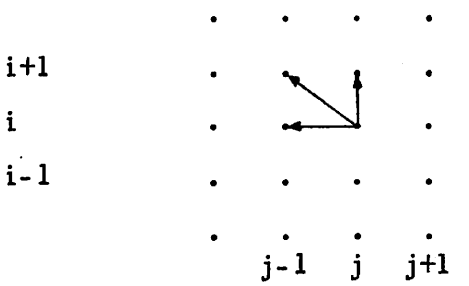
where Q is found to be a $N \times N$ tridiagonal, symmetric Toeplitz matrix for stationary (or piecewise stationary) images. For real images, Q is a positive definite matrix with bounded eigenvalues. Also, the eigenmatrix of Q is completely independent of its tridiagonal parameters and depends only on its structure (which in turn depends on the nearest neighbors chosen). Also, the components of the eigenmatrix (say H), can be related to the Fourier coefficients via a linear transformation S containing zeros



(a) nearest neighbors model



(b) line scanner model



(c) alternate model

Figure 4.4-1. Various image model configurations.

and ones only (selection matrix), so that the FFT algorithm can be used to determine H. Now, the observation vector can be written as

$$y_i = x_i + \eta_i \,, \tag{4}$$

where η_i is the additive Gaussian noise of known statistics. From Eqs. (3) and (4) one can directly write the vector Kalman filtering equations. Since Eq. (3) is a second order equation, the associated Riccati equation will be a $2N \times 2N$ matrix equation. Moreover, since Q is positive definite, the one step predictor and estimator equations are unstable. Such difficulties are inherent in elliptic systems. Thus, it might seem that the nearest neighbors model adds stability problems to the already complex problem of large dimension. However, by formulating the maximum likelihood function, the equations for the optimal interpolator (denoted by x_i itself) are given by

$$x_{i+1} = Qx_i - x_{i-1} + \sigma^2 b^2 v_{i+1} \,, \tag{5}$$

where

$$v_i = Q^T v_{i+1} + w_{i+1} + \frac{1}{\sigma_n^2} (y_i - x_i) \,, \tag{6}$$

$$w_i = -v_{i+1} \,, \tag{7}$$

subject to the boundary conditions

$$v_{N+1} = w_{N+1} = 0 \,, \tag{8}$$

$$x_i = \sigma^2 v_1 \,, \tag{9}$$

$$x_0 = \frac{\sigma^2}{1 + b^2 \sigma^2} \left[w_1 + \frac{b}{\sigma_n^2} y_0 \right] \tag{10}$$

Equations (5) to (10) give a linear boundary value problem. This can be converted to a stable nonlinear initial value problem by a Riccati transformation giving $N \times N$ matrix equations (as compared to $2N \times 2N$ in Kalman

filtering). Since eqs. (5) to (10) are linear one can write

$$x_i = r_i v_i + t_i w_i + s_i \tag{11}$$

Using Eq. (11) in Eqs. (5) and (6) and rearranging terms one can also write

$$x_i = d_i v_{i+1} + t_i w_{i+1} + g_i , \tag{12}$$

where d_i and t_i are related to r_i and t_i (in fact $t_i d_{i-1}$). It has been shown that r_i and t_i satisfy stable initial value Riccati equations and the matrices $h r_i h$, $h t_i h$, etc., are diagonal. Also, the equations for s_i and g_i are equations of one step predictor and estimator respectively. Since all the filter equations decouple under the transformation h, only scalar operations need to be performed.

One Step Interpolator and Real Time Considerations In order to compute the optimal interpolator, the entire transformed data y_i (i=1,...,M) has to be stored. However, experimentally it has been found that a one step interpolator, requiring a one step delay buffer storage gives an estimate which is close to the optimal interpolator in terms of S/N ratio improvement (see Table 1 and Figures 2 and 3). Figure 4 shows the block diagram structure for such an interpolator for an on-line operation. The total number of computational operations required is of the order of $N^2 \log_2 N$ ($8N^2$ for N = 256), compared to on the order of N^4 computations required for direct application of Kalman filtering results. The scalar filter equations lend themselves to a certain degree of parallel computation, so that a parallel processor structure for the filter could be defined. This is currently under investigation [4].

Extensions Preliminary results of implementation the above algorithm on 256 x 256 images as shown (Figure 5) have indicated that image representation by elliptic models can be exploited to develop real time methods for:

TABLE 1

	Ratio	σN^2	1-Step Interpolator Improvement in Db	Optimal Interpolator Improvement in Db
Square Signal	7/3	9	7.98	8.94
S Signal	6.8/3	9	7.08	7.57

(i) Adaptive image enhancement to account for nonstationary statistics of images;

(ii) Frame to frame (or multi-sensor) image restoration; and

(iii) Enhancement of images degraded by elliptic point spread functions and additive noise.

These aspects are currently being studied.

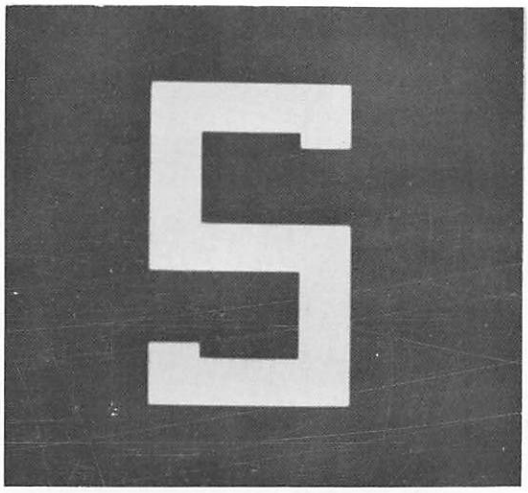
References

1. A. K. Jain and E. Angel, "Image Restoration, Modelling and Reduction of Dimensionality," USC Tech. Report No. 72-50, (submitted for publication).

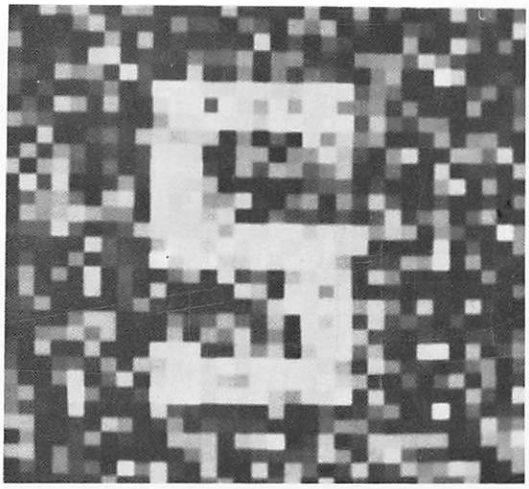
2. E. Angel and A. K. Jain, "A Dimensionality Reducing Model for Distributed Filtering," IEEE Trans. Aut. Control, Vol. AC-18, No. 1, Feb., 1973, pp. 59-62.

3. E. Angel and A. K. Jain, "Filtering of Multidimensional Diffusion Processes," 6th Asilomar Conference on Systems, Nov., 1972.

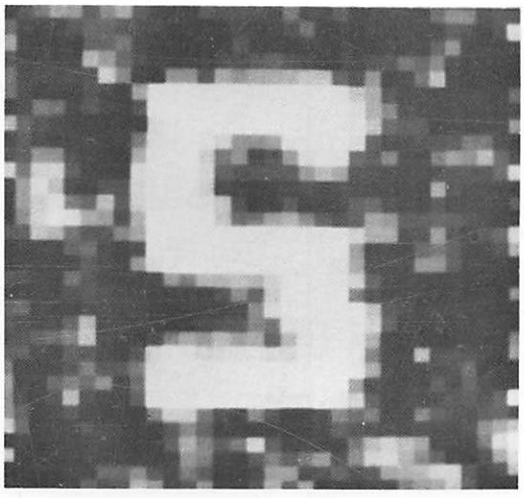
4. A. K. Jain, "Real Time Recursive Image Enhancement," (in preparation).



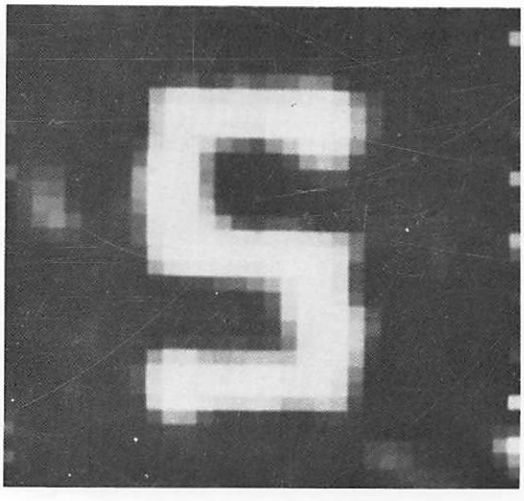
(a) original image



(b) original image corrupted by noise of variance 9

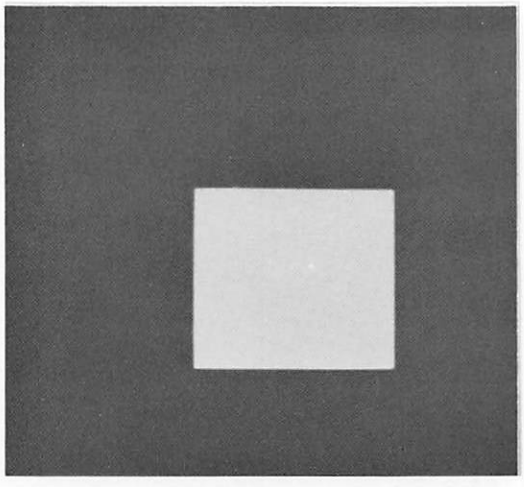


(c) one step interpolation

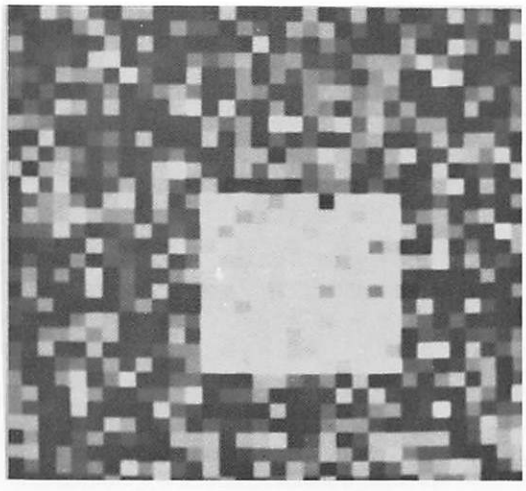


(d) optimal interpolation

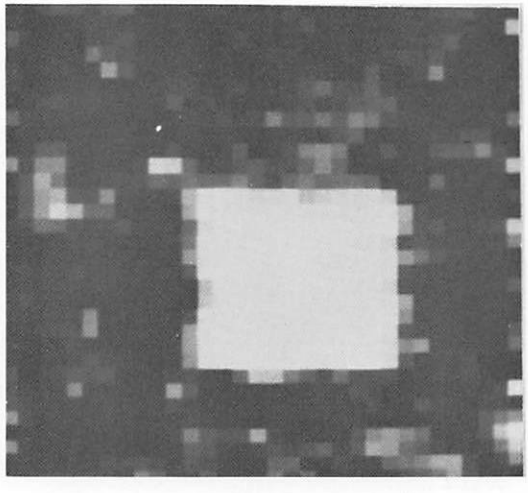
Figure 4.4-2. Filtering of the "S" signal.



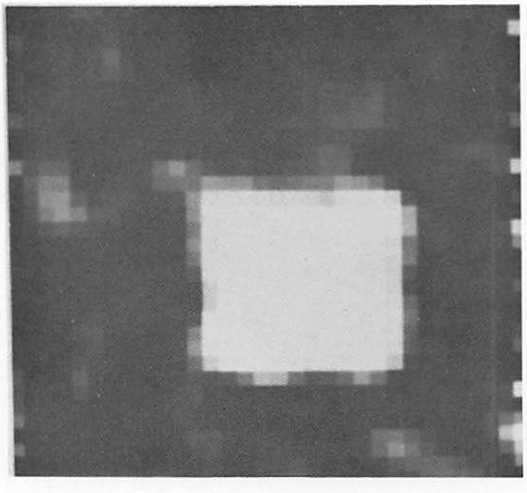
(a) original image



(b) original image corrupted by noise of variance 9



(c) one step interpolation



(d) optimal interpolation

Figure 4.4-3. Filtering of the square signal.

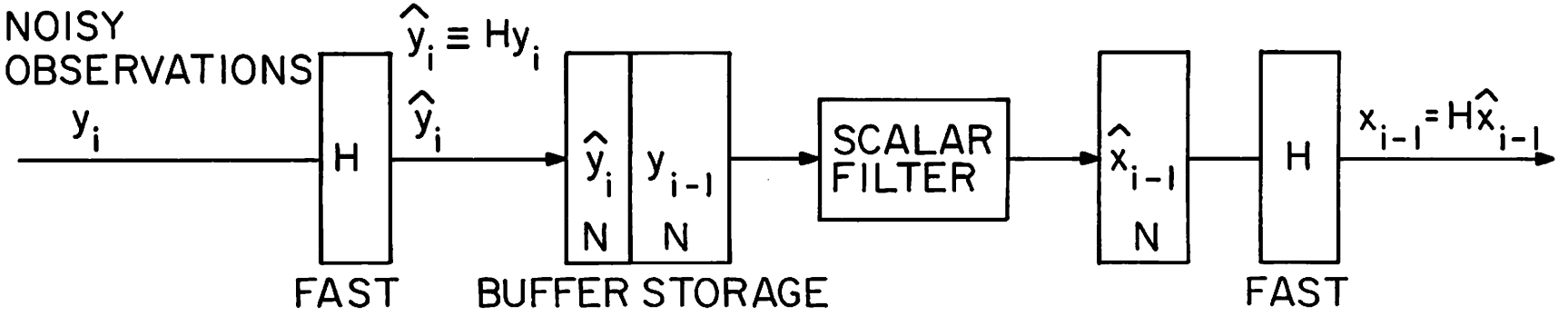


Figure 4.4-4. On-line one step interpolator.

(a) noisy image, S/N = 2

(b) one step interpolator
11.5db improvement

(c) noisy image, S/N = 2

(d) one step interpolator
10db improvement

Figure 4.4-5. Implementation of fast recursive filter on 256 x 256 images.

4.5 Nonstionary Recursive Image Restoration

Nasser E. Nahi and Touraj Assefi

The role of Kalman filtering in image restoration has now been established [1]-[6]. It has been shown that Kalman filtering is a very suitable method of image restoration when used in conjunction with appropriate dynamic models of the image processes. Additional improvement in estimation performance seems possible if the information concerning the position of the scanner with respect to the picture boundary is directly utilized by the estimator. A method to implement this idea follows.

Let a horizontally scanned picture be decomposed into M vertical strips and let s(t) be the scanner output and $s_m(t)$ be the portion of the output associated with m^{th} strip. Statistical properties of $s_m(t)$ have been determined for all m, $1 \leq m \leq M$. It has been shown that if the original image has a stationary (two dimensional) statistical property, the statistics of $s_m(t)$ are all identical except for $s_M(t)$. Each random process $s_m(t)$ has been modeled as in [1] or [2] and corresponding Kalman filters derived. In the implementation, as the scanner enters the m^{th} strip at j^{th} horizontal line, the initial condition for the estimator is updated using the preceding estimate from the $(m-1)^{st}$ strip and its covariance. Consequently, the main additional complexity over previous methods is in the fact that it is necessary to store the estimates and their covariances associated with the preceding scan line (one horizontal line preceding the present scanner position). This is only a slight added complexity. At present this estimator is being implemented to verify the expected improvements especially in the vicinity of the edges of the picture.

References

1. N. E. Nahi, "Role of Recursive Estimation in Statistical Image Enhancements," Proceedings of IEEE, Vol. 60, No. 7, July 1972, pp. 872-877.

2. N. E. Nahi and T. Assefi, "Bayesian Recursive Image Estimation," IEEE Transactions on Computers, Vol. C-21, No. 7, July 1972, pp. 734-738.

3. S. R. Powell and L. M. Silverman, "Modeling of Two-Dimensional Covariance Functions with Application to Image Restoration," to be presented at the 1972 Conference on Decision and Control.

4. A. Habibi, "Two-Dimensional Bayesian Estimate of Images," Proceedings of IEEE, Vol. 60, No. 7, July 1972, pp. 878-884.

5. N. E. Nahi and C. A. Franco, "Application of Kalman Filtering to Image Enhancement," 1972 Conference on Decision and Control, Nes Orleans, Louisiana.

6. N. E. Nahi and C. A. Franco, "Recursive Image Enhancement: Vector Processing," IEEE Transactions on Communications, May 1973.

4.6 Nonlinear Recursive Image Restoration

Nasser E. Nahi and Mohammed Jahanashahi

A typical image in the absence of any distortion or noise consists of an object of interest within a background. The object, such as a human face, usually contains detailed information which is essential in the quality of the final enhanced image. In addition to this detailed information content, the object represents a shape defined by its boundary. Generally, both the detailed information and boundary information may be characterized by statistical measures such as the mean and autocorrelations. This is the only a priori knowledge available to the image restoration system.

If one now attempts to represent the entire image (the object within a background) by first and second order moments, the statistics of the boundary will usually overwhelm the statistical information on the details of the object. It is then expected that while the minimum mean square estimate of the object given a noisy observation provide good restoration of the object as an area within the image, it does result in undesirable

blurring of the detail within the object. This is clearly apparent in references [1] -[4] especially in Figure 8 of the latter reference. Consequently, the estimate of the object should be composed of two estimates: 1) estimate of the object boundary, and 2) estimate of object detail.

Let the output of the scanner be represented by S(t). Furthermore, let $S(t) = A(t)\,U(t)$ where

$$U(t) = \sum_{j=0}^{M} [\,u(t-a_j) - u(t-b_j)]$$

with

$A(t)$ = gray level of the image at scanning time t,

M = number of lines in the image,

$u(\cdot)$ = step function,

a_j = start of the object in line j,

b_j = end of the object in line j.

The justification for choosing S(t) as above is due to the apriori knowledge that there exists an object of interest in the image and the problem is to determine the location of the object. Sampling S(t) at intervals $t = 0, 1, 2, \ldots, N$, where N denotes the number of pixels in the entire image, will result in $S(k) = A(k)U(k)$, $k = 0, 1, \ldots, N$. Now given the observation sequence $Y(k) = S(k)+V(k)$, with V(k) as the observation noise, one may estimate the unknown process A(k), $k = 0, 1, \ldots, N$, and a_j, b_j, $j = 0, 1, \ldots, M$, based on maximum aposteriori probability criterion. Let A(k) be independent of a_j and b_j. Assume A(k) is a zero mean Gaussian random process with $E[A(i)A(j)] = P(i,j)$, and V(k) is a zero mean Gaussian white noise with $E[V(i)V(j)] = o^2\Delta(i-j)$, where $\Delta(\cdot)$ denotes the Kronecker delta. A map estimate of A(k), a_j, b_j is found to satisfy the following relationships:

$$\hat{A}(k) = L(k,k)Y(k) + \sum_{m=0}^{k-1} L(k,m)[\,Y(m) - \hat{A}(m)]\,, \qquad (1)$$

where

$$L(k,m) = \frac{P(k,m)\,\hat{U}(m)}{o^2 + P(k,k)\hat{U}(k)} \quad , \tag{2}$$

and

$$\hat{U}(m) = \sum_{j=0}^{M} [\,u(m-\hat{a}_j) - u(m-\hat{b}_j)] \tag{3}$$

$$\text{minimum} \quad \frac{1}{o^2} \sum_{m=0}^{k} [\,\hat{A}^2(m) - 2\hat{A}(m)Y(m)]U(m) - 2\,\text{Log}_e f(a_1,\ldots,a_M,b_1,\ldots,b_M)$$

$$a_1,\ldots,a_M$$

$$b_1,\ldots,b_M \tag{4}$$

where $f(\cdot)$ = joint density function. It can be shown that $\hat{A}(k)$ for $\hat{a}_j \leq k \leq \hat{b}_j$, $j = 0,1,\ldots,M$, takes the same values as estimates of $S(k)$ obtained by a Kalman Filter as in [1], [2]. The next step is to develop an algorithm for computing $\hat{a}_j$ and $\hat{b}_j$ from relation (4) above. It seems that, due to the nature of the minimization, it will be possible to develop simple procedures to accomplish this objective. A number of procedures are under consideration. Once $\hat{a}_j$ and $\hat{b}_j$ are determined for $j = 0,1,\ldots,M$, one can obtain an estimate of the size of object boundary. Furthermore, values of $\hat{A}(k)$ should result in a high detail restoration.

References

1. N. E. Nahi, "Role of Recursive Estimation in Statistical Image Enhancement," Proceedings of IEEE, Vol. 60, No. 7, July 1972, pp. 872-877.

2. N. E. Nahi and T. Assefi, "Bayesian Recursive Image Estimation," IEEE Transactions on Computers, Vol. C-21, No. 6, July 1972, pp. 734-738.

3. S. R. Powell and L. M. Silverman, "Modeling of Two-Dimensional Covariance Functions with Application to Image Restoration," presented at the 1972 Conference on Decision and Control.

4. A. Habibi, "Two-Dimensional Bayesian Estimate of Images," Proceedings of IEEE, Vol. 60, No. 7, July 1972, pp. 878-884.

4.7 Space-Variant Image Motion with Different Geometrical Distortion Models

Alexander A. Sawchuk

The problem of a posteriori image quality improvement is simplified when a particular system model of degradation can be assumed. Although incoherent optical systems are linear in intensity, the blurring of each point of the object by the system generally varies with position so that a space-variant point-spread function (SVPSF) $h(\underline{x}, \underline{u})$ in the superposition integral

$$\mathcal{I}(\underline{x}) = \int_{-\infty}^{\infty} h(\underline{x}, \underline{u})\, \mathcal{O}(\underline{u})\, d\underline{u} \tag{1}$$

must be assumed. Here $\mathcal{O}(\underline{u})$ is the original object intensity function, $\mathcal{I}(x)$ is the image intensity recorded by the system, and $h(\underline{x}, \underline{u})$ is the response as a function of image coordinates $\underline{x} = (x_1, x_2)$ to a unit impulse at $\underline{u} = (u_1, u_2)$ in the object coordinates [1], [2]. The difference variables $\underline{x}$ and $\underline{u}$ are used because the object and image may be measured in different coordinate systems.

Space-variant point-spread functions often result from moving optical systems with geometrical distortions. These distortions are fixed coordinate transformations, so that distances, locations and intensities measured in one system may not be preserved when measured in another. In the absence of motion, the geometrical coordinate transformation c uniquely relates the location of a point $\underline{x}$ in one system to the location of its conjugate point in another system by the parametric equations

$$u_1 = c_1(x_1, x_2) \tag{2a}$$

$$u_2 = c_2(x_1, x_2) \tag{2b}$$

which may be expressed in vector form by

$$\underline{u} = c(x) \tag{3}$$

The distortion is a memoryless one-to-one mapping between points and may be a nonlinear function. With these properties, the system (2) may be inverted and solved to produce

$$\underline{x} = c^{-1}(\underline{u}) \tag{4}$$

with corresponding parametric form. Since these transformations are memoryless point mappings, there is no smearing, averaging or integration due to the distortions alone.

Depending on the nature of the physical process which produces the distortions, two different expressions for intensity functions which undergo distortion are obtained. The first expression arises from the imaging of diffuse planar objects (such as by aerial photographic systems) or from electronic or digitally computed distortions which simply manipulate point locations without energy conservation requirements (as in certain display devices).

Distortions described by this first physical model occur in particular when the direction of observation by the recording imaging system is not perpendicular to the plane on which the object function intensity is plotted. For extended sources such as the object, radiative intensity is denoted by a quantity called the radiance, which depends on how the object is illuminated, how it reflects, transmits and absorbs, and on the direction from which it is viewed [3]. To simplify this, an assumption for the first model is that the object function is self-radiating and that it is a lambert, or perfectly diffusing surface. The lambert surface has the same radiance as a function of location on the plane no matter what the viewing direction, so that the only effect on the intensity as seen by the imaging system is a change in the geometrical shape and relative location of object points, without any measured intensity change. The same physical model also holds for certain electronic or digital display or graphics devices which perform shape distortions without intensity changes. Denoting an object intensity function in a rectilinear space $\underline{u} = (u_1, u_2)$ by $\mathcal{O}(u_1, u_2)$, the

distorted function $\mathcal{D}(x_1, x_2)$ plotted in the $\underline{x}$ coordinates is given by

$$\mathcal{D}(x_1, x_2) = \mathcal{O}[c_1(x_1, x_2),\ c_2(x_1, x_2)] = \mathcal{O}(u_1, u_2) \tag{5}$$

where one simply substitutes from (2) and equates the intensity levels at conjugate point locations.

The second distortion model is appropriate when distortions arise from aberrations in optical systems which image parallel entrance and exit planes, or when distortion is caused by nonlinearities in image sensors and related electronics for certain scanned imaging systems. With this model, eq. (5) describing the distortion must be modified from a single point movement by adding a multiplicative Jacobian term

$$|c'(\underline{x})| = \begin{vmatrix} \dfrac{\partial u_1}{\partial x_1} & \dfrac{\partial u_1}{\partial x_2} \\ \dfrac{\partial u_2}{\partial x_1} & \dfrac{\partial u_2}{\partial x_2} \end{vmatrix} \tag{6}$$

giving

$$\mathcal{D}(x_1, x_2) = |c'(\underline{x})|\, \mathcal{O}[\, c_1(x_1, x_2),\ c_2(x_1, x_2)] \tag{7}$$

for the distorted image. This expression is derived using a constant energy imaging model.

With an object function moving in the field of view of the recording optical system during exposure, an equivalent linear SVPSF can be derived [1], [5], [6]. A motion function uniquely relating any point $\underline{u}$ in one frame to the location of its image point $\underline{x}$ in the other frame is assumed known at any time instant in the exposure interval $[0, T]$. This mechanical description is written parametrically as

$$u_1 = g_1(x_1, x_2; t) = g_1(\underline{x}; t) \tag{8a}$$

$$u_2 = g_2(x_1, x_2; t) = g_2(\underline{x}; t) \tag{8b}$$

and may be derived from the distortion properties of the system with knowledge of the relative movement. Omitting the details, the expression

$$\mathcal{L}(\underline{x}) = \int_{\underline{u} = \underline{g}(\underline{x}; 0)}^{\underline{u} = \underline{g}(\underline{x}; T)} \frac{\mathcal{O}(\underline{u}) \, |J_g(\underline{x}; k_1(u_1; \underline{x})|}{\left[\left(\frac{\partial g_1}{\partial t}\right)^2 + \left(\frac{\partial g_2}{\partial t}\right)^2\right]^{\frac{1}{2}}} \, ds_u \Bigg|_{t = k_1(u_1; \underline{x}) = k_2(u_2; \underline{x})} \tag{9}$$

is obtained for the system SVPSF assuming the energy conserving model. The various functions in (9) are the Jacobian $J_g(\underline{x}; t)$ and the function

$$t = k_1(u_1; \underline{x}) = k_2(u_2; \underline{x}) \tag{10}$$

obtained by inverting each motion function (8) to show the dependence of t on $\underline{u}$ for a fixed $\underline{x}$. Equation (10) specifies a path of integration followed in the $\underline{u}$ plane to obtain the image intensity at point $\underline{x}$, and may be multiple-valued when the motion retraces or crosses its own path. The denominator of the integrand in (9) is the speed of movement of an object point, so that the amplitude of the response is inversely proportional to this quantity. The absolute value brackets in (9) are included to ensure that the PSF remains ≥ 0 as required by the incoherent model. For the lambert imaging model of distortion, the overall expression is the same as (9) except that the Jacobian factor $J_g(\underline{x}; t)$ is omitted from the numerator.

It is possible to considerably simplify Eq. (9) for special cases of moving imagery. For line images or two dimensional images in which the blur occurs in a straight line along one spatial dimension, a one-dimensional model is useful [6]. A different simplification using either distortion model occurs when object and image planes are parallel and undergo a relative translation during exposure [4], [6]. The motion functions of Eq. (8) become

$$u_1 = x_1 - m_1(t) = g_1(\underline{x}; t) \tag{11a}$$

$$u_2 = x_2 - m_2(t) = g_2(\underline{x}; t) \tag{11b}$$

and (9) reduces to a space-invariant operation. Some photographs illustrating this process have been presented previously.

It has also been shown previously that the motion functions describing certain types of aerial imaging take the form

$$u_1 = c_1(x_1, x_2) - m_1(t) \tag{12a}$$

$$u_2 = c_2(x_1, x_2) - m_2(t) \tag{12b}$$

where the c_1 and c_2 represent various types of distortion. From this description an overall system SVPSF may be found, and efficient techniques for image restoration follow immediately [1], [7]. These motion functions correspond to an integration operation followed by a geometrical distortion.

Another kind of space-variant blurring occurs when the order of these operations is reversed. Motion functions for this type of system take the form

$$b_1(u_1, u_2) = x_1 - m_1(t) \tag{13a}$$

$$b_2(u_1, u_2) = x_2 - m_2(t) \tag{13b}$$

where the $b_i(\underline{u})$ represent distortion functions, and Figure 1 shows the overall system. This model for motion blur describes the imaging by a distorting system onto a moving image plane, as in some types of photographic recording and motion picture systems.

It may be shown that the overall effect of the system is described by the space-variant operation

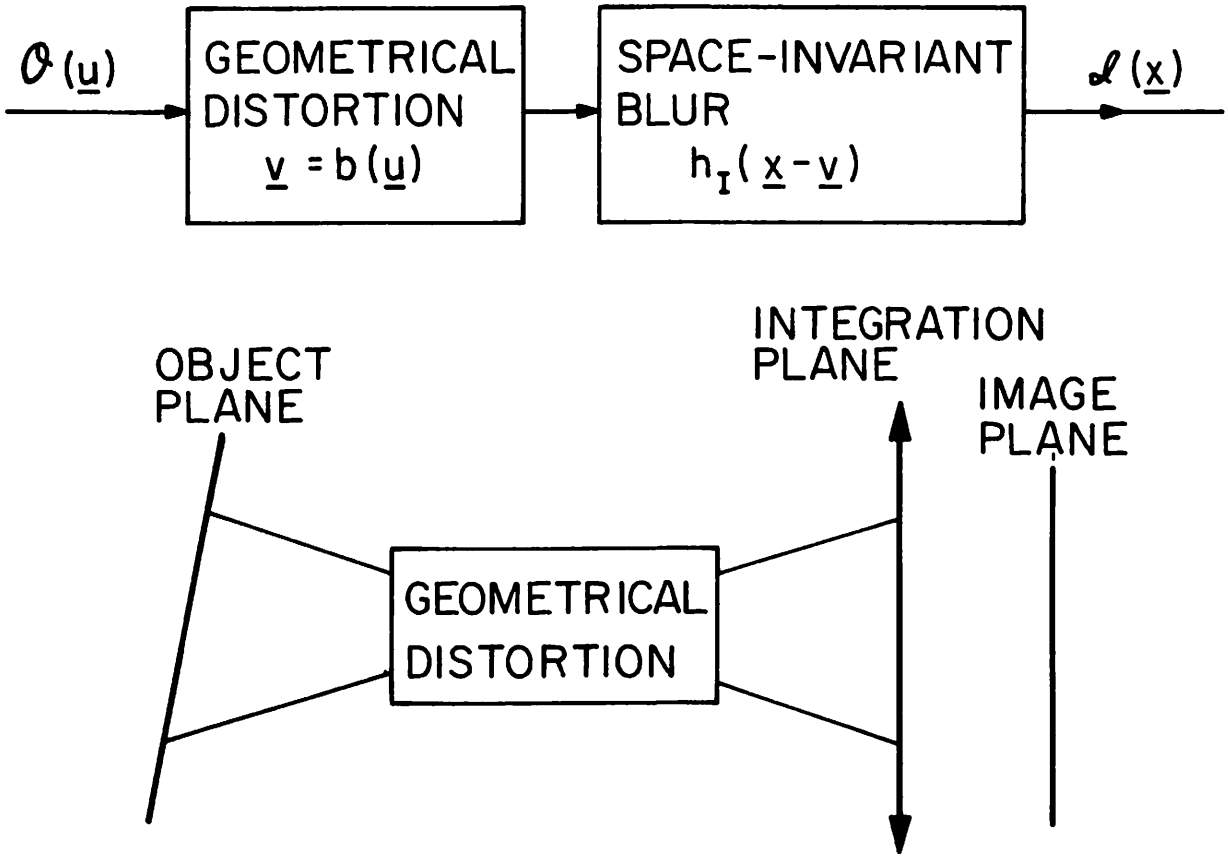


Figure 4.7-1. Distortion - integration cascade.

$$\mathcal{L}(\underline{x}) = \int_{\underline{v} = \underline{x}-m(0)}^{\underline{v} = \underline{x}-m(T)} \frac{(b^{-1}(\underline{v})|b'^{-1}(\underline{v})|\, ds_v}{\left[(\dot{m}_1(t))^2 + (\dot{m}_2(t))^2\right]^{\frac{1}{2}}} \qquad (14)$$

$$t = m_1^{-1}(x_1 - v_1)$$

$$= m_2^{-1}(x_2 - v_2)$$

assuming the energy conservation model.

If invertible coordinate transformation preceding and following an integration are allowed, a general class of space-variant point-spread functions is obtained which includes the two previously described types. When motion functions for a moving system can be written in the form

$$b_1(u_1, u_2) = c_1(x_1, x_2) - m_1(t) \qquad (15a)$$

$$b_2(u_1, u_2) = c_2(x_1, x_2) - m_2(t) \qquad (15b)$$

then the overall space-variant operation $h(\underline{x}, \underline{u})$ may be decomposed into a cascade of space-invariant PSF's and geometrical distortions [2], [4]. This description is of significant value in the restoration of images degraded by space-variant operations [1].

References

1. A. A. Sawchuk, "Space-Variant Image Motion Degradation and Restoration," Proceedings of the IEEE, Vol. 60, p. 854, 1972.

2. A. A. Sawchuk, "Image Restoration by Space-Variant Decomposition," USC Semiannual Technical Report 425, 1 March 1972 - 31 August 1972, pp. 75-82.

3. M. Born, E. Wolf, Principles of Optics, (4th Ed., Pergamon Press, London, 1970), p. 181.

4. A. A. Sawchuk, "Restoration of Motion Degraded Images," USC Semiannual Technical Report 425, 1 March 1972 - 31 August 1972, pp. 83-87.

5. A. A. Sawchuk, "Linear Space-Variant Motion Degradation and Restoration," Journal of the Optical Society of America, Vol. 62, p. 726, 1972, Spring Meeting, Optical Society of America, New York, 1972.

6. A. A. Sawchuk, "Space-Variant System Analysis of Image Motion," submitted to Journal of the Optical Society of America.

7. A. A. Sawchuk, "Coordinate Transformations in Space-Variant Image Enhancement and Restoration," Journal of the Optical Society of America, Vol. 62, p. 1337, 1972, Annual Meeting, Optical Society of America, San Francisco, 1972.

4.8 Rotational Motion Degradation and Restoration

Alexander A. Sawchuk

An important example of space-variant motion degradation is the case of rotation between parallel object and image planes in an aerial system. Assuming rotation at constant angular velocity ω, the motion functions

$$u_1 = x_1 \cos \omega t + x_2 \sin \omega t \qquad t \in [0, T];\ \omega > 0 \qquad (1a)$$

$$u_2 = -x_1 \sin \omega t + x_2 \cos \omega t \qquad (1b)$$

describe a counter-clockwise rotation of the object coordinates (u_1, u_2) with respect to the image coordinates (x_1, x_2). Setting the exposure time T equal to $\pi/3\omega$, rotation of the parallel planes proceeds through $\pi/3$ radians during exposure, and space-variant point-spread function

$$h(\underline{x}, \underline{u}) = \begin{cases} \dfrac{1}{\omega\left[u_1^2 + u_2^2\right]^{\frac{1}{2}}}, & x_1^2 + x_2^2 = u_1^2 + u_1^2,\ \dfrac{\sqrt{3}u_1}{2} - u_2/2 \le x_1 \le u_1, \\ & u_2 \le x_2 \le \dfrac{u_1}{2} + \dfrac{\sqrt{3}}{2} u_2 \\ 0, & \text{elsewhere} \end{cases} \qquad (2)$$

is obtained. The function is plotted in Figure 1 and shows how the response decreases inversely with object distance from the origin. The point spread function (PSF) is non-zero only on the paths of motion of the points in image space.

If coordinate distortions can be found which make all object points follow the same path in the image plane, then the restoration of this space-variant degradation is possible [1],[2]. Such distortions are the simple polar coordinate distortions

$$x_1 = r_z \cos \theta_z \tag{3a}$$

$$x_2 = r_z \sin \theta_z \tag{3b}$$

$$u_1 = r_v \cos \theta_v \tag{3c}$$

$$u_2 = r_v \sin \theta_v \tag{3d}$$

Equation (3) and the inverse distortions

$$\theta_z = \tan^{-1}\left(\frac{x_2}{x_1}\right) \tag{4a}$$

$$r_z = \left(x_1^2 + x_2^2\right)^{\frac{1}{2}} \tag{4b}$$

$$\theta_v = \tan^{-1}\left(\frac{u_2}{u_1}\right) \tag{4c}$$

$$r_v = \left(u_1^2 + u_2^2\right)^{\frac{1}{2}} \tag{4d}$$

can be used to transform object, image and any vertical rotation PSF's to space-invariant form. The PSF of Eq. (2) for the constant ω example transforms to

$$h_I(r, \theta_z - \theta_v) = \begin{cases} \frac{1}{\omega}, & 0 \leq \theta_z - \theta_v \leq \omega T \\ 0, & \text{elsewhere} \end{cases} \tag{5}$$

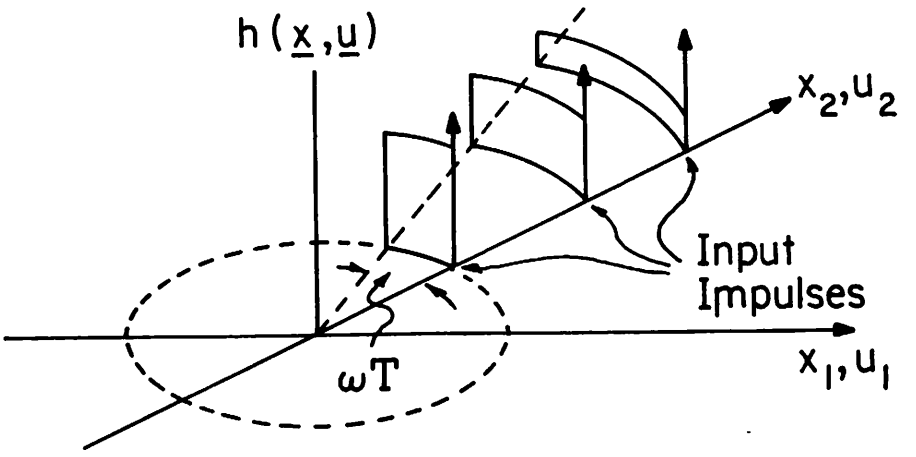


Figure 4.8-1. SVPSF for constant velocity rotation.

where $r = r_z = r_v$ since there is no blur in the radial direction. This transformed PSF is just a uniform space-invariant motion blur in the θ direction. The equation

$$\mathcal{J}_z(r, \theta_z) = \int_{\theta_v = \theta_z - \omega T}^{\theta_v = \theta_z} \frac{\theta_v(r, \theta_v)}{\omega} \, d\theta_v \tag{6}$$

expresses the blur where $\mathcal{J}_z(\cdot)$ and $\mathcal{O}_v(\cdot)$ are transformed object and image functions. The polar mapping is applied so the $h_I(\cdot)$ is defined only for $0 \leq \theta \leq 2\pi$ while $\mathcal{J}_z(\cdot)$ and $\mathcal{O}_v(\cdot)$ have values for all $\theta_v > 0$, $\theta_z \geq 0$, and are periodic with 2π.

For more complicated cases in which $\theta(t)$ is a constant angular acceleration, for example, $\theta(t) = t^2$, then a rather involved expression for $h(\underline{x}, \underline{u})$ results although the decomposition method for describing the response still applies.

Once the degradation is converted to space-invariant form, the usual inverse filtering or statistical estimation techniques [1] may be used for restoration. The overall system dimensionality is reduced by a factor of two and the space-invariant filter processing can be done on a line-by-line basis with filter coefficients which are functions only of the coordinate difference. Following the restoration process, a reverse polar coordinate transformation is used to return to the estimated object in the original rectangular coordinate system.

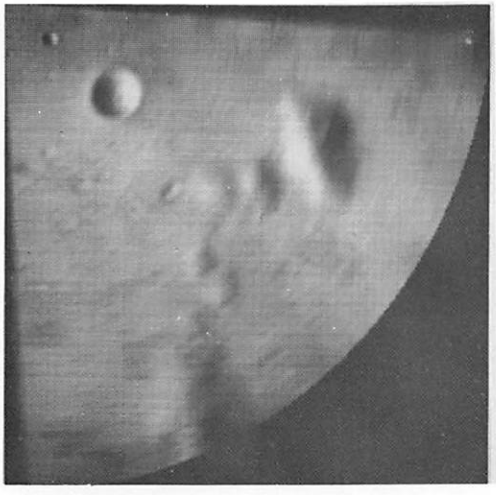
The effects of rotational blur have been simulated using aerial object scenes. Figures 2a and 3a show the effect of a constant velocity rotational blur of 3.9^o about the upper left hand corner of the object. The blur which increases with distance from the center of rotation is quite noticeable. Applying the polar coordinate decomposition of Eqs. (3) to (4), the intermediate object $\mathcal{J}_z(\underline{z})$ shown in Figures 2b and 3b is obtained. In this case $\mathcal{J}_z(\underline{z})$ is just an image blurred by a linear smear in the θ direction. An inverse filter using the fast Fourier transform is applied

to remove the blur, and the intermediate object $\mathcal{O}_v(\underline{v})$ is shown in Figure 2c and 3c. A final reverse geometrical distortion produces the restored object shown in Figure 2d and 3d.

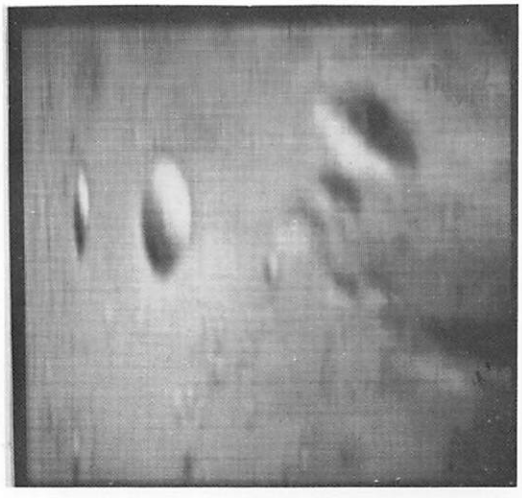
References

1. A. A. Sawchuk, "Space-Variant Motion Degradation and Restoration," Proc. IEEE, 60, 854, (1972).

2. A. A. Sawchuk, "Coordinate Transformations in Space-Variant Image Enhancement and Restoration," J. Opt. Soc. Amer., 62, 726 (1972).

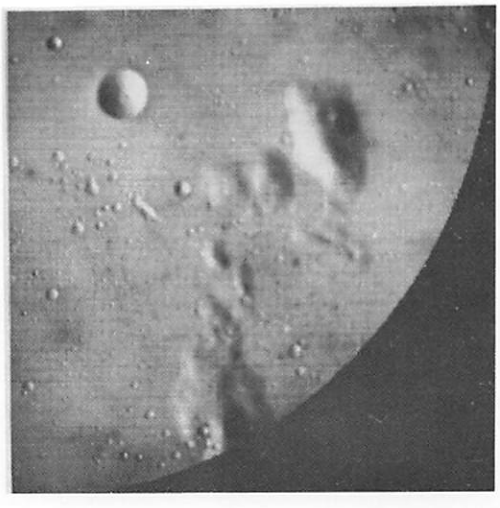


(a) images blurred by rotational motion



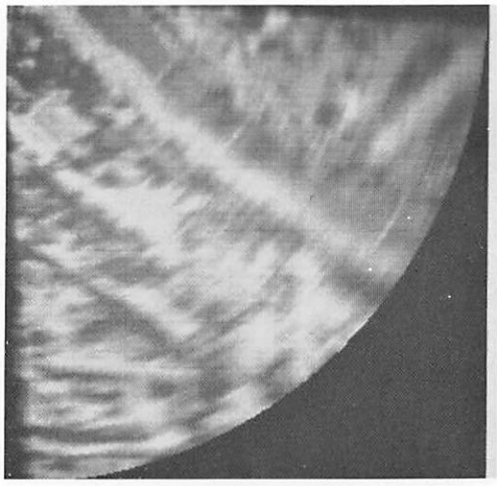
(b) image after polar coordinate distortion

(c) restoration by space-invariant inverse filter

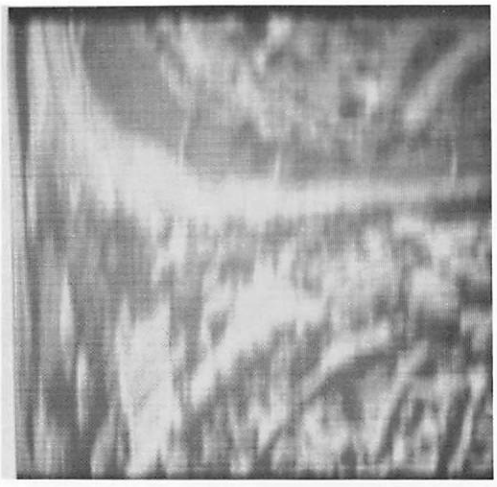


(d) restored image after inverse distortion

Figure 4.8-2. Example of image motion blur restoration - moonscape.

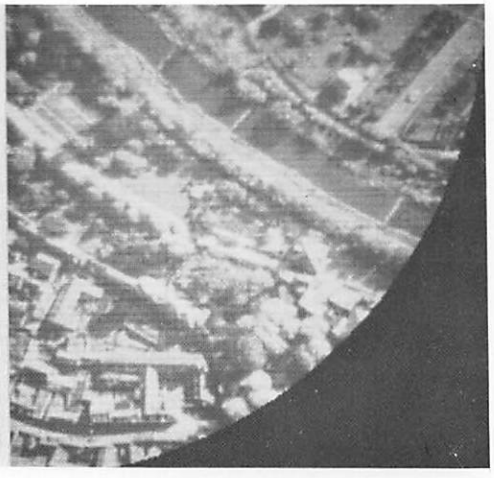


(a) images blurred by rotational motion



(b) image after polar coordinate distortion

(c) restoration by space-invariant inverse filter



(d) restored image after inverse distortion

Figure 4.8-3. Example of image motion blur restoration - river scene.

5. Image Data Extraction Projects

Image data extraction is a new name given to the collection of projects that are concerned with the detection of features within an image and methods of measuring these features. Allied techniques of image formatting, such as automatic spatial registration and rectification, necessary as a precursor to image data extraction, are also included in this section.

The first report describes a study of digital and optical techniques of texture discrimination. Texture, as defined in this context, is the local arrangement of pixel values that gives the appearance of a semi-repetitive pattern. The digital techniques studied involve several spatial, statistical measures of texture. A coherent optical system which measures texture in terms of the spatial frequency spectrum of an image section has also been studied. A successful application of both measures has been obtained in a particular application--the automated detection of coal miner's black lung disease by texture discrimination of chest radiographs.

The second project is an investigation of correlation techniques of image registration for pairs of images with unknown relative translation. Most classical solutions to this problem are based upon the cross-correlation measure between the images. This technique has been extended by performing a preprocessing operation to improve the sensitivity of the cross-correlation measure. This technique leads to an order of magnitude improvement in misregistration detection, and thus permits registration of independently noisy images.

5.1 Image Textural Discrimination

Richard P. Kruger

Rosenfeld and Troy [1] describe image texture ideally as the "repetitive arrangement of a unit pattern over a given area." They also state that in natural imagery it would be difficult to identify such unit

patterns or determine their repetitive arrangement. Therefore, the previous description should be used only as a guide in the analysis of natural imagery.

Since the approach to textural feature extraction will of necessity be experimental, some probable statistical indices will be explored. With statistical analysis it is hoped that useful measures of the organization and arrangement of the textures in question will be derived without having to focus attention on the specific structural properties. For instance, these statistical measurements will not attempt to recognize or trace textural structures per se, but merely compute quantitative impressions which will characterize them.

Textural discrimination of images types will be attempted in both the spatial and spatial frequency domain. The specific approach, therefore, is to initially consider a two class situation. The task is to experimentally compute several textural features, and subsequently apply a feature selection and supervised statistical pattern recognition technique to obtain a two class classification.

Data Management and Textural Feature Extraction from Digital Images Initial prototype images consist of the manual selection of image fields from two or more known classes. A monotonic transformation of gray levels is performed which will produce an image with 8 equally likely gray levels [2,3,4]. This preprocessing step is designed to negate the effect of additive and multiplicative constants introduced due to inconsistency in photographic and/or digitization of the original images by constraining all inputs to the feature extractor to be identical with respect to first order probability of gray level occurrences. The textural feature extraction measures are all based on spatial gray level dependence matrices [5,6] under the assumption that visual texture-context information in each image is contained in the spatial relationship between image picture elements at several fixed distances and angular orientations. More specifically, it shall be assumed that this texture-context information

is adequately specified by the symmetric matrix of relative frequencies p(i, j) with which two neighboring pixels are separated by a distance (d) and an angle (a) for each (i, j) gray level pair in the space. For this application, d is the number of image lines separating the two pixels of interest. Therefore, an 8 by 8 symmetric count matrix is formed for each selected prototype image and the count matrix is normalized to create a matrix of relative frequencies as a function of a and d. The parameter set is given by:

$$i = 0, 1, \ldots, 7$$

$$j = 0, 1, \ldots, 7$$

$$a = 0^{o}, 45^{o}, 90^{o}, 135^{o}$$

$$d = 1, 3, 7, 11$$

The following five textural measurements $T_k(a,d)$ $k = 1, \ldots, 5$ are computed for each matrix

$$T_1(a,d) = \sum_{i=0}^{7} \sum_{j=0}^{7} i \cdot j \; p(i,j,a,d) \tag{1}$$

$$T_2(a,d) = \sum_{i=0}^{7} \sum_{j=0}^{7} (i-j)^2 \; p(i,j,a,d) \tag{2}$$

$$T_3(a,d) = \sum_{i=0}^{7} \sum_{j=0}^{7} \frac{p(i,j,a,d)}{i+(i-j)^2} \tag{3}$$

$$T_4(a,d) = \sum_{i=0}^{7} \sum_{j=0}^{7} p(i,j,a,d) \log p(i,j,a,d) \tag{4}$$

$$T_5(a,d) = \sum_{i=0}^{7} \sum_{j=0}^{7} |i-j| \; p(i,j,a,d) \tag{5}$$

T_1 is an autocorrelation measure designed to meausre image coarseness. T_2 is a dissimilarity measure often called the moment of inertia. T_3 measures the extent to which the same or similar gray levels tend to be

neighbors. T_4 is a conditional entropy measure and measures image homogeneity. T_5 is another dissimilarity measure which is similar to T_2. A total of 80 textural measures are extracted from each prototype image with each textural feature a function of angle (a) and distance (d). The number of textural features was reduced from 80 to 60 by calculating the mean ($\bar{M}$) variance (V) and range (R) at a given distance d for each of four angles (a). These statistics are defined as

$$\bar{M}_k(d) = \frac{1}{4} \sum_{a=1}^{4} T_k(a,d) \tag{6}$$

$$R_k(d) = \max T_k(a,d) - \min T_k(a,d) \quad a = 0^o, 45^o, 90^o, 135^o \tag{7}$$

$$V_k(d) = \frac{1}{4} \sum_{a=1}^{4} (T_k(a,d) - \bar{M}_k(d))^2 \tag{8}$$

Fourier Transform Domain Feature Extraction Using a Coherent Optical Approach: The textural features extracted in the previous section were derived from digital spatial domain data. The Fourier domain measures to be presented treat each of the prototype images as an entity, and as such, measure more global aspects of visual texture. Figure 1 describes in simplest form, the Recognition Systems Inc. ROSA-3 used to extract the spatial frequency measures of visual texture. A helium-neon laser emits a light which passes through a collimating lens and then through the input film image. The transmitted light from the film next passes through a positive thin lens which performs the Fourier transformation. The transformed image is then projected onto a detector and appropriate energy measurements are obtained.

The Fourier transform equation is

$$\hat{F}(u,v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \exp\{j2\pi(xu+yv)\}\, dx\, dy \tag{9}$$

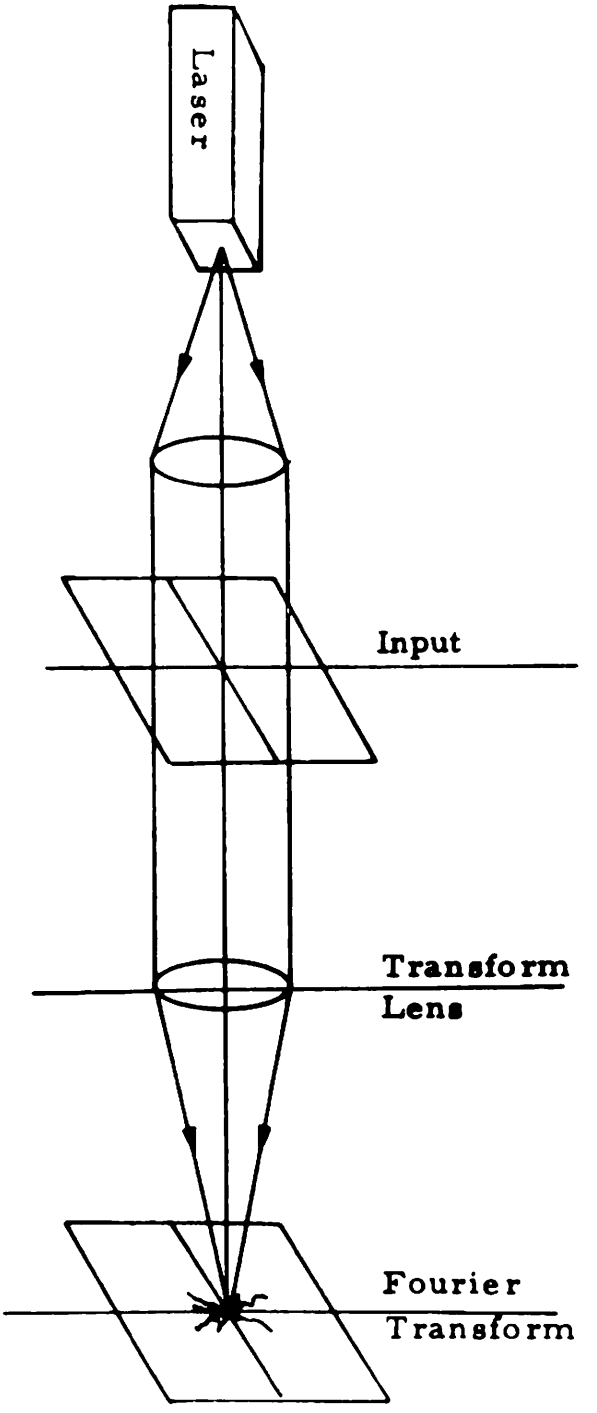


Figure 5.1-1. Optical feature extraction device.

where f(x,y) is the illuminated film region and F(u,v) is the transform of that region.

It is well known that high frequency information pertains to the amount of sharpness of edge information in an image. A detector in the transform plane shown in Fig. 2 which has 32 annular rings and 32 angular wedges, is used to obtain both the total contribution of 32 radial frequencies which would accurately reflect the amount of edge by the relative strength of the higher energy annulli, as well as any angular dependencies in the prototype images. A 20 wedge shaped band rejection filter in the detector serves as a read out path for the detector. The spectral measurements were normalized to unit energy by dividing each measurement by the total energy in the transform which in this case was the sum of the energy in all the annular rings. The normalized energies in the rings were then logarithmically transformed to create distributions which were more nearly Gaussian.

Textural Feature Selection and Classification Approach: Both sets of textural feature measurements will contain both redundancy and features which are of little value in separating the classes. For a classifier to work successfully, these features must be removed. In addition, the larger the set of measurements the classifier must deal with, the greater the numerical inaccuracy in computing discriminant functions will be. Thus, it is advantageous to make the feature space as low dimensional as possible, determining which of the original measurements contain the most useful information for the classifier.

In order to select meaningful features, a measure of the value of a feature must be defined. For a two class problem where the classifier assumes a Gaussian distribution of features, the "Divergence" [7] is such a measure. The divergence of a set of features is defined as

$$J(S_1,S_2) = \int_{-\infty}^{\infty} [\, p(\bar{x})|S_1) - p(\bar{x}\,|S_2)] \ln \left[\frac{p(\bar{x}\,|S_1)}{p(\bar{x})|S_2)} \right] d\bar{x} \qquad (10)$$

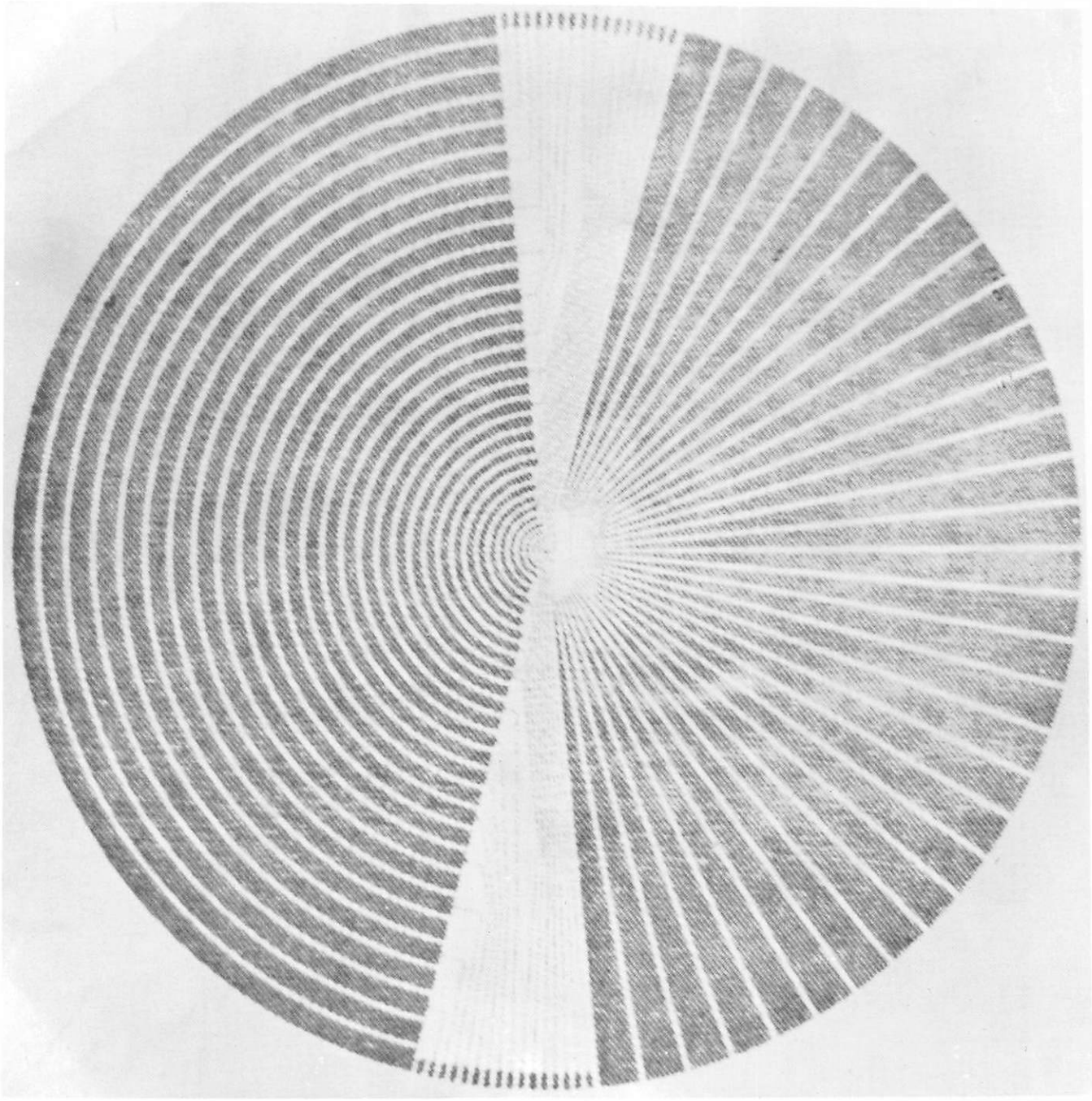


Figure 5.1-2. Ring and wedge Fourier transform plane energy detector.

where $S_1 \rightarrow$ class 1, $S_2 \rightarrow$ class 2.

If classes S_1 and S_2 are assumed multivariate Gaussian distributed, the R dimensional $\bar{x}$ vector is distributed

$$p(\bar{x}/S_k) = \frac{\exp\left[-\frac{1}{2}(\bar{x}-\mu_k)^t[\Phi_k]^{-1}(\bar{x}-\mu_k)\right]}{(2\pi)^{R/2}|[\Phi_k]|^{\frac{1}{2}}} \qquad (12)$$

$$k = 1,2$$

where μ_k is a mean vector and $[\Phi_k]$ is the covariance matrix for each class. The divergence distance is optimized over all linear transforms $[T]$, of dimension N by R, which implies Gaussian distributions in the transformed space as well. Thus

$$p(\bar{x}/S_k[T]) \sim N(\mu_k^T, [\Phi_k^T]) \qquad k = 1,2 \qquad (13)$$

$$\mu_k^T = [T]\mu_k\,[\Phi_k^T] = [T]^t[\Phi_k][T]$$

Therefore, the divergence measure becomes a function of the linear transformation and will be denoted $J(S_1,S_2,T)$. The divergence can therefore be expressed as

$$J(S_1,S_2,T) = \tfrac{1}{2}\,\mathrm{tr}\,([\Phi_2^T]^{-1}[\Phi_1^T] + [\Phi_1^T]^{-1}[\Phi_2^T] - 2[I]) +$$

$$\tfrac{1}{2}\,\mathrm{tr}\,(([\Phi_1^T]^{-1} + [\Phi_2^T]^{-1})((\mu_1^T - \mu_2^T)(\mu_1^T - \mu_2^T)^t)) \qquad (14)$$

where tr is the trace of the matrix.

It would be desirable to find which set of features, taken together, would be optimal. However, in practice this is not possible. A compromise is to calculate the divergence of R features one at a time and then to choose N of those with the highest divergence value. If Gaussian statistics

are assumed and if features are analyzed one at a time, the divergence measure of the ith feature becomes

$$J(S_1,S_2,i) = \frac{\left(\sigma_i^{(1)} - \sigma_i^{(2)}\right)^2 + \left(\sigma_i^{(1)} + \sigma_i^{(2)}\right)\left(\mu_i^{(1)} - \mu_i^{(2)}\right)^2}{2\,\sigma_i^{(1)}\,\sigma_i^{(2)}} \qquad (15)$$

where $\sigma_i^{(k)}$ and $\mu_i^{(k)}$ are the variance and mean of feature i for class k and [T] is a 1×R matrix that has one non-zero term of unity value at location (1,i). In this manner an R dimensional pattern space is reduced to an N dimensional feature space.

For the extracted features it is often discovered that textural feature measures for the several prototype classes exhibit a large variation. This seems to preclude a distribution free classification approach and has led to a statistical approach. Statistical classifiers make assumptions about the underlying distributions of features. The most common assumption made about feature statistics is that they are multivariate Gaussian. If a classifier is desired which maximizes the likelihod of correct classification, then the discriminant functions become

$$g_k(\bar{x}) = -\tfrac{1}{2}\,\bar{x}^T[\varphi_k]^{-1}\,\bar{x} + \bar{x}^T[\varphi_k]^{-1}\mu(k) - \tfrac{1}{2}\,\mu(k)^T[\varphi_k]^{-1}\mu(k)$$

$$+ \ell n\, P(S_k) - \tfrac{1}{2}\,\ell n\,(\det[\varphi_k]) \qquad k = 1,2 \qquad (16)$$

where φ_k is the N×N covariance matrix of the feature distributions and $P(S_k)$ is the a priori probability of a sample being from diagnostic class k.

Automated Diagnosis of Abnormal Lung Fields from the Routine Chest Radiograph The digital and optical textural measures described above have been applied to a problem of automatically detecting abnormal lung vascularity resulting from coal workers pneumoconeosis (black lung disease).

The features selected by the divergence measure for the digitally derived features were $R_4(3)$, $\bar{M}_3(1)$, $V_4(3)$, $\bar{M}_5(3)$, $V_5(3)$ and $\bar{M}_5(1)$ in order of preference. This yielded a feature vector length of $N = 6$. A separate analysis on the optical features selected annular rings 5, 22, 7, and 14 in order of preference yielding a $N = 4$ dimensional feature vector.

Computer Classification Results The computer diagnostic procedure consisted of a removing one sample from the data base, training on the remaining samples and resubmitting the withdrawn sample for reclassification. This is a fair test since the classifier does not "see" the withdrawn sample until it is asked to diagnostically assign it to a class. A second more severe test was also performed. This test consisted of removing one-half of the data from each class and training on the remaining data. The removed half was then submitted to the classifier for diagnosis. This was repeated twice so that all data was classified in a test situation. In many respects these two testing procedures are logical extremes. In the first test only one sample is withdrawn and as such the test must be repeated 141 times for the optical and 298 times for the digital data bases, respectively. The second test is only repeated twice for each of the data bases. The problem of manually detecting and grading simple pneumoconioses opacities for radiographs appears to be largely one of discrimination between normal pulmonary vascularity (lung markings) pattern and partial or complete obliteration of this normal tree like pattern by opacities of various sizes and profusions which themselves exhibit a more or less textural nature. The data base consisted of 141 such delineated lung zones from all six lung zones. Figure 3 contains a photograph illustrating computer outlined lung zones.

Data Management and Textural Feature Extraction from the Digital Images It was decided that visual diagnosis of simple pneumoconiosis lesions is often arrived at by inspecting the lung regions between the more visually prominent posterior ribs. This seemed logical since it is in these inter-rib spaces that normal vascularity is least obstructed by visual

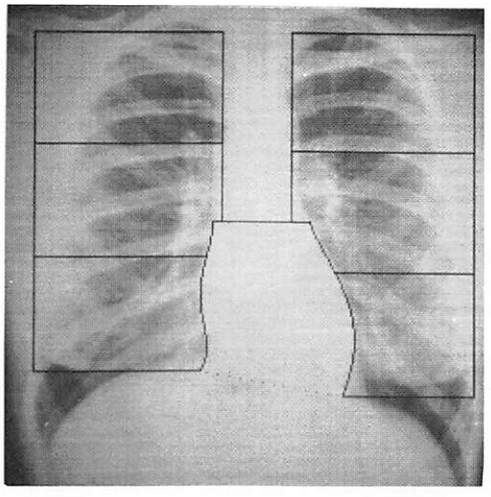


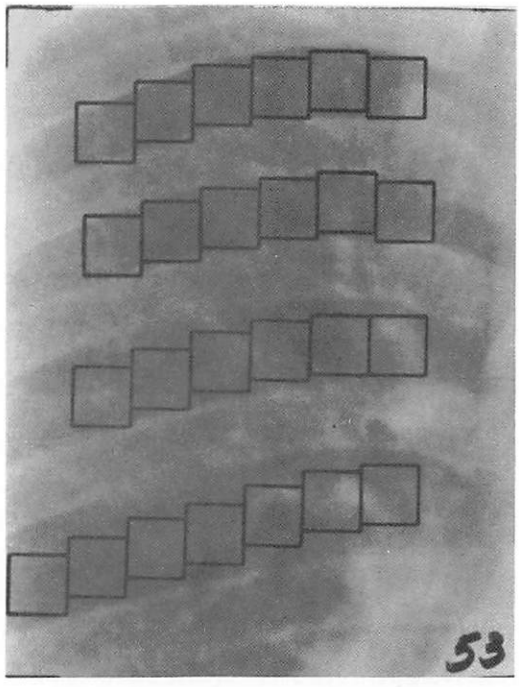
Figure 5.1-3. Computer outlined cardiac projections and lung zones.

interference from more dense radiographic structures. These posterior inter-rib spaces were therefore manually extracted in the following manner. First a computer generated grid was superimposed upon each of the digitized images. This grid allowed visual extraction of from 4 to 10, 100 by 100 pixel squares from each inter-rib space. These squares are delinated for the zonal film shown in Figure 4a. A p(i,j) matrix was found for each inter-rib space. Since there was 3 to 4 such spaces per lung zone, the complete digital data base consisted of textural measurements from 298 inter-rib spaces. The optical data base consisted of the spectral measurements from an illuminated circular 2.5 inch aperture in each of 141 lung zone films. A typical aperture illuminated region is shown in Figure 4b.

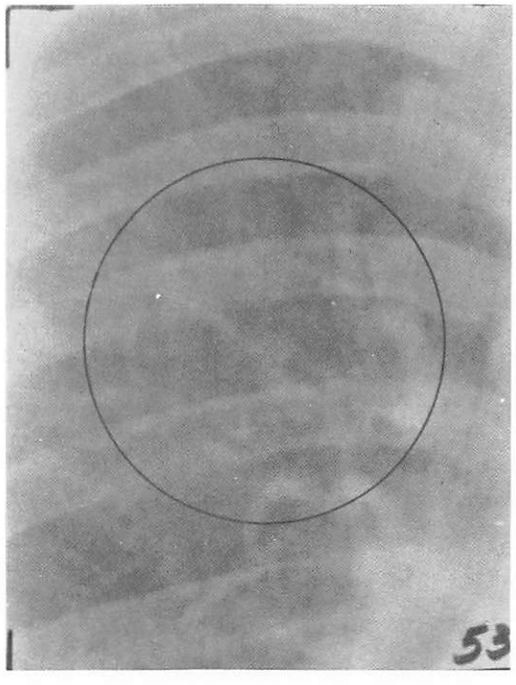
It was decided to consider each posterior inter-rib space as a separate input to the diagnostic classifier to create an automated classification technique which would not be sensitive to any inter-rib space nor any specific lung zone and would constitute a local region for textural analysis.

The normal-abnormal training diagnostic rate was computed to be 95.9%. It should also be noted that the false positive rate 5.3% versus a false negative rate is only 3.6%. This, of course, is the conservative medical diagnosis for a mass screening situation. On a per film basis only 2 films in 95 were missed. The one at a time removal test procedure yielded a normal-abnormal diagnostic rate of 95.2%. On a per-film basis 3 films were missed for a corresponding diagnostic rate of 96.9%. When the more severe second test was performed the normal-abnormal rate was 92.9%. On a per-film basis this was 96.8%.

One quite obvious conclusion is that the normal-abnormal diagnostic rate is quite stable, using digitally derived textural features. The two class transform domain results will now be presented. A corresponding digital rate will be given for the 95 films common between the two film bases. The normal-abnormal training diagnostic rate was 93.6%. When



(a) zonal film with inter-rib squares inscribed



(b) zonal film with laser aperture super-imposed in the input plane

Figure 5.1-4. Zonal films.

only films common between the digital and optical data bases are examined, the normal-abnormal diagnostic rate is 97% representing 3 misses in 95 films. The one at a time removal test procedure yielded the normal-abnormal rate of 90.8%. When a common data base with digital data is compared, 5 films were missed for a diagnostic normal-abnormal rate of 95%. The more severe test procedure yielded the normal-abnormal diagnostic rate of 88.7%. Using the data base of the digital measurements indicated 5 misses out of 95 for a normal-abnormal rate of 94.8%.

Physician Diagnosis Six radiologists were requested to diagnose the identical 141 lung regions submitted for automatic analysis. Two of the six readers originally selected the films in this study. As a group these six radiologists represent over 130 man years of radiological reading experience.

The physicians used the entire P-A radiograph with appropriate rectangular lung zones within which they were to make their diagnosis labelled and numbered. This allowed the readers to grade the films within an anatomical context.

When all the $6 \times 141 = 846$ physician observations are averaged the normal-abnormal rate was 93.4%. In summary, the false positive rates for the six readers ranged from 95.0% to 2.9% with an average of 17.5%. Correspondingly the false negative rate ranged from 1.0% to 6.9% with an average of 2.6%. The averaged physician rates showed no significant change when computed on the basis of the 95 films submitted for digital analysis.

Conclusions A study was undertaken to determine the feasibility of using textural measures for the possible automated mass diagnostic screening of pneumoconiosis radiographs. Two distinct textural feature extraction methods involving digital and coherent optical approaches were undertaken. The performance of the two automated diagnostic systems was determined and analogous results were obtained for diagnosis obtained from experienced radiologists asked to analyze the same films given to the

automated systems.

At the present time a study is underway to apply the previously described methods to the analysis of aerial photography. Initial study will center about automatic discrimination between man-made and naturally occuring terrain types.

References

1. Rosenfeld, A., Troy, E., "Visual Texture Analysis," Proceedings of the UMR-Kelly Communications Conference, Rolla, Mo., October, 1970.

2. Andrews, H. C., Tescher, A. G., and Kruger, R. P., "Image Processing by Digital Computer," IEEE Spectrum, Vol. 9, No. 7, July 1972.

3. Rosenfeld, A., Troy, E., "Visual Texture Analysis," Technical Report 70-116, June 1970, University of Maryland.

4. Haralick, R., Anderson, D., "Textural-Tone Study with Applications to Digitized Imagery," Technical Report 182-2, University of Kansas, November, 1971.

5. Rosenfeld, op. cit., 1.

6. Haralick, op. cit., 4.

7. Andrews, H. C., Introduction to Mathematical Techniques in Pattern Recognition, John Wiley and Sons, Inc., New York, 1962.

5.2 Correlation Techniques of Image Registration

William K. Pratt

In many image processing applications it is necessary to form a pixel-by-pixel comparison of two images of the same object field obtained from different sensors, or of two images of an object field taken from the same sensor at different times. To form this comparison it is necessary to spatially register the images and thereby correct for relative translational shifts, magnification differences, and rotational shifts, as well as

geometrical and intensity distortions of each image. Often it is possible to eliminate or minimize many of these sources of misregistration by proper static calibration and compensation of the image sensor; in some applications misregistration detection and subsequent correction must be performed dynamically for each pair of images.

Consideration is given here to the single problem of registering images subject to translational differences. The results can be applied to the detection of rotational and magnification differences by increasing the dimensionality of the problem, or by a proper transformation of coordinates (e.g., a rotational shift is equivalent to a translational shift in polar coordinates).

A classical technique for registering a pair of functions is to form a correlation measure between the functions and determine the location of the maximum correlation. In applying this technique to two dimensions, let $f_1(j,k)$ and $f_2(j,k)$ represent two discrete images to be registered. In its simplest form the correlation measure is defined as

$$R(u,v) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} f_1(j,k) f_2(j-u,k-v)}{\left[\sum_{j=1}^{J}\sum_{k=1}^{K} f_1^2(j,k)\right]^{1/2}\left[\sum_{j=1}^{J}\sum_{k=1}^{K} f_2^2(j-u,k-v)\right]^{1/2}}$$

where (j,k) are indices in a J x K point window area, W, that is located within an MxN point search area, S. Figure 1 illustrates the relationship between the search and window areas. In general, the correlation function R(u,v) must be computed for all (M-J+1)(N-K+1) possible translations of the window area within the search area to determine its maximum value and obtain a mixregistration estimate.

There are two basic problems with this simple correlation measure. First, the correlation function may be rather broad, making detection of the peak difficult. It should be noted that the simple correlation measure

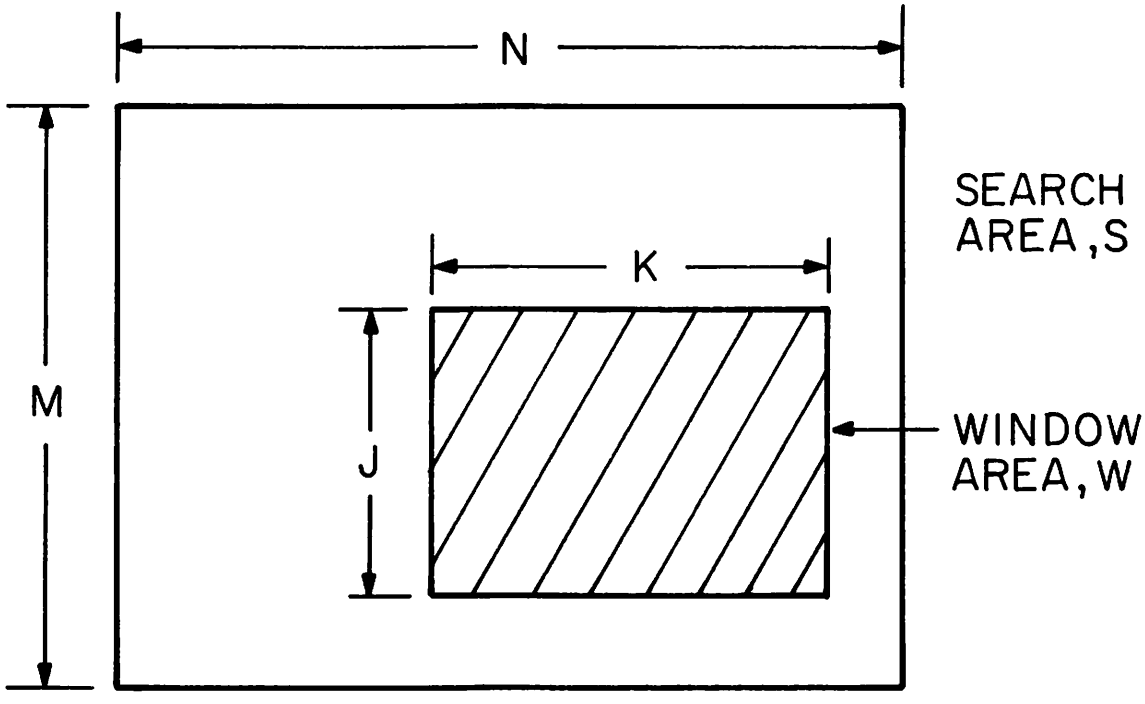


Figure 5.2-1. Relationship between search and window areas.

ignores the spatial relationship of points within each image. Second, image noise may mask the peak correlation. Both problems can be alleviated by extending the correlation measure to consider the statistical properties of each image function $f_1(j,k)$ and $f_2(j,k)$. The statistical correlation measure is defined as

$$R_S(u,v) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} g_1(j,k)\,g_2(j-u,k-v)}{\left[\sum_{j=1}^{J}\sum_{k=1}^{K} g_1^2(j,k)\right]^{1/2}\left[\sum_{j=1}^{M}\sum_{k=1}^{N} g_2^2(j-u,k-v)\right]^{1/2}}$$

where the $g_i(j,k)$ are obtained by spatially convolving the sampled images $f_i(j,k)$ with spatial filter functions $D_i(j,k)$. Thus,

$$g_i(j,k) = f_i(j,k) \circledast D_i(j,k)$$

The spatial filter function is chosen to maximize the correlation peak ratio

$$C_P = \frac{R_S(\delta_x,\delta_y)}{R_S(u;v)} \qquad \begin{array}{l}\text{for all } u \neq \delta_x \\ \text{for all } v \neq \delta_y\end{array}$$

Determination of the optimum spatial filter function is facilitated by a vector space representation of each image. Let the column vector $\underline{Q}$ represent the image function $f_1(j,k)$ when the image is scanned in a vertical raster fashion. Similarly, let $\underline{P}_{u,v}$ represent the column scanned image $f_2(j+u,k+v)$. The elements of $\underline{Q}$ and $\underline{P}_{u,v}$ will be highly correlated spatially since $f_1(j,k)$ and $f_2(j,k)$ are each spatially correlated to a significant extent for natural imagery. The first step in the spatial filter design process is to decorrelate or "whiten" each image vector by whitening filter matrices $\underline{H}_Q$ and $\underline{H}_P$. Thus, let

$$\underline{A} = (\underline{H}_Q)^{-1} \underline{Q}$$

$$\underline{B}_{u,v} = (\underline{H}_P)^{-1} \underline{P}_{u,v}$$

where $\underline{H}_Q$ and $\underline{H}_P$ are obtained by a factorization of the image covariance matrices

$$\underline{K}_Q = \underline{H}_Q \underline{H}_Q^T$$

$$\underline{K}_P = \underline{H}_P \underline{H}_P^T$$

The simple correlation operation is now performed on the whitened vectors $\underline{A}$ and $\underline{B}$ yielding the statistical correlation measure

$$R_S(u,v) = \frac{\underline{A}^T \underline{B}_{u,v}}{\left(\underline{A}^T \underline{A}\right)^{1/2} \left(\underline{B}_{u,v}^T \underline{B}_{u,v}\right)^{1/2}}$$

which can be written as

$$R_S(u,v) = \frac{\left(\underline{K}^T\right)^{-1} \underline{Q}^T \underline{P}_{u,v}}{\left(\left(\underline{K}^T\right)^{-1} \underline{Q}^T \left(K^T\right)^{-1} \underline{Q}\right)^{1/2} \left(\underline{P}_{u,v}^T \underline{P}_{u,v}\right)^{1/2}}$$

where

$$\underline{K} \equiv \underline{H}_P \underline{H}_Q^T$$

In this formulation the statistical correlation measure is obtained by filtering one of the image vectors $\underline{Q}$ a single time with a filter matrix

$$\underline{G} = \left(\underline{K}^T\right)^{-1}$$

and then evaluating the simple correlation measure between the filtered image and the other image for each potential misregistration.

In order to assess the performance of the statistical correlation measure, a computer simulation was performed to determine the misregistration of an image with itself. In the simulation, a window of 16 x 16 pixels was employed in a 32 x 32 pixel search area. The image correlation matrix was assumed to be of Markov form with an adjacent pixel correlation, ρ, ranging from 0 to 1. In all cases the misregistration was detected. Figure 2 contains a plot of the measured correlation function for a vertical shift of four pixels for several values of the parameter ρ. With $\rho = 0$, the statistical correlation measure reduces to the simple correlation measure. The relatively small peak of the correlation function is apparent from the figure in this case. As ρ increases toward unity, the correlation peak becomes more pronounced, and permits detection of the misregistration with greater accuracy.

The operational performance of the statistical correlation system can be justified by a heuristic argument: If the images to be registered are each highly correlated, then they will each contain large areas of nearly constant brightness. These large common areas will give a high correlation value for all amounts of misregistration, and hence mask the true correlation peak. By prefiltering the original images, their edges will be enhanced with respect to the backgrounds. Then, the edges of the two images are correlated to obtain a more pronounced measure of the misregistration.

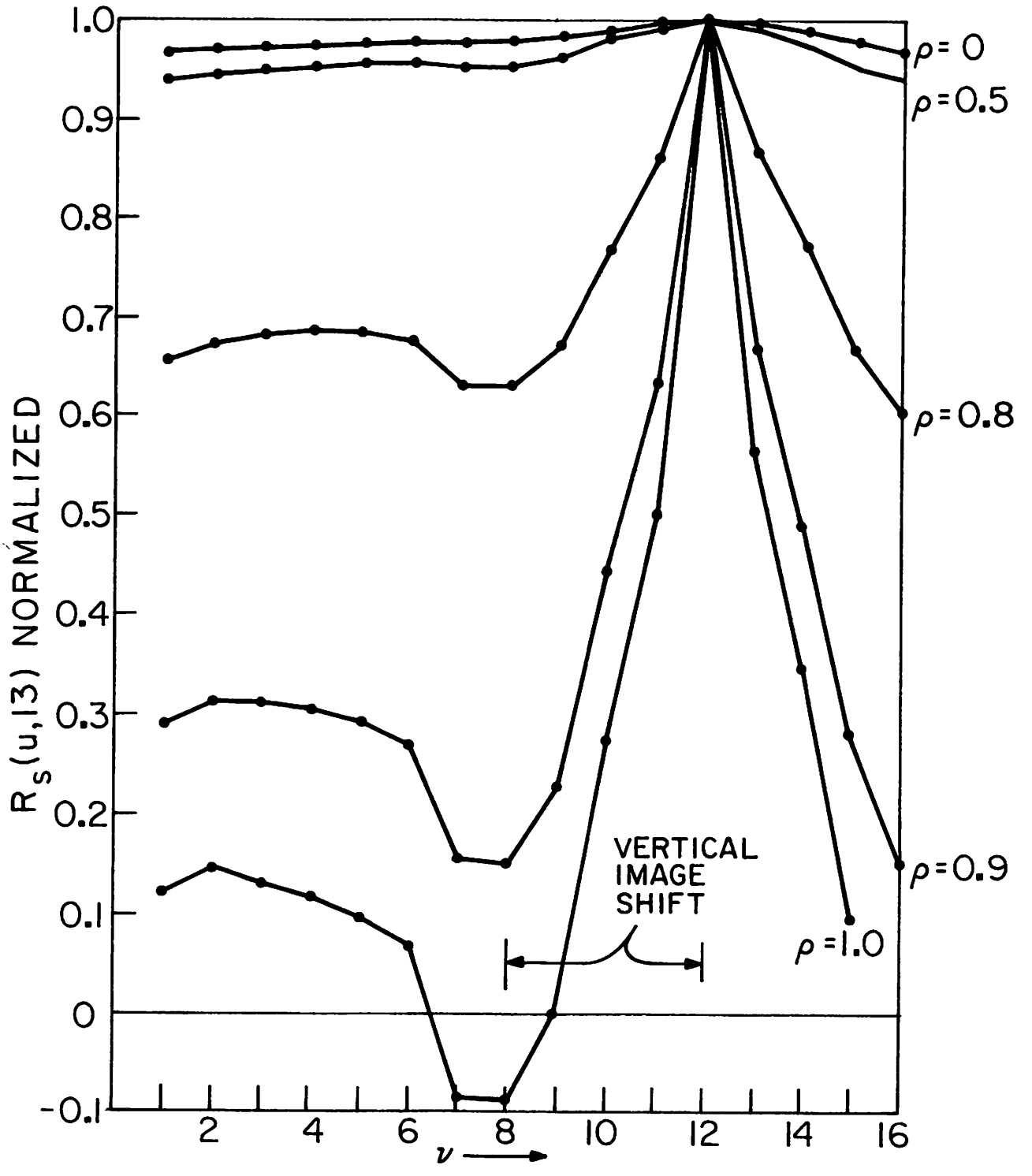


Figure 5.2-2. Simulation results of correlation measure.

6. Image Analysis Projects

The image analysis projects are concerned with the background technology necessary to effectively design image coding, restoration and enhancement, and data extraction systems. Of particular interest is the development of quantitative measures of image fidelity and intelligibility. Other projects included in this classification are studies directed toward the characterization of images and image acquisition and display devices.

In the first report further advances in the development of a geodesic color measure are discussed. The color measure under consideration yields a quantitative metric of color distance between two colors with known luminances and chromaticity coordinates. The present research effort has been directed toward the evaluation of the color measure as a means of accurate color difference.

The next section describes the application of Fourier transform techniques for the performance analysis of an image scanner. These techniques have been found useful in isolating noise and interference processes present in the scanner system. Also, the analysis gives an indication of methods that can be employed for restoration and enhancement.

The final report represents a preliminary discussion of a technique of representing images and other forms of two dimensional fields. This representation offers promise of a significant dimensionality reduction that can be exploited for image coding and restoration processes.

6.1 Color Measures in Verification of Schrodingers Theory of Color Vision

A. K. Jain

In many situations such as scene analysis, color image restoration, image evaluation, etc., where a visual model for color preception or color discrimination is utilized, the role of a color difference measure becomes

very important. Recently, some efforts have been made to compare various color difference formulas on the basis of their relative agreements with observed data. Results of these comparisons indicate that the FMC1 (Friele-MacAdam-Chickering No. 1) formula is superior to the other formulas recommended by the CIE. This section presents a discussion of this formula in verification of Schrodinger's concept [1,2] in the theory of color vision. The success of this verification could be taken as an indication of the performance of the FMCI formula.

Geodesics and Schrodingers' Criterion Given two colors with coordinates x_i and $x_i + dx_i$, i = 1,2,3 the FMCI formula describes the color difference metric as

$$[ds^2] = \sum_{i=1}^{3} \sum_{j=1}^{3} C_{ij}\, dx_i\, dx_j , \qquad (1)$$

where x_i, i=1,2,3 are linearly related to the X, Y, Z color coordinates and C_{ij} are the sensitivity coefficients of average human perception of colors and depend on x_i, i=1,2,3. Using Eq. (1) as the color difference metric, the color difference between any two arbitrary colors (c_1 and c_2) is given by [4],

$$s = \min_{x_i} \int_{c_1}^{c_2} ds \qquad (2)$$

In other words, the color distance between the colors c_1 and c_2 is equal to the distance measured according to Eq. (1), along the curves of least distance between these colors. This set of curves of least distance $x_i(t)$, (i=1,2,3, t = parameter along the curves) is called the geodesic between c_1 and c_2. If $(\bar{x}, \bar{y}, \bar{Y})$ and $(\hat{x}, \hat{y}, \hat{Y})$ represent the chromaticity coordinates (x, y) and luminances (Y) of the colors c_1 and c_2 respectively, then according to Schrodinger's theory, the color c_2 should appear equally as bright as c_1 if for fixed $\bar{x}, \bar{y}, \bar{Y}$, $\hat{x}$ and $\hat{y}$, the luminance $\hat{Y}$ is such that the

color distance between c_1 and c_2 is a minimum. This means that for the fixed color c_1 and fixed chromaticities (x, y) in the x, y, Y space, Y is such that the geodesic between c_1 and c_2 has minimum color distance. Mathematically, this implies

$$s(\bar{x}, \bar{y}, \bar{Y}; \hat{x}, \hat{y}, \hat{Y}) = \min_{\substack{\alpha \\ \alpha \in \mathcal{Y}}} \{ s(x, y, Y; \hat{x}, \hat{y}, \alpha) \} \tag{3}$$

where $\mathcal{Y}$ is the set of all possible luminance values in the x, y, Y space. Figure 1a illustrates this meaning. The broken lines show the various geodesics from c_1 to various colors on the constant chromaticity line through (x, y). The solid line is the geodesic satisfying Eq. (3). This is also called the constant brightness geodesic. Clearly, for any fixed c_1 and any chromaticity value (x, y), there is a unique constant brightness geodesic. The loci of all these geodesics is called the constant brightness surface. Based on Schrodingers' theory described above, the following observations can be made:

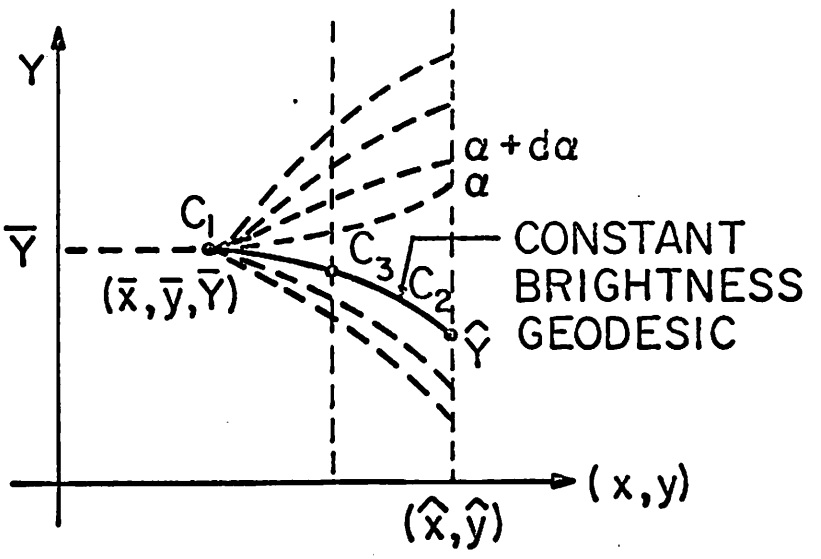
(a) Any two colors on a constant brightness geodesic are colors of equal brightness;

(b) All constant brightness geodesics are horizontal at white (or any gray).

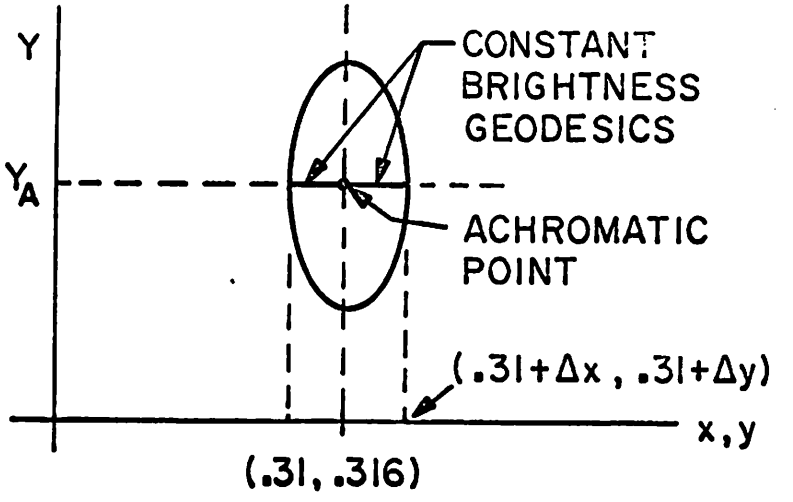
The first observation can be justified by invoking Bellman's Principle of Optimality and the following conclusion can be made:

<u>Any two colors on the surface of constant brightness are colors of equal brightness (according to Schrodinger's criterion) and the geodesic connecting them is a constant brightness geodesic.</u>

This can be used to verify Schrodinger's theory via the FMCI formula. The second observation can be justified by applying the Shcrodinger criterion to the color difference metric, and the fact that any color



(a) Geodesics between a fixed color c_1 and a constant chromaticity line. Broken lines show a geodesic with arbitrary value of Y on the constant chromaticity line. Solid line satisfies Schrodingers' criterion.



(b) Equi-color difference ellipsoid at a gray.

Figure 6.1-1. Color geodesic diagrams.

difference metric (including FMC1) described by an equation such as (1) shows that the equi-color distance infinitesimal ellipsoids with center on white or any gray are vertical with no tilt in the x, y, Y space (Figure 1b). For details see [1]. The slope of the constant brightness geodesics at any other point depends on the tilt of the equi-color difference ellipsoid at that point.

Implementation Figure 2 shows the projection of the constant brightness geodesics (generated between the achromatic point Y = 50 and the spectral colors) on the x, y plane. These geodesics map the constant brightness surface, and according to Schrodinger's criterion developed in the last section, if a geodesic is generated between any two cross spectrum colors on this surface, then their intersection points with the geodesics in Figure 2 should have the same luminance (Y) values. Figure 3 shows the cross spectrum geodesics with the intersecting constant brightness geodesics.

Conclusions Based on the results obtained in the study reported here it can be concluded that:

(a) Schrodinger's criterion of constant brightness colors is valid;

(b) FMC1 formula verifies this criterion uniformly, nearly everywhere within the color solid.

There are a few intersection points which give rise to relatively high values of ϵ. These should be re-examined by generating constant brightness surfaces with higher accuracy of the Y values at the end points. It is possible to determine analytically the end conditions for the constant brightness geodesics, and these conditions should be used in solving eq. (3) for more accurate results. This aspect is being studied now and the results will be reported in the near future.

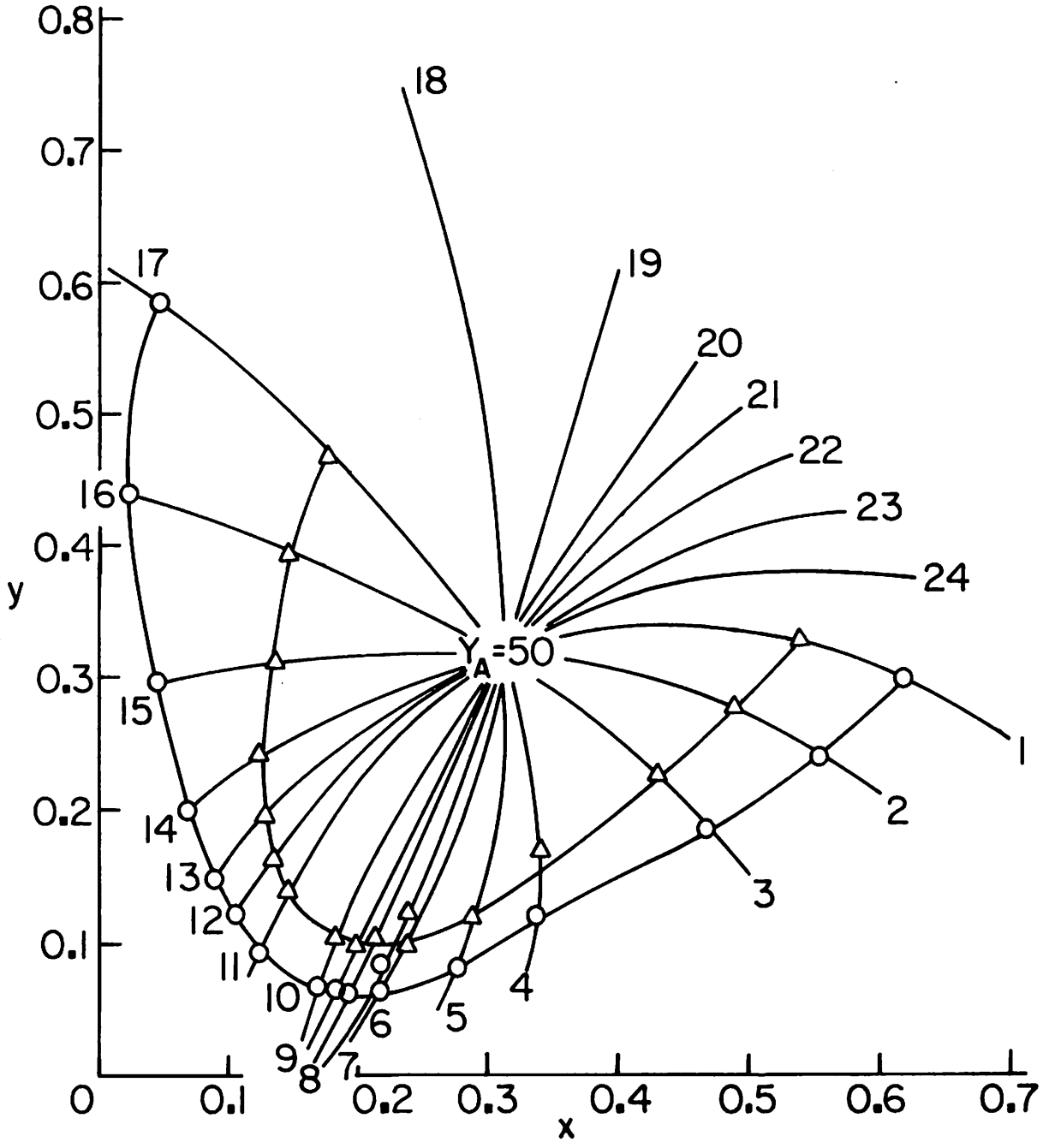


Figure 6.1-2. Projection on the x, y plane of the constant brightness surface generated by FMC1 formula. Points marked Δ have Y = 45, points marked O have Y = 40.

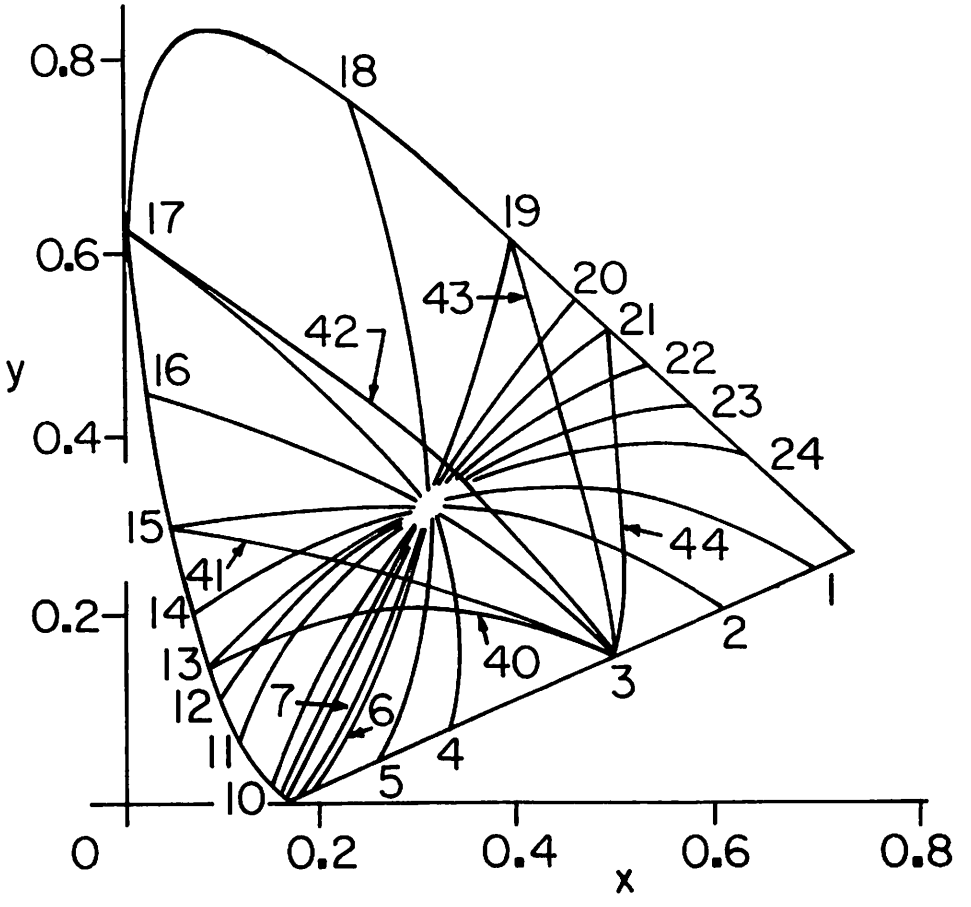


Figure 6.1-3. Intersection points of the cross spectrum geodesics with the constant brightness geodesics of Figure 3 in x, y plane.

References

This study was suggested by Dr. D. L. MacAdam of Kodak Research Labs., Rochester, New York.

1. Anil K. Jain, Role of Geodesics in Schrodingers' Theory of Color Vision, (submitted for publication).

2. D. L. MacAdam, "Role of Luminance Increments in Small Color Difference," in Color Metrics, edited by J. J. Vos, L.F.C. Friele, and P. L. Walravon, (AIC/Holland, c/o Institute of Perception TNO, Soesterberg, Netherlands, 1972), p. 160.

3. Anil K. Jain, "Color Distance and Geodesics in Color - 3 Space," J. Opt. Soc. Am. 62, 1267 (1972).

6.2 Scanner Evaluation Using the Fourier Domain

Harry C. Andrews

In late 1972, a magnetic tape of digitized images was supplied to USC IPI for purposes of providing the university community with "real world" data for picture processing purposes. This section of this report presents some preliminary results utilizing the supplied images. Figure 1 presents the block diagram of the processing on three sections of the view of an aircraft. The three images used are: a) "Blurred wing section"; b) "Center fuselage section", and; c) "Entire Aircraft". The processing implemented is described in each box and the circled numbers refer to the particular photographic image in Figures 2 - 4. The experiments implemented were selected to demonstrate a scanner evaluation capability, and were not necessarily optimized for a specific restoration or enhancement objective. The labeling, scaling, clipping, etc., routines are standard preprocessing procedures often utilized for imagery work. Figure 2 presents the results of the processing on the "Blurred Wing Section". The original image presented in Figure 2a appears to contain some scanner synchronization artifacts. In the processing sequence the original is framed, selectively clipped (for removal of background noise),

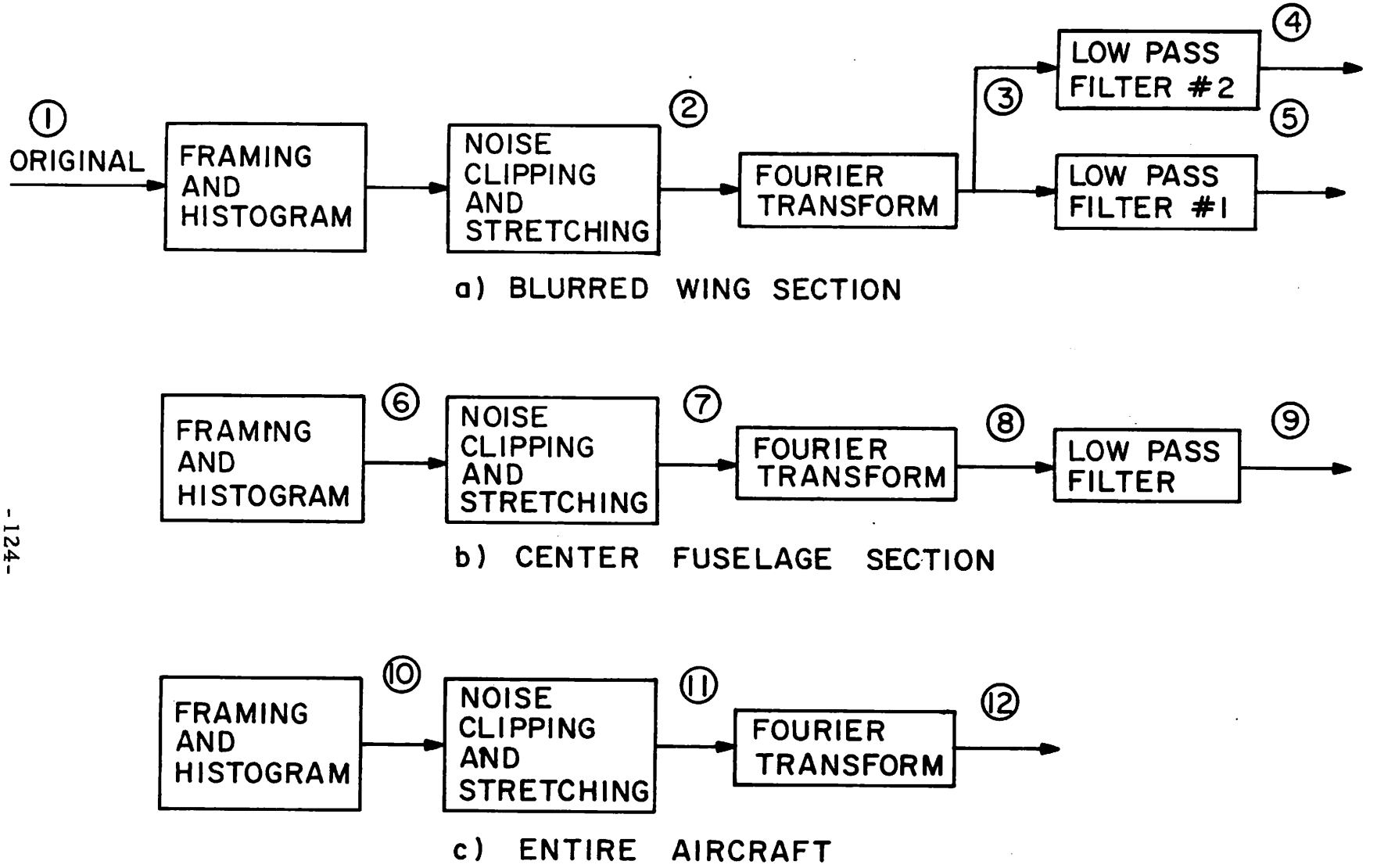
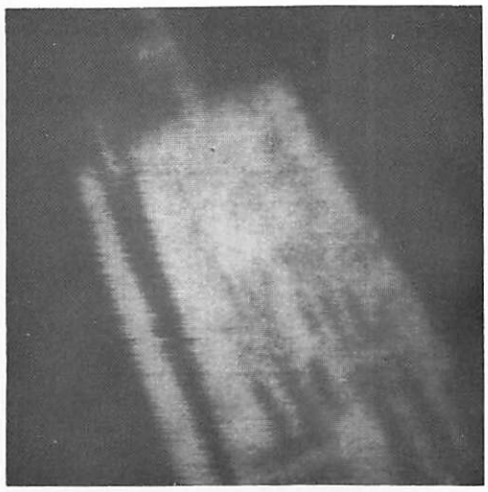


Figure 6.2-1. YOUTWO block diagrams.

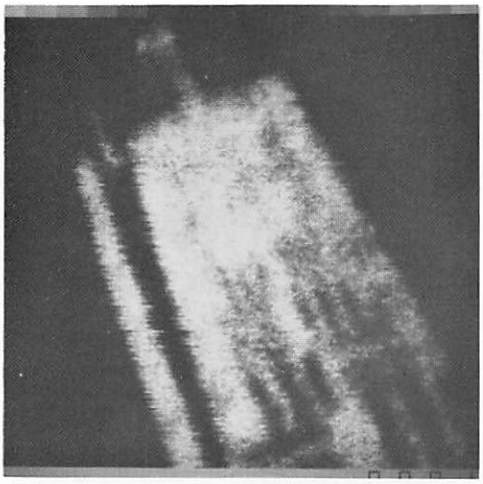
labeled 0001 and presented in Figure 2b. The Fourier transform and various low pass versions of the "Blurred Wing Section" illustrate high frequency noise removal. Figure 3 shows the results of the processing concentrating on the "Center fuselage section" with noise clipping in this case, providing a higher contrast image. Again, the Fourier and a low pass image are present. Figure 4 shows the three images associated with the "Entire Aircraft" scene. Here noise clipping removes the interference in the scanner to a certain extent.

The Fourier domain of all three images demonstrate two interesting phenomena concerning the device used for scanning. First, there is quite a bit of image energy along the horizontal axis in the Fourier domain. There are only three means by which such energy can consistently fall on axis in the Fourier plane. One way for energy on either the vertical or horizontal axis to occur is from actual imagery perfect aligned with the scanning device (highly doubtful). A second way for the Fourier axis to have so much energy is for one of the borders of the image to be brighter than its opposite border. Thus, in the "Center fuselage section" the right and left borders of the window have different imagery present. This accounts for the faint vertical line in picutre 1003. Notice that such a line is noticeably absent in picture 0004 and 2002 because right and left window borders in these images are essentially the same. The same phenomena will occur for top and bottom window borders but is masked in these cases by all the high energy falling on the horizontal axis of the Fourier domain of all three images. This axis is displaying energy exactly normal to the scan axis in the original image. Because there is jitter on the sync signal for the start of each line of the scanner, artificial line shifts are introduced thereby introducing artificial edges which are perfectly aligned normal to the scanning axis. Of course, such artificial structure should be removed if possible by a selective filter in the Fourier domain.

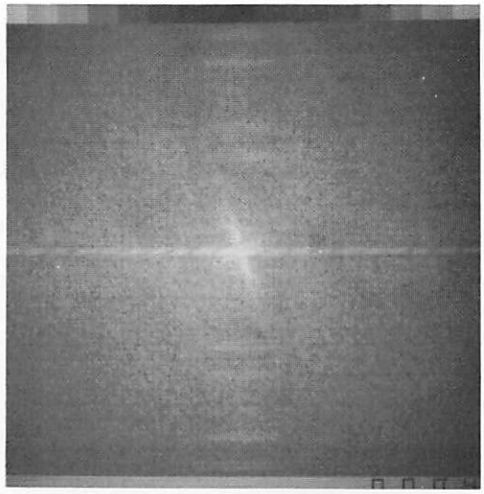
The second unanswered phenomena displayed by the Fourier images is the periodic faint horizontal stripes that occur along both horizontal and



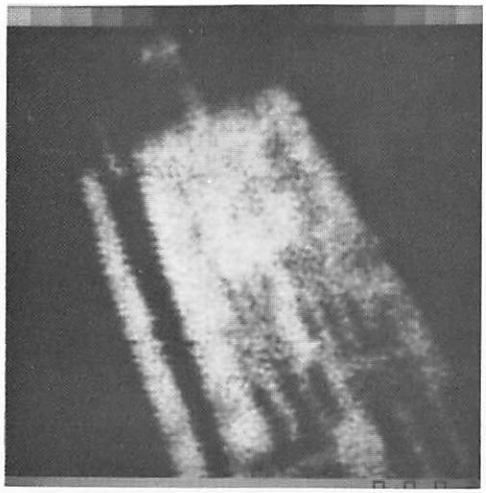
(a) original - ①



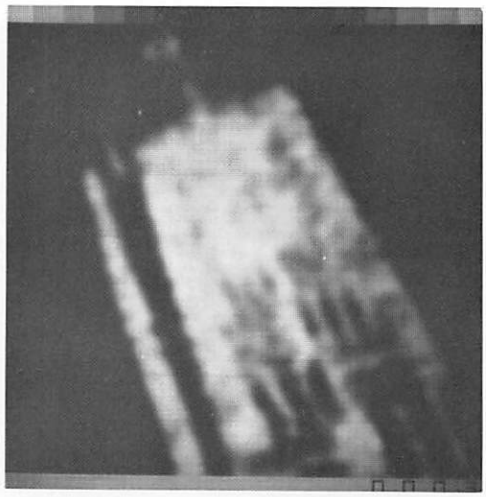
(b) noise clipped - ②



(c) Fourier transform - ③

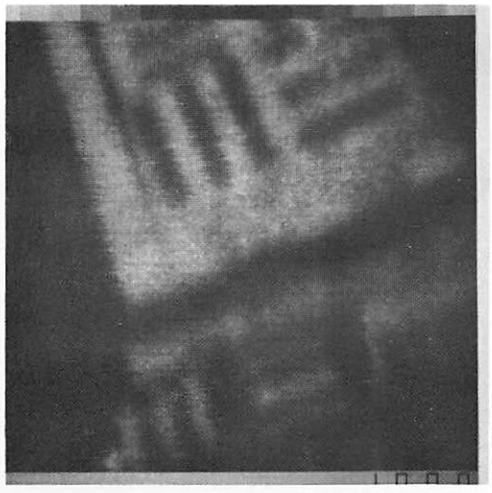


(d) low pass filter - ④

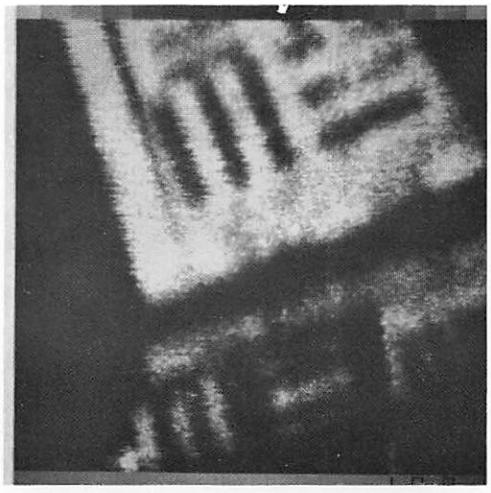


(e) more severe low pass filter - ⑤

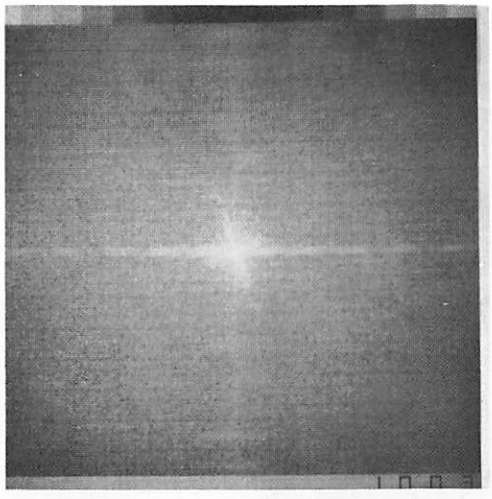
Figure 6.2-2. Blurred wing section.



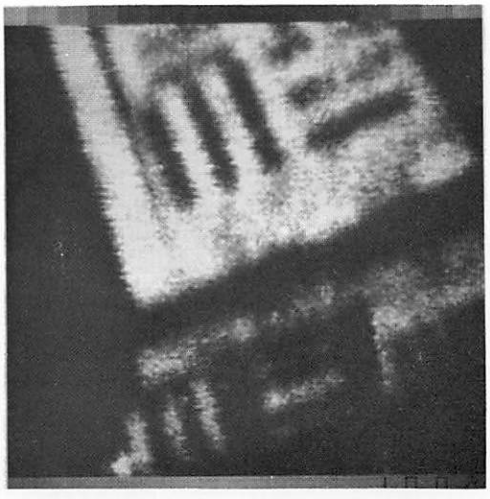
(a) original framed - ⑥



(b) noise clipped - ⑦

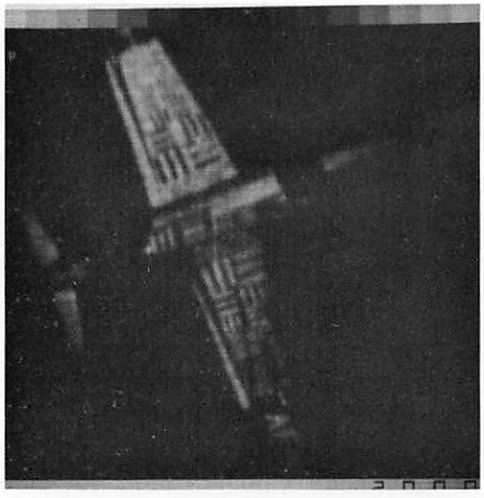


(c) Fourier transform - ⑧

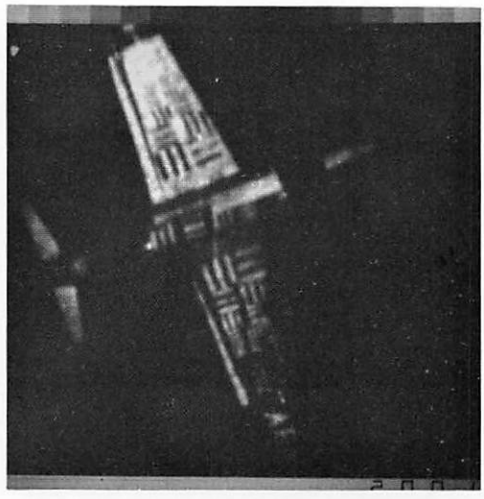


(d) low pass filter - ⑨

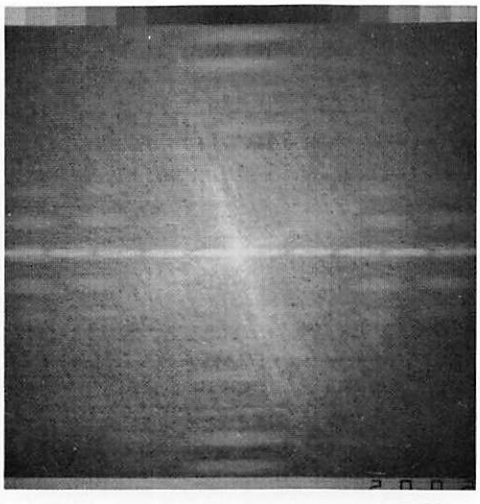
Figure 6.2-3. Center fuselage section.



(a) original framed - (10)



(b) noise clipped - (11)



(c) Fourier transform - (12)

Figure 6.2-4. Entire aircraft

vertical axes (stronger in the vertical for pictures 0004 and 1003). This structure likely is due to some aperture effect associated with the scanner since the stripes are aligned with the axes. Possibly bearing wobble or wear in the rotating mirrors of the scanner might account for this effect. Indeed the above two phenomena graphically demonstrate the power of the use of the Fourier domain for scanner evaluation.

6.3 Picture Decomposition

Ronald Hershel

There is considerable utility in decomposing an N x M array p of picture elements into separable products of eigenvectors. Specifically, let the decomposition be given by

$$p_R = \sum_{i=1}^{R} \lambda_i \underline{u}_i \underline{v}_i^t \qquad (1)$$

where $\underline{v}_i^t$ and $\underline{u}_i$ are eigenvectors of $p^t p$ and pp^t respectively. For a complete representation of p a total of $R \leq M \leq N$ pairs of vectors are required in (1). For many bandwidth-compression and restoration applications, it is interesting to attempt to reduce the number of separable products required to give an accurate but not exact representation of p. This type of representation is defined as

$$p_L = \sum_{i=1}^{L<R} \lambda_i \underline{u}_i \underline{v}_i^t$$

where the associated mean squared error is given by

$$\epsilon_L = |p_R - p_L|^2 = \sum_{i=L+1}^{R} \lambda_i^2 \qquad (2)$$

Since $\{\lambda_i^2\}$ $i=1,\ldots,R$ are the eigenvalues of pp^t, and order in decreasing value, then

$$\lambda_1^2 \geq \lambda_L^2 \geq \cdots \geq \lambda_R^2 > 0$$

The procedure of selecting the first L separable products is optimum in a least squares sense. Unlike other transform techniques, the orthogonal set used for reconstruction is matched to the particular picture at hand. Hence, bandwidth compression must take into account the number of bits required for $\underline{u}_i$ and $\underline{v}_i$ in Eq. (2).

However, for certain applications in frame to frame coding only the relative magnitude λ_i of the components need be calculated each frame where updating of $\underline{u}_i$ and $\underline{v}_i$ is required only when correlation develops; i.e., $\underline{u}_i^t p \underline{v}_j \neq 0$ for $i \neq j$.

In restoring pictures with significant x-y symmetry, the decomposition technique proves extremely powerful both in discriminating unsymmetries (including noise) and in reducing computation times for both linear and nonlinear deconvolution. For separable degradation matrices, the imaging equation for object A and image B becomes

$$S_y A S_x = B \tag{4}$$

If

$$B = \sum_{i=1}^{L} b_i \underline{x}_i \underline{y}_i^t$$

then eq. (4) reduces to finding the pseudo inverse A^*, which must be expressed as a separable product

$$A^* = \sum_{i=1}^{L} b_i \underline{x}_i \underline{y}_i^t$$

where $S_x \underline{x}_i = \underline{u}_i \quad S_y \underline{y}_i = \underline{v}_i \quad i=1,L$

and hence it is possible to solve for $\underline{x}_i$ and $\underline{y}_i$ to reconstruct A.

Numerical Methods Obviously the major shortcoming of this technique is having to solve the eigenvalues and eigenvectors of large matrices since the number of computations for this task grow as N^3. However, since only a relatively few number of components (in particular the largest eigenvalues) are of interest, there are a number of iterative algorithms available to obtain approximate solutions. One promising scheme, developed by Lunczos, uses the matrix P as an operation on a random vector $\underline{b}$ to obtain a sequence of vectors

$$\underline{p}_1, \underline{p}_2, \cdots, \underline{p}_L, \underline{q}_1, \underline{q}_2, \cdots, \underline{q}_L$$

defined by

$$\underline{p}_n = 4P_o^t \underline{q}_{n-1} - 2\underline{p}_{n-1} + \underline{p}_{n-2}$$

$$\underline{q}_n = P_o^t \underline{p}_n$$

By squaring the Fourier transform of each row of a matrix formed by the $\underline{p}$ vectors, a map of the eigenvalues is obtained from $\lambda^2 = 0$ to $\lambda^2 = 1$ (P is normalized by its largest eigenvalue to form P_0). This operation goes as LN^2 for general N x N matrices and considerably faster for sparse or highly redundant pictures.

Since the resolution of the eigenvalue distribution is proportional to L, sufficient separation of nearly degenerate eigenvalues may require more iterations. However, noting that this distribution must be positive, restoration techniques can be employed to more accurately locate the eigenvalues of interest. Once $\{\lambda_i^2\}$ $i=1,2,\ldots J \leq L$ have been estimated the vectors $\underline{p}_i$ and $\underline{q}_i$ can be simply transformed into the required estimates of the eigenvectors pairs $(\underline{u}_i, \underline{v}_i)$ associated with P. Equation (6) suggests an adaptive scanning procedure for real images where a controlled slit transmission $p_i(x)$ scans P in the y direction to yield $q_i(y)$ and then in the x direction to yield $p_{i+1}(x)$, hence considerably increasing S/N as well as providing nearly real time picture decomposition.

7. Image Processing Support Projects

The image processing support projects include hardware and software projects supportive of the image processing research effort.

The first report discusses the development of a real time color image display terminal designed to be used for image reception over the ARPANET. The basic display unit is capable of storing a 256 x 256 byte image received from the network, and presenting it for real time display on an industrial grade television monitor. Extensions to the display will allow natural color and pseudocolor displays of up to 512 x 512 byte resolution.

The following report details progress on the image processing software systems. At present, a library of digitized pictures is accessible over the network. Also, members of the research community may mail hard copy imagery to be digitized, and retrieve the digitized images over the network. Work is progressing on making the VICAR image processing language available over the network.

7.1 Development of Real Time ARPANET Image Display

John E. Tahl

An inexpensive digital image display/printer/scanner for use on the ARPANET is presently under development. This device has been separated into two basic units: the display and the printer/scanner. At present the display section has been designed and is in the process of fabrication. After a working display has been produced, on approximately July 1, 1973, the development of the high resolution printer/scanner section will be undertaken.

The display section will be a completely self-contained unit, including an input processing and decoding section, main display refresh memory, output data processing section and color display monitor. The

initial capability of the display will be as follows:

1. Receive, from an ARPANET TIP, digital picture information, with brightness resolution of up to 64 levels (6 bits) and at input rates up to 19.2K baud.

2. Store the received data in an array of up to 256 x 256 six bit picture values.

3. Present the received and stored information at standard television rates, to high speed output digital to analog converters, for application to the red, green and blue inputs of a color monitor. In this mode the image presentation will be shades of gray. At a 19.2K baud input rate, a complete 256 x 256 image can be received and displayed within approximately 18 seconds.

4. In addition to the black and white display, the input will have full pseudo coloring capability, using a high speed random access memory inserted between the output of the refresh memory and the digital to analog converters. The random access memory will be able to be programmed by transmissions from the TIP or by local switch control. Over 4096 different color combinations of hue saturation and luminescence will be available for pseudo coloring.

5. In addition to the black and white and pseudocolor, the display unit will be capable of being programmed, by transmission from the TIP or by local switch control, for sequential full color presentations. In this mode, red, green or blue definition is assigned to each image and displayed in the corresponding color on the screen of the display. A triple exposure, by a camera, photographing the sequential display of three primary color images, will produce a full color photograph. The displayed image

presentation will occupy only the center one half of the monitor, to provide a pleasing presentation without a blocked appearance.

The second phase of the development will use the storage capability of the display refresh memory, as a TIP buffer, for the transmission and reception of high spatial resolution image information for the future scanner/printer section. Figure 1 is a block diagram of the display portion of the digital image display/printer/scanner.

7.2 USC/ARPANET Image Processing System

James Pepin

In the past six months software development has concentrated in three areas: network telnet support, file transfer support, and Vicar implementation. This software is now readily accessible by a user from a remote network site.

The first area to be discussed will be the telnet server. The purpose of the server is to allow a user at a remote site to log-in and use programs and devices present on the USC IPI system. This service is implemented as a foreground partition in USCPS. This foreground program can support six users in the present configuration. This number can be increased by adding more core to the partition. The types of programs that have been implemented in this monitor are a line editor remote job submission subsystem, a job printing subsystem, a Tektronix display program, directory scanning routine, and communication with the IBM 360/44 operator routine. This set of subsystems allow a remote user to create, submit and display the output of a program run on the 44. This is the basic package. It is implemented and working now. In the next few months a user FTP text editor 360-44 console support and some other features will be added to the capabilities already mentioned.

The server FTP has had extensive usage lately. The Institue has

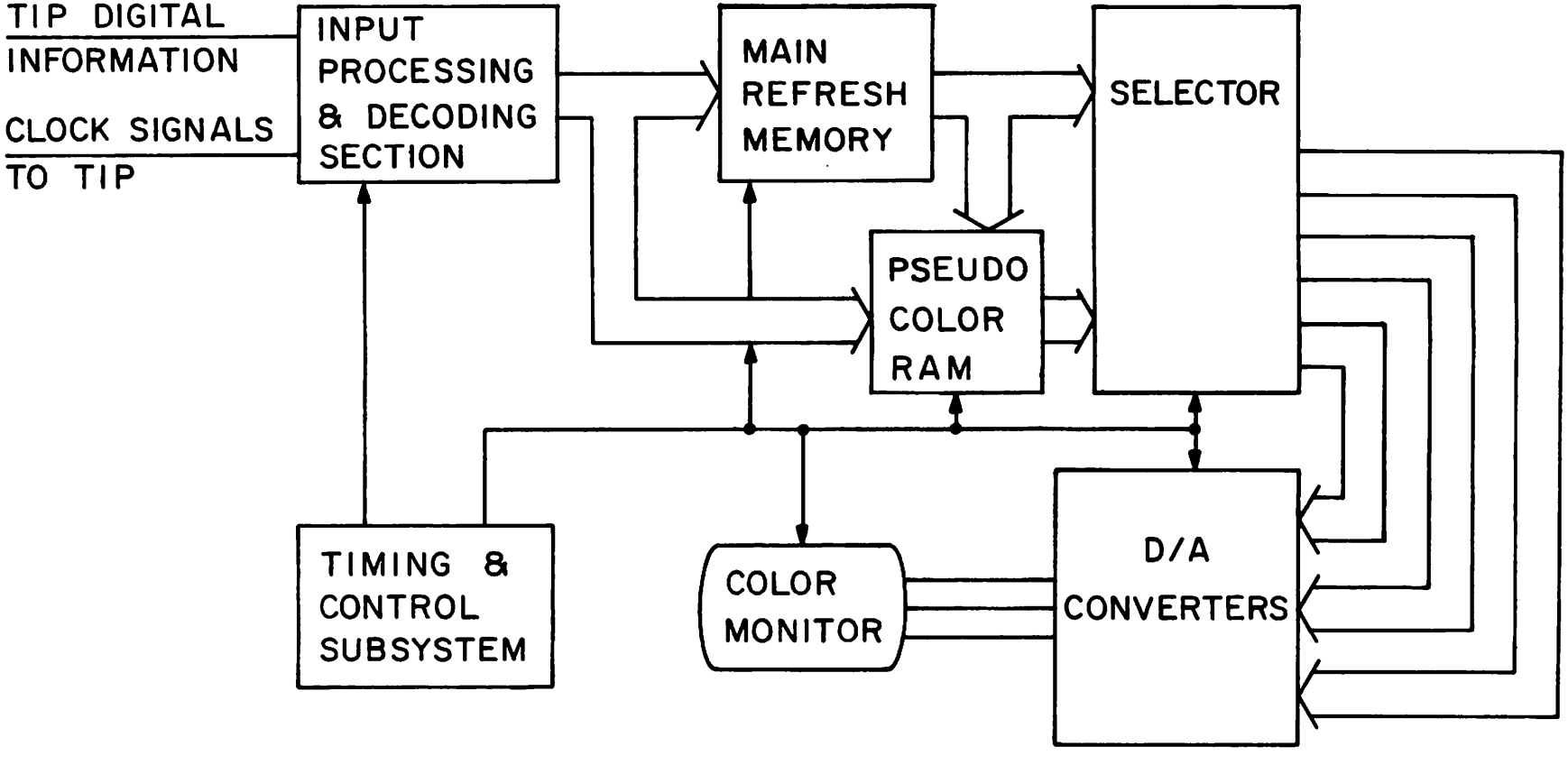


Figure 7.1-1. ARPANET digital image display.

been digitizing and sending data to several network sites, including Carnegie, USC-ISI, and UCSB. The pictures transmitted have ranged in size from 65,000 bytes to 256,000 bytes. Some disappointments have been experienced in data transfer rates, however. The rates obtainable with TENEX systems have been in the 4kb range. This is because of the need to transmit 8 bit bytes in image mode. This is one area of projected effort in the near future. An attempt will be made to increase data rates up to at least 25kb/ This does not seem too impractical since that rate has been realized in communication with other 360 type machines. Work to be done in the FTP area includes implementing of all the modes and types called for in the protocal and inclusion of the new mail capability. Also called for will be the permanent inclusion of the FTP server. As it exists now, FTP is only up on demand of a user and for increased use it should be up regularly.

The work done in the VICAR implementation area basically falls into two categories: implementation of VICAR on the 360/44 and the interfacing of it to the existing VICAR present already on the campus IBM 370/155. The implementation of VICAR under USCPS was not a great difficulty since VICAR, as originally implemented by the Jet Propulsion Laboratory, was on a small 360/44. The main thrust of the effort was in the smoothing out of the problems that the 370/155 system has solved. These problems included the efficient allocation of disk data sets on different physical devices to allow for minimization of disk arm contention. Another area of effort in the 44 VICAR project was to implement the device types allowed by the OSVICAR. This results in the ability to run VICAR programs virtually unchanged on either computer. This is useful for debugging, since turnaround is significantly better on the 44 but the 155 is faster and better suited for production work. Our next area of concern was the interface between the two VICAR processes. The present effort is to make the machine the VICAR programs run "transparent to the user." In this configuration the VICAR language processor will scan the work requested for such items as programs needed, core and CPU usage, and

determine what tapes need to be copied, then proceed to perform the job on the best CPU. This is the major effort at the present time. Preliminary results look very promising for such applications as running the VICAR language processor on the 44, then perhaps collecting the necessary picture files, sending them to a large machine on the net, then returning the results.

8. New Research Projects

The following are summary descriptions of new research studies that will be initiated during the next six months in addition to the continuation of topics presented earlier:

8.1. Generalized Spectrum Extrapolation

William K. Pratt

In transform image coding systems one technique of achieving a bandwidth compression is to discard transform coefficients of actual or expected small value. At the decoder an inverse transform is usually taken with zeros substituted for the discarded components. A new technique has been discovered for estimating the values of the discarded coefficients prior to the inverse transformation. This technique offers the promise of increased resolution and decreased restoration error at no increase in channel bandwidth. Studies will be performed to quantitatively evaluate the performance of this technique and to assess its implementation requirements.

8.2. Facsimile Image Representation

Lloyd R. Welch

The representation of two-level gray scale images by contour information is a promising technique for certain classes of images. In particular, handwritten text should be suitable for such representation. The research divides naturally into two parts: 1) the conversion of the physical image into an array of zeros and ones; and 2) analysis of the contour structure in the resulting array.

The line width of handwritten material with a medium width mechanical pencil lead is about the same as the width of a pixel element of the Muirhead facsimile scanner. Therefore, the magnitude of a pixel sample will vary according to what fraction of the pixel region is covered by pencil lead (or ink). To obtain a reasonable digital representation it will be necessary to do some data processing. Possibly an interpolation algo-

rithm will be required to obtain a finer grid than the Muirhead scanner provides. Then the algorithm for replacing pixel samples with zeros and ones may require more sophistication than simply establishing a threshold.

After this phase of research, the contours will be investigated for structural and statistical regularities with a goal of data reduction.

8.3. Space Variant Point Spread Function Inversion Through Singular Values Decomposition

Harry C. Andrews

The commonly employed linear model for space variant imaging is

$$g(x,y) = \iint h(x,y,\zeta,\eta)\, f(\zeta,\eta)\, d\zeta\, d\eta$$

The corresponding model in discrete form is given by

$$\underline{g} = [H]\, \underline{f}$$

In most imaging situations the image matrix H possesses a relatively low number of degrees of freedom. Advantage of this fact may be utilized to achieve a dimensionality reduction in describing the image model by use of a singular value decomposition (SVD). The SVD is defined as

$$H = \sum_{i=1}^{I} \lambda_i^{\frac{1}{2}}\, \underline{u}_i\, \underline{v}_i^t$$

where

$$HH^t\, \underline{u}_i = \lambda_i\, \underline{u}_i$$

and

$$H^tH\, \underline{v}_i = \lambda_i\, \underline{v}_i$$

Hopefully, I will be much smaller than N^2. Then utilizing the pseudo inverse for inverting H, one obtains

$$H^{-1} = \sum_{i=1}^{I} \lambda_i^{-\frac{1}{2}} \underline{u}_i \underline{v}_i^t$$

Then an estimate of the object, f(x, y), is given by

$$\hat{f} = H^{-1} \underline{g}$$

$$= \sum_{i=1}^{I} \lambda_i^{-\frac{1}{2}} \underline{u}_i \underline{v}_i^t \underline{g}$$

which becomes a sum of separable dot products. To guarantee a positive restoration it is suggested that a vector-wise or point-wise recursive algorithm be developed within the above summation as done in the modified Van Cittert technique.

The advantages of the SVD technique are minimal storage requirement, i.e. I x N locations, and rapid pseudo inverse calculations.

8.4. Vector Scanning Model of Image Motion

Alexander A. Sawchuk

The difficulty of restoring images blurred by motion, particularly space-variant motion, has led to the use of special computation techniques for this purpose [1] - [3]. One promising technique known as vector scanning effectively converts the two spatial dimensions of an image into a sequence of vectors with one spatial dimension to simplify the processing. This project attempts to find useful physical models for the motion degradation in a vector scanning context. Both space-invariant and space-variant motion blur will be considered.

References

1. A. A. Sawchuk, "Space-Variant Image Motion Degradation and Restoration," Proceedings of the IEEE, Vol. 50, p. 854, 1972.

2. N. E. Nahi and T. Assefi, "Bayesian Recursive Image Enhancement," IEEE Transaction on Computers, July 1972.

3. N. E. Nahi, "Role of Recursive Estimation in Statistical Image Enhancement," Proceedings of the IEEE, Vol. 60, July 1972.

8.5. Two Level Statistical Representation of Images and Nonlinear Kalman Filtering

N. E. Nahi

The objective of this study will be to represent a picture by a combination of: 1) statistics of the background; 2) statistice of the object detail (the features included within the object); and 3) the statistics representing the size of the object and its location within the picture. An important factor considered is the fact that the object, with its own statistics, replaces a portion of background having different statistics, where only statistical information on the size and location of the replaced portion is available. A dynamic model of such a process is clearly a nonlinear stochastic system. The enhancement then requires development of appropriate nonlinear recursive filters [1].

Reference

1. N. E. Nahi, Estimation Theory and Applications, Wiley, 1969.

8.6. Textural Measures Applied to Terrestrial Features

Richard P. Kruger

In the near future an effort will be made to gather a data base of terrestrial images for processing using textural measures presently applied to biomedical imagery. This data will consist of satellite or high altitude aerial reconnaissance imagery. Initial classification will attempt to discriminate man made from naturally occurring terrain features with further subdividing of these major categories to await initial results.

8.7 Imaging Systems Inversion

Ronald S. Hershel

In many imaging systems there is a great need in exploring

approaches to solving nonlinear simultaneous equations of the form

$$\mathbb{D}\,(\vec{x})\;\vec{x} = \vec{y}$$

when explicit representation of $\mathbb{D}$ or $\mathbb{D}^{-1}$ are not required. Application of such algorithms to positive and nonlinear restoration techniques are immediate. In the study emphasis will be placed on convergence speed and stability.

9. Publications

The following is a list of papers, articles, and reports published or accepted for publication during the past six months, that have resulted from ARPA sponsored research.

E. Angel and A. K. Jain, "Filtering of Multidimensional Diffusion Processes," Proceedings Sixth Asilomar Conference, Pacific Grove, Cal., November, 1972.

W. Chen and W. K. Pratt, "Slant Transform Monochrome and Color Image Coding," Picture Coding Symposium, University of Southern California, Los Angeles, Cal., January, 1973.

W. Chen and W. K. Pratt, "Color Image Coding with the Slant Transform," Symposium on Applications of Walsh Functions, Washington, D.C., April, 1973.

F. Davarian and W. K. Pratt, "Differential Transform Image Coding," Picture Coding Symposium, University of Southern California, Los Angeles, Cal., January, 1973.

L. D. Davisson, "Universal Coding," Picture Coding Symposium, University of Southern California, Los Angeles, Cal., January, 1973.

W. Frei, "Color Image Statistics," Picture Coding Symposium, University of Southern California, Los Angeles, Cal., January, 1973.

W. Frei, "Multi-Dimensional Redundancy Reduction of Color Image Information," Picture Coding Symposium, University of Southern California, Los Angeles, Cal., January, 1973.

A. Habibi, "Coding Color Images by Differential Pulse Code Modulation," Proceedings International Telemetering Conference, Los Angeles, Cal., October, 1972.

A. Habibi, "Unitary Transformations and DPCM Coding of Pictorial Data," Picture Coding Symposium, University of Southern California, Los Angeles, Cal., January, 1973.

B. R. Hunt and H. C. Andrews, "Comparison of Different Filter Sturctures for Restoration of Images," Sixth Hawaii International Conference on System Science, Honolulu, Hawaii, January, 1973.

A. K. Jain, "Color Distance and Geodesics in Color 3 Space," Journal of the Optical Society of America, Vol. 62, No. 11, November, 1972.

A. K. Jain, "Linear and Nonlinear Interpolation for 2 Dimensional Image Enhancement," Proceedings of the Conference on Decision and Control, New Orleans, Louisiana, December, 1972.

A. K. Jain and W. K. Pratt, "Color Image Quantization," IEEE 1972 National Telecommunications Conference Record, Houston, Texas, December, 1972, p. 34D-1.

W. K. Pratt, "Binary Symmetric Channel Error Effects on PCM Color Image Transmission," IEEE Transactions on Information Theory, Vol. IT-18, No. 5, September, 1972.

W. K. Pratt and R. P. Kruger, "Image Processing over the ARPA Computer Network," Proceedings International Telemetering Conference, Los Angeles, Cal., October, 1972.

A. A. Sawchuk, "Coordinate Transformation in Space-Variant Image Enhancement and Restoration," Journal of the Optical Society of America, Vol. 62, p. 1337, 1972, Annual.