

THEORETICAL AND EXPERIMENTAL PROCESSING OF LUNAR TELEVISION PICTURES

Final Report

SGC 203R-4

N65-17214
NASA CR-60754

November 29, 1962

C. S. Lorens

A.M. Boehmer

J. Gallagher

Prepared for the

JET PROPULSION LABORATORY

of the

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

4800 Oak Grove Drive
Pasadena, California

Under Contract No. NAS 7-100(950249)

Walter F. Storer, Cognizant Engineer

Prepared by

SPACE-GENERAL CORPORATION

9200 E. Flair Drive
El Monte, California

This work was performed for the Jet Propulsion Laboratory, California Institute of Technology, sponsored by the National Aeronautics and Space Administration under Contract NAS7-100.

APPROVED:

RM Stewart

Robert M. Stewart
Associate Manager
Research and Advanced
Development Division

1480

Abstract

It is estimated that large quantities of lunar planetary pictures are to be obtained from forthcoming experiments. The processing of pictures through digital computer techniques offers one possibility for handling the large number of pictures which are to result from these experiments. This report is the final report on a theoretical and experimental study of techniques for processing pictures.

An outline is presented of possible linear and non-linear theoretical work which can be made applicable to processing lunar and planetary pictures. The theoretical material of this report is limited to linear processing. The material deals with exploiting the geometry of pictures principally through the assumption of statistically stationary properties under translation and rotation. Pictures are defined as functions of one index where the index is a 2-dimensional vector corresponding to the 2-dimensional coordinates of the picture. This notation makes it possible to handle pictures in the conventional 1-dimensional notation of classical processing theory while preserving the 2-dimensional properties of the picture. The report deals with representation of the pictures, linear processing, matrix products, correlation functions, stationary and symmetry properties optimum linear processing, trivial processing, orthogonal preprocessing, minimum error, minimization with a constraint, preprocessing with prediction, quantization, and convergence of iterative computations.

The experimental work of this report tests the theoretical results for complexity, usefulness, and correctness. An experimental capability was developed capable of producing pictures with up to 750 elements in 6 shades of brightness corresponding to the 6 faces of the cube onto which the shades of brightness were pasted. This capability was then used in the experimental representation of the Craters Eratosthenes and Archimedes in 15 x 19 fields of elements. Artificial pictures were also constructed having controlled correlation functions. A number of experiments were performed with these pictures to extract in an optimum manner another desired picture. A variety of FORTRAN programs are included for eventually performing the processing computations on a digital computer.

Introduction

The successful extraction of significant data from lunar and planetary television pictures requires the survey and comprehension of presently known techniques as well as the development of new techniques particularly directed at television processing. This report deals with theoretical and experimental work in the field of processing pictures. Much of the material of the report comes from 3 previous reports¹⁹.

The theoretical work is directed at producing techniques for extracting data either before transmission or subsequent to transmission. Pre-processing of television pictures has the possibility of minimizing the transmission channel requirements or maximizing the use of the transmission channel when the channel is fixed and inflexible. Subsequent processing of television pictures after transmission permits an organization of the data into a form which is more meaningful as a final product and also permits hypothesis testing for the extraction of hypothetical data from the pictures.

The quantity of data to be gathered from lunar picture experiments requires both automatic reduction and the automatic design of prototype reduction programs. Many of the pictures will be used solely for mapping areas where only rough maps are presently available. Scientific efforts, however, will be directed at measuring the density distribution of crater radii and typing the debris around the craters as well as within the crater. Knowledge of the crater structure has the possibility of distinguishing between comet and asteroid impacts as well as the possibility of the recognition of volcanic craters. Similar crater studies will be performed on the scientific experiments directed at the Planet Mercury.

Geological studies will be interested in the uniformity and roughness of the surface structure as well as the classification of the surface geology.

Theoretical Studies

Television pictures are normally thought of as a sequence of pictures capable of displaying moving objects. The use of television pictures in lunar and planetary experiments is much broader than this concept. Many of the pictures will be individually distinctive and will require appropriate processing. Some of the pictures will be related to others through observation of substantially the same experiment from different directions and different times. These sets of pictures require simultaneous processing which is considerably different than that required for the processing of a sequence of pictures of moving objects.

Many optimum processing techniques are presently known for the processing of telemetry. Very few of these methods have ever been used in the processing of pictorial data. The following theoretical areas develop a number of techniques directly applicable to processing television pictures with particular emphasis on the use of digital computation facilities for processing pictures both before and after transmission. The areas are listed in order of increasing difficulty.

Linear Special Filtering based on previously measured correlation functions can be used to preprocess data to reduce its quantity as well as to post-process data to remove random irregularities such as noise and to enhance various effects for possible later detection. Most of the theoretical work presented in this report deals with this area. Some of the original work was done by Bell Telephone Laboratories and is reported in references 7, 13, 14 and 24. Much of the theoretical work in 1 dimension as presented in references 11 and 16 is changed into 2 dimensions in this report. The concepts of rotation and translation have an expanded meaning in 2 dimensions.

4

Linear Learning Processes introduce the concept of measuring parameters from sets of hypothetically characteristic data. The development of an intelligent learning process has the possibility of replacing the heuristic guessing which has normally accompanied picture processing in the past. The ideas of References 1, and 4 have the possibility of being applicable. Other approaches are contained in References 8, 15, and 25.

Nonlinear Spatial Filtering and Learning is directed at processing the inherent nonlinear properties of a picture. The shades-of-gray scale is defined as the logarithm of the brightness scale to correspond to the human preception of relative changes in the grey levels rather than the absolute changes in the levels. The brightness scale also has an upper and lower bound. In order to handle these nonlinear effects and possibly others, it is necessary to seriously study techniques of nonlinear processing. Much of the basic theory is contained in Reference 27. Examples of the implementation of this theory will be found in References, 3, 6, 9, 17, and 20.

Sequential Filtering of Pictures which are correlated with one another in sequence is an expansion of the amount of data used in the processing of a single point of a picture. The theory is applicable to scanning systems where it may be desirable to process the movement and change of objects in the pictures. The material of References 12 and 23 has possible application.

Extraction of the presence or absence of data in a picture from only a hypothetical knowledge of what the data is to look like is an area particularly directed toward the discrete recognition or rejection of the presence of data in the picture under observation. The techniques of this area are also applicable to control of experiments and the acceptance of valid commands. The material of references 5, 8, and 18 can be applied to this subject.

Linear Analog Techniques provide an alternate means of processing pictures. Most theoretical work results in discrete processing on a digital computer. A number of multi-dimensional analog techniques also exist which have the possibility of producing simple linear equipment which may work at considerably higher speeds than can be obtained with digital equipment. The original work is contained in Reference 26. Further work is in References 2, and 28.

The following theoretical and experimental work develops and tests a number of linear techniques to handle the two-dimensional aspects of pictures so that the geometry of the picture can be retained in two dimensions rather in one dimension where a considerable amount of theory is presently well known.

Representation of a Picture

The following theoretical development has 2 objectives: 1. to represent a picture and its processing in classical vector and matrix notation, and 2. to preserve the 2-dimensional properties of the picture.

A sampled picture is represented by the row vector x having elements $x_{1,i}$ corresponding to the samples of the picture where the indicies $l = (1,1)$ and $i = (a,b)$ are 2-dimensional vectors. In classical notation, it is customary to write the elements of a row vector sequentially in one row. However, in the notation of this report a row vector is written in a geometrical form, each of the elements of the index vector indicating the geometrical position of the corresponding row vector element. The elements of the representation of x then correspond geometrically to the picture that they represent. A possible representation of a row vector is then

$$x = [x_{1,i}] = [x_i] = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \end{bmatrix}$$

Linear Processing

A second picture y can be obtained from the picture x by linearly weighting and adding the elements of x to form the elements of the picture y . Notationally, the new picture is computed by the matrix formula

$$y = xh + k$$

where

$$\left[y_{1,j} \right] = \left[\sum_i x_{1,i} h_{i,j} + k_{1,j} \right]$$

The summation is over all the values of the index i . In reality, this is a double summation. It is represented here as a single sum because of its simplicity and its similarity to the classical matrix notation of the matrix product.

Linear processing can be used to produce an isolated point in a picture or to approximate some desired result based on the data contained in the processed picture.

Matrix Transposition

The row vector x is a degenerate form of a matrix h having elements $h_{i,j}$ where the indices i and j are 2-dimensional vectors. The transpose h^t of the matrix h has elements $h_{j,i}$.

$$h = \left[h_{i,j} \right] \quad h^t = \left[h_{j,i} \right]$$

The transpose x^t of the row vector $x = [x_{1,i}]$ is defined to be a column vector. Note that the indices have been transposed, not the elements of the vector representing the indices. The representation of the column vector x^t is the same as for the row vector x .

$$x^t = \left[x_{i,1} \right] = \left[x_i \right] = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \end{bmatrix}$$

The i th row of the matrix h is the row vector $[h_{i,j}]$ containing all the elements $h_{i,j}$ for which i is the first index. In a like manner the j th column of the matrix h is the column vector $[h_{i,j}]$ containing all the elements $h_{i,j}$ for which j is the second index.

The vector indices $i = (a,b)$ and $j = (c,d)$ are equal when their corresponding elements are equal. The addition of indices is also element by element.

equality of indices: $i = j \iff a = c \text{ and } b = d$

addition of indices: $k = i+j \iff k = (a+c, b+d)$

The sum, f , of 2 matrices h and k in this notation is the matrix of elements equal to the sum of elements having identical indices.

$$f = h + k$$

where

$$[f_{i,j}] = [h_{i,j} + k_{i,j}]$$

The sum of the two pictures x and y is the sample-by-sample sum of the pictures.

Product of Linear Transformations

The sequential processing of pictures by the relations

$$y = xh \text{ and then } z = yg$$

leads to the product of the matrices h and g since substitution of y into the equation for z produces the relation

$$z = xhg = xc$$

where $c = hg$.

In terms of elements, the matrix equations are

$$\begin{bmatrix} y_{1,j} \end{bmatrix} = \begin{bmatrix} \sum_i x_{1,i} h_{i,j} \end{bmatrix}$$

$$\begin{bmatrix} z_{1,k} \end{bmatrix} = \begin{bmatrix} \sum_j y_{1,j} g_{j,k} \end{bmatrix}$$

so that by substitution

$$\begin{aligned} \begin{bmatrix} z_{1,k} \end{bmatrix} &= \begin{bmatrix} \sum_j \sum_i x_{1,i} h_{i,j} g_{j,k} \end{bmatrix} \\ &= \begin{bmatrix} \sum_i x_{1,i} c_{i,k} \end{bmatrix} \end{aligned}$$

where

$$\begin{bmatrix} c_{i,k} \end{bmatrix} = \begin{bmatrix} \sum_j h_{i,j} g_{j,k} \end{bmatrix}$$

The product of the matrices is thus a sum over the column index j of the first matrix h and the row index j of the second matrix g .

A square matrix h is one which has an equal number of row and column indices. The diagonal of a square matrix is the set of elements for which the row and column indices are identical. The identity matrix I is then defined to be the one which is 1 on the diagonal and 0 elsewhere.

$$\begin{bmatrix} I_{1,j} \end{bmatrix} = \begin{bmatrix} \delta_{i,j} \end{bmatrix} \quad \text{where} \quad \delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Correlation Functions

The cross-correlation function ϕ_{xz} between the pictures x and z is defined as the matrix product.

$$\phi_{xz} = \overline{x^t z}$$

where the bar denotes the ensemble average of each element in the matrix product $x^t z$. Due to the transposition, there is only one index, $l = (1,1)$, over which the summation is performed.

$$\left[\phi_{xz} \ i, j \right] = \left[\overline{x_{i,1} z_{1,j}} \right] = \left[\overline{x_{i,1} z_j} \right]$$

It should be noted that the two pictures x and z can have a different number of samples and are not restricted to being the same geometrical size. The cross-correlation function of a picture x with itself is known as the auto-correlation function.

$$\left[\phi_{xx} \ i, j \right] = \left[\overline{x_{i,1} x_{1,j}} \right] = \left[\overline{x_{i,1} x_j} \right]$$

In this 2-dimensional notation, the correlation functions have many of the same properties as the classical one-dimensional correlation functions. The transposition of the cross-correlation function between x and z produces the cross-correlation between z and x .

$$\phi_{xz}^t = \left[\overline{x^t z} \right]^t = \left[\overline{x_{i,1} z_{1,j}} \right]^t = \left[\overline{z_{i,1} x_{1,j}} \right] = \overline{z^t x} = \phi_{zx}$$

The auto-correlation is also its own transpose.

$$\phi_{xx}^t = \phi_{xx}$$

Statistical independence between the two pictures produces the results that the cross-correlation function is the matrix product of the mean values of the pictures

$$\varphi_{xz} = \overline{x^t z} = \overline{x^t} \overline{z} \quad x \text{ and } z \text{ statistically independent}$$

where the mean value of the picture x is the row vector made up of the mean value of each sample of the picture. In the particular case where the pictures are statistically independent and one of them has a zero mean, the cross-correlation function is zero.

$$\varphi_{xz} = 0.$$

The auto-correlation of the sum of two pictures x and y is the sum of the individual auto-correlations and the two cross-correlations.

$$\varphi = \overline{(x+y)^t (x+y)} = \varphi_{xx} + \varphi_{yy} + \varphi_{xy} + \varphi_{yx}$$

Where the pictures are statistically independent and one has a zero mean, the auto-correlation of the sum is then just the sum of the individual auto-correlation functions.

$$\varphi = \varphi_{xx} + \varphi_{yy}$$

The correlation function of a picture made up of independent elements all having the same standard deviation σ and mean m is an example which has many uses.

$$\overline{x_i x_j} = \sigma^2 \delta_{i,j} + m^2$$

The correlation function is then

$$\phi_{xx} = \sigma^2 I + m^2 U$$

where I is the identity matrix and U is the unit matrix of all 1's.

The cross-correlation function between two pictures y_1 and y_2 linearly obtained from pictures x_1 and x_2 is a matrix operation on the cross-correlation functions between x_1 and x_2 .

$$\begin{aligned} y_1 &= x_1 h_1 + k_1 & y_2 &= x_2 h_2 + k_2 \\ \phi_{y_1 y_2} &= \overline{y_1^t y_2} & &= \overline{(x_1 h_1 + k_1)^t (x_2 h_2 + k_2)} \\ &= h_1^t \phi_{x_1 x_2} h_2 + h_1^t \bar{x}_1^t k_2 + k_1^t \bar{x}_2 h_2 + k_1^t k_2 \end{aligned}$$

In the case where the pictures x_1 and x_2 have zero mean the cross-correlation function $\phi_{y_1 y_2}$ has the particularly simple form

$$\phi_{y_1 y_2} = h_1^t \phi_{x_1 x_2} h_2 + k_1^t k_2$$

A further simplification occurs in the computation of the auto-correlation function ϕ_{yy} of the picture y obtained from the picture x when the picture x is made up of independent, zero-mean elements. In this case, $\phi_{xx} = \sigma^2 I$ so that the auto-correlation function of the picture y is

$$\phi_{yy} = \sigma^2 h^t h + k^t k.$$

An example of the use of linear processing theory is in the construction of an artificial picture from a random number table. Artificial pictures are often used in experiments where it is necessary to know and control the properties of the pictures.

If it is desired to construct an artificial picture with a prescribed auto-correlation function ϕ_{yy} , a one-dimensional technique is available for symmetrically factoring the matrix ϕ_{yy} to obtain a linear process h which will transform a statistically independent picture into an artificial picture having this correlation function. The two-dimensional implications of this technique are not presently known.

A later section describes experiments in artificially creating correlated pictures from random number tables and linear processing.

Positive Definite Condition

An ensemble of pictures x is defined to be linearly dependent whenever there exists a nontrivial linear process h of the picture x about its mean \bar{x} which will produce a one-element picture $y = x'h$ having a zero correlation function.

$$\phi_{yy} = \overline{y^2} = 0.$$

$$x' = x - \bar{x}.$$

A set of pictures in which the same two samples are always identical is an example of an ensemble of linear dependent pictures. When the ensemble of pictures is not linearly dependent, it is defined to be linearly independent. In the case of an ensemble of linear independent pictures, every linear process h producing a one-element picture has a positive valued correlation function

$$\phi_{yy} = h^t \phi_{x'x'} h > 0 \text{ for every } h.$$

An auto-correlation matrix thus is defined to be positive definite whenever it is obtained from an ensemble of linearity independent pictures. Where the matrix is positive definite, the determinant $|\phi_{xx}|$ will be non-zero.

The implication of the picture x being linearly dependent is that at least one sample x_1 is essentially computable from the rest of the picture. By definition, if x is linearly dependent, there exists a non-trivial h such that

$$y_1 = \sum_i x_{1i} h_{i1}$$

where

$$\overline{y_2} = 0.$$

This implies that y is essentially zero.

$$y = 0.$$

Since at least one of the elements of h has to be non-zero, assume that $h_{1,1} \neq 0$. This produces the result that

$$x_{1,1} = \frac{1}{h_{1,1}} \sum_{i \neq (1,1)} x_{1,i} h_{i,1}$$

Thus, in the case where the auto-correlation matrix ϕ_{xx} has a zero determinant $|\phi_{xx}| = 0$, the auto-correlation matrix is non-positive definite so that the samples of the picture are linearly dependent. At least one of the samples can be produced from the others. In the case of an auto-correlation function ϕ_{xx} having a zero determinant $|\phi_{xx}| = 0$, it is always possible to drop the dependent samples with a linear process h so as to produce a picture $w' = x'h$ with a smaller number of linear independent samples and an auto-correlation function $\phi_{w',w'}$, which is positive definite and having a non-zero determinant.

$$|\phi_{w',w'}| \neq 0.$$

The new picture w' will contain all of the information contained in the original picture x since the linearly dependent samples are computable from the linearly independent samples.

Stationary Correlation Functions

The initial definition of the correlation function

$$\phi_{xz} = \left[\overline{x_i z_j} \right]$$

permitted each element in the matrix ϕ_{xz} to have a different value. In many practical situations the correlation function will be independent of both translation and rotation, and dependent only on the distance of separation between the elements x_i and z_j .

The correlation function ϕ_{xz} is defined to be stationary when the elements of its matrix are functions only of the difference between the indices. A stationary correlation function is thus independent of translation but not of rotation. Whenever $\phi_{xz} = \left[\overline{x_i z_j} \right]$ is stationary, there exists $\phi_{xz}(k)$ such that $\phi_{xz}(k) = \overline{x_i z_{i+k}}$ is independent of the index i . The addition and subtraction of indices is a vector addition and subtraction. A stationary cross-correlation function satisfies the relation

$$\left[\phi_x(k) \right] = \left[\overline{x_i z_{i+k}} \right] = \left[\overline{z_i x_{i-k}} \right] = \left[\phi_{zx}(-k) \right]$$

The stationary auto-correlation function ϕ_{xx} satisfies the relation

$$\phi_{xx}(k) = \phi_{xx}(-k).$$

Symmetric Correlation Functions

The four elements x_{11} , x_{12} , x_{31} , x_{32} of the picture

$$\begin{bmatrix} \underline{x_{11}} & \underline{x_{12}} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \underline{x_{31}} & \underline{x_{32}} & x_{33} \\ x_{41} & x_{42} & x_{43} \end{bmatrix}$$

normally have a correlation function such that

$$\overline{x_{11} x_{12}} = \overline{x_{31} x_{32}} \quad \text{horizontal}$$

$$\overline{x_{11} x_{31}} = \overline{x_{12} x_{32}} \quad \text{vertical}$$

$$\overline{x_{11} x_{32}} = \overline{x_{12} x_{31}} \quad \text{diagonal}$$

The first two relations are stationary relations. In the horizontal relation the difference in indices (1,2) - (1,1) on the left and (3,2) - (3,1) on the right, are the same. The third relation is not a stationary relation since the difference of the indices on the left is (2, 1) and on the right (2,-1).

In order to include the diagonal symmetry of the correlation, a stationary correlation function is defined to be symmetric whenever the correlation function is a function of the magnitude of the elements of the differences between the indices. That is, if the stationary correlation function $\phi(k)$ is symmetric then $\phi(k_1) = \phi(k_2)$ whenever $k_1 = (a_1, b_1)$ and $k_2 = (a_2, b_2)$ such that $|a_1| = |a_2|$ and $|b_1| = |b_2|$. When this property holds for the entire correlation function, it is said to be symmetric.

Distance

Further restrictions can be placed upon a stationary correlation function by specifying that the correlation is only a function of the distance between the elements being correlated. The distance d between the picture elements having indices $i = (i_1, i_2)$ and $j = (j_1, j_2)$ is defined to be

$$d = \sqrt{(i_1 - j_1)^2 + (i_2 - j_2)^2}$$

Two correlation elements are defined to be equal whenever they are the correlation function of picture elements separated by the same distance.

A further modification can be made by specifying the distance measured in terms of an ellipse. This modification should have some use in the consideration of scanned television pictures where the scanning process introduces a distortion into the correlation function.

Stationary and Symmetric Matrix Products

It would seem intuitive that the matrix product of stationary and symmetric matrices would also be stationary and symmetric. It is fortunate that this is the case since it makes it possible to construct invariant symmetric processing devices which are a function of the distance and direction from the particular picture element which is to be constructed.

Whenever the matrices h and g are stationary, their matrix product $c = hg$ is also stationary. That is if

$$\begin{bmatrix} c_{ik} \end{bmatrix} = \begin{bmatrix} \sum_j h_{ij} g_{jk} \end{bmatrix}$$

where h is a function of the difference $j-i$ and g is a function of the difference

$k-j$, then c is a function of the difference $k-i$.

$$\begin{aligned} c_{i+a,k+a} &= \left[\sum_j h_{i+a,j} g_{j,k+a} \right] \\ &= \left[\sum_j h_{i+a,j+a} g_{j+a,k+a} \right] \\ &= \left[c_{i,k} \right] \end{aligned}$$

In terms of the differences between the indices $k-i = r$ and $k-j = s$ the matrix product becomes the familiar convolution formulas.

$$c(r) = \begin{bmatrix} \sum_s h(r-s) g(s) \\ \sum_{s'} h(s') g(r-s') \end{bmatrix}$$

A slightly more useful formular in computation is

$$c(r) = \begin{bmatrix} \sum_s h(r+s) g(-s) \\ \sum_s h(-s) g(r+s) \end{bmatrix}$$

Thus to compute the element $c(r)$, the field of elements of g is reflected through the origin and correspondingly multiplied and summed with the field of elements h starting at the appropriate element $h(r)$. For example where the matrices are

$$h = \begin{bmatrix} 0 & h_{0,1} & 0 \\ h_{-1,0} & h_{0,0} & h_{1,0} \\ 0 & h_{0,-1} & 0 \end{bmatrix} \quad g = \begin{bmatrix} 0 & g_{0,1} & g_{1,1} \\ 0 & g_{0,0} & g_{1,0} \\ 0 & 0 & 0 \end{bmatrix}$$

the field of g is reflected to give

$$g(-s) = \begin{bmatrix} 0 & 0 & 0 \\ g_{1,0} & g_{0,0} & 0 \\ g_{1,1} & g_{0,1} & 0 \end{bmatrix}$$

The term $c_{0,0}$ of the product is obtained by directly multiplying the fields and summing.

$$c_{00} = \begin{bmatrix} 0.0 & h_{0,1}^0 & 0.0 \\ h_{-1,0} g_{1,0} & h_{0,0} g_{0,0} & h_{1,0}^0 \\ 0 g_{1,1} & h_{0,-1} g_{0,1} & 0.0 \end{bmatrix}$$

$$= h_{0,0} g_{0,0} + h_{-1,0} g_{1,0} + h_{0,-1} g_{0,1}$$

The calculation of the element $c_{1,1}$ of the product is obtained by moving the field $g(-s)$ to the element $h_{1,1}$, multiplying the fields, and summing.

$$c_{11} = \begin{bmatrix} 0.0 & h_{0,1} g_{1,0} & 0 g_{0,0} \\ h_{-1,0}^0 & h_{0,0} g_{1,1} & h_{1,0} g_{0,1} \\ 0.0 & h_{0,-1}^0 & 0.0 \end{bmatrix}$$

$$= h_{0,0} g_{1,1} + h_{0,1} g_{1,0} + h_{1,0} g_{0,1}$$

Further calculations would produce the matrix

$$c = \begin{bmatrix} & c_{0,2} & c_{1,2} & \\ c_{-1,1} & c_{0,1} & c_{1,1} & c_{2,1} \\ c_{-1,0} & c_{0,0} & c_{1,0} & c_{2,0} \\ & c_{0,-1} & c_{1,-1} & \end{bmatrix}$$

where only the non-zero terms have been indicated.

These are convenient formulas for both hand computation and machine computation.

From the matrix expression it can be shown that the product of two matrices h and g which are their own transpose ($h^t = h$ and $g^t = g$), that the product $c = hg$ is not necessarily also its own transpose.

$$c^t = g^t h^t = gh \neq hg = c$$

However, when the matrices are dependent only on the difference between their indices and thus stationary, the product will be its own transpose when the factors are their own transposes. A stationary matrix $h = h(s)$ is its own transpose h^t when $h(s) = h(-s)$. Thus, from the convolution formulas

$$\begin{aligned} [c(-r)] &= \left[\sum_s h(-r-s) g(s) \right] \\ &= \left[\sum_{s'} h(-r+s') g(-s') \right] \\ &= \left[\sum_{s'} h(r-s') g(s') \right] \\ &= [c(r)] \end{aligned}$$

where the first step is the change of variable $s' = -s$ and the second step is the use of the self transpose property on the factors h and g .

A similar result is that the product of stationary symmetric matrices is also symmetric. A stationary matrix is symmetric when it is invariant under reflection of any of the elements of the vector representing the index difference. That is, where the indices $s = (a,b)$ and $s' = (-a,b)$, the stationary matrix $g = [g(s)]$ is symmetric in the first dimension when $[g(s)] = [g(s')]$. For example

$$g = \begin{bmatrix} 0 & 3 & 0 \\ 1 & 5_{0,0} & 1 \\ 0 & 2 & 0 \end{bmatrix}$$

is symmetric in the horizontal dimension and not in the vertical dimension.

Where both the stationary factors of a product have a particular symmetry, the product also has that symmetry. In particular where the indices are

$$\begin{aligned} s &= (a,b) & r &= (c,d) \\ s' &= (-a,b) & r' &= (-c,d) \end{aligned}$$

the matrix product is

$$\begin{aligned} [c(r')] &= \left[\sum_s h(r'+s) g(s) \right] \\ &= \left[\sum_{s'} h(r'+s') g(s') \right] \\ &= \left[\sum_s h(r+s) g(s) \right] \\ &= [c(r)] \end{aligned}$$

where the first step is a change in the order of summation and the second step is the use of the symmetry property of the factors h and g.

It should be noted that fully symmetric stationary matrices have the self transpose property. Thus, in the matrix product, the fields can be multiplied directly together without a reflection where both of the factors are symmetric in both dimensions.

Construction of a Correlated Picture

An interesting example of the product of symmetric stationary matrices is the processing of a zero-mean picture x whose elements are uncorrelated to obtain the picture $y = xh$. The correlation function ϕ_{xx} for the picture x is

$$\phi_{xx} = \sigma^2 I$$

The correlation function ϕ_{yy} of the second picture y is then

$$\begin{aligned}\phi_{yy} &= h^t \phi_{xx} h \\ &= \sigma^2 h^t h\end{aligned}$$

A specific example of a stationary symmetric matrix h is

$$h = \begin{bmatrix} & w & & \\ w & 1 & w & \\ & & & \\ & & & w \end{bmatrix}$$

The matrix multiplication then produces the correlation function

$$\phi_{yy} = \sigma^2 \begin{bmatrix} & & & & w^2 & \\ & & & & 2w^2 & 2w & 2w^2 \\ & & & & 2w & 1+4w^2 & 2w & w^2 \\ & & & & 2w^2 & 2w & 2w^2 & \\ & & & & & & & w^2 \end{bmatrix}$$

This correlation is then both stationary and symmetric. The normalized correlation is plotted in Figure 1 for several values of w. The maximum $\frac{\phi(1)}{\phi(0)}$ occurs for $w = 1/2$.

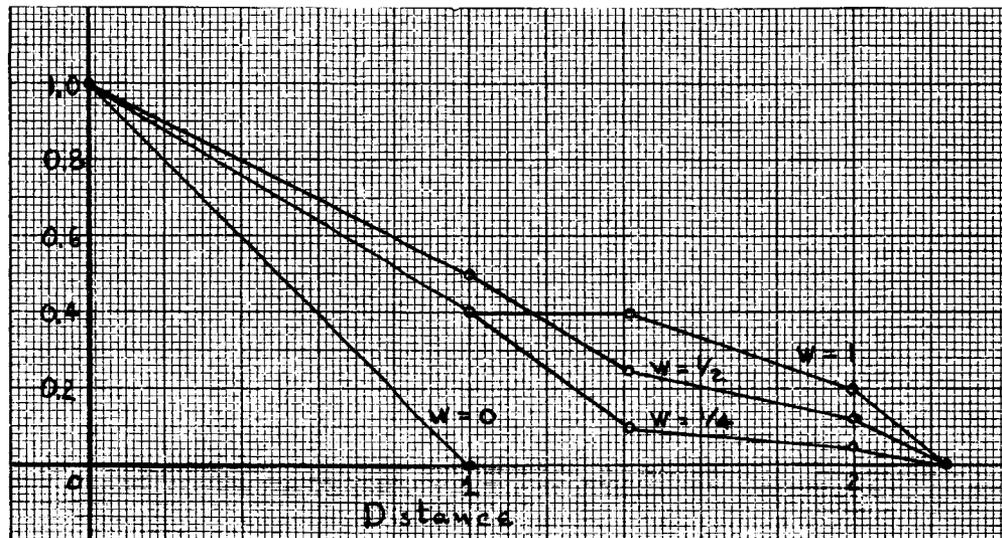


Figure 1, Theoretical Correlation as a Function of Distance for various Weighting Coefficients of the Linear Process $y = xh$

Experimentally the uncorrelated picture of Figure 2 was processed with the stationary processor

$$h = \begin{bmatrix} & 1/2 & \\ 1/2 & 1 & 1/2 \\ & 1/2 & \end{bmatrix}$$

corresponding to the weighting coefficient $w = 1/2$. This processing produced the correlated picture in Figure 3. The smoothness of Figure 3 is an indication of the correlation in contrast to the sharpness of the independent picture of Figure 2.

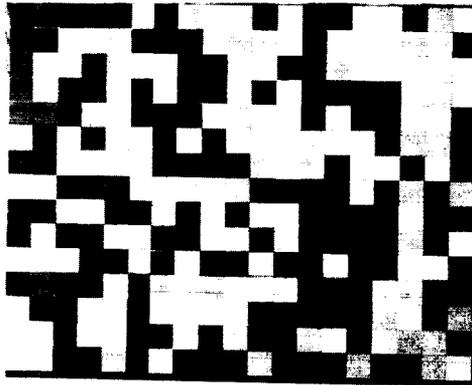


Figure 2. An Uncorrelated Picture

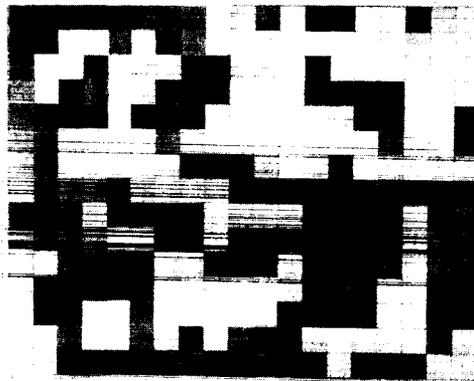


Figure 3. A Correlated Picture Obtained by Processing the Uncorrelated
Picture of Figure 2

Desired Pictures

Linear processing of the picture x produces another picture y which may have one sample or many. Normally, the processing of the picture x is directed at obtaining a third picture z , which is known as the desired picture, as in Figure 4.

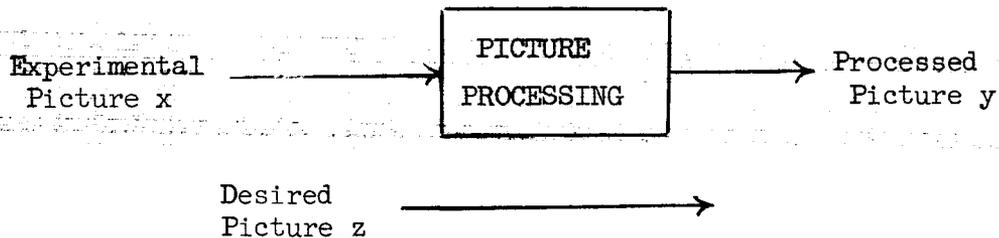


Figure 4. Picture Processing Directed at Obtaining the Desired Picture z .

Exact construction of the desired picture z from the picture x is normally prohibited by the inherent random differences existing between x and z . The error picture in the processing is defined as the difference between the pictures y and z .

$$e = y - z.$$

The mean squared error of each sample of the error picture is the diagonal of the auto-correlation of the error picture

$$\phi_{ee} = \overline{(y-z)^t (y-z)}$$

Optimum Processing

One approach to the construction of a picture y as close to the picture z is to make the squared error between each sample of the two pictures y and z as small as possible. In the case where the process is a linear process, the process can be represented by the matrix relation

$$y = x'h + k$$

where h and k are optimum matrices to be specified and x' is the difference between the picture x and the mean picture \bar{x} , $x' = x - \bar{x}$. Variation of the mean square error

$$\varphi_{ee} = \overline{(y-z)^t (y-z)}$$

produces the relation

$$\begin{aligned} \partial \varphi_{ee} &= 2 (\partial h^t x'^t + \partial k^t) (x'h + k - z) \\ &= 2 \partial h^t (\varphi_{x'x'}^h - \varphi_{x'z}) \\ &\quad + 2 \partial k^t (k - \bar{z}) \end{aligned}$$

The two relations

$$\begin{aligned} \varphi_{x'x'}^h &= \varphi_{x'z} \\ k &= \bar{z} \end{aligned}$$

are sufficient to minimize each diagonal element in the error correlation matrix φ_{ee} . This implies that in the mean square, each element of the produced picture y is as close to each element of the desired picture z as can be possible using linear processing.

The processing which produces a minimum mean squared error for each element is thus

$$y = x' \varphi_{x'x'}^{-1} \varphi_{x'z} + \bar{z}$$

This solution assumes that the auto-correlation $\varphi_{x'x'}$ has an inverse. In the case where the auto-correlation $\varphi_{x'x'}$ has no inverse, the determinant $|\varphi_{x'x'}|$ is zero indicating that it is not positive definite. In this case, there exists a linear process h which eliminates a set of samples linearly dependent upon the

rest of the samples. The new set of linearly independent samples w contains all of the information of the original set of samples. The optimum process is then

$$\begin{aligned} y &= w' \varphi_{w'w}^{-1} \varphi_{w'z} + \bar{z} \\ &= x' h [h^t \varphi_{x'x} h]^{-1} h^t \varphi_{x'z} + \bar{z} \end{aligned}$$

where use has been made of the relation

$$w' = x'h$$

Simple Example

A simple example of linear processing results from the consideration of pictures z which have had statistically-independent zero-mean noise added to them. The picture to be processed is then

$$x = z + n$$

and the desired picture is z . The auto-correlation of the picture is the sum of the auto-correlation functions of the desired picture z and the noise n .

$$\varphi_{xx} = \varphi_{zz} + \varphi_{nn}$$

and the cross-correlation φ_{xz} between the picture x and the desired picture z is

$$\varphi_{xz} = \overline{x^t z} = \overline{z^t z} + \overline{n^t z} = \varphi_{zz}$$

The mean value of the picture x is equal to the mean value of the desired picture z

$$\bar{x} = \bar{z}.$$

The optimum process is thus,

$$\begin{aligned} y &= x' (\phi_{z'z'} + \phi_{nn})^{-1} \phi_{z'z'} + \bar{z} \\ &= x' (I + \phi_{z'z'}^{-1} \phi_{nn})^{-1} + \bar{z} \end{aligned}$$

where

$$z' = z - \bar{z} \text{ and } x' = x - \bar{x}.$$

For small values of noise the process is approximately

$$y = x - x' \phi_{z'z'}^{-1} \phi_{nn}$$

This is a useful result in picture processing since it essentially represents a slight "touching up" of the original picture x by the picture $-x' \phi_{z'z'}^{-1} \phi_{nn}$. Under normal operation of the transmission facilities, noise is usually small and statistically-independent, zero-mean.

At the other extreme, where the picture is predominantly lost in the noise, the optimum processor is

$$y = x' \phi_{nn}^{-1} \phi_{z'z'} + \bar{z}$$

In the case of a one sample picture, the processing should be

$$y = x' \frac{\rho_{11}}{\rho_{11} + \sigma_{11}} + \bar{z}$$

where

$$\phi_{z'z'} = [\rho_{11}] \text{ and } \phi_{nn} = [\sigma_{11}].$$

In the case of a two sample picture, the process is the following operation in terms of classical matrix operations.

$$[y_1, y_2] = \frac{1}{\Delta\phi} [x'_1, x'_2] \begin{bmatrix} \rho_{22} + \sigma_{22} & -(\rho_{12} + \sigma_{12}) \\ -(\rho_{12} + \sigma_{12}) & \rho_{11} + \sigma_{11} \end{bmatrix} \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{12} & \rho_{22} \end{bmatrix} + [\bar{z}_1, \bar{z}_2]$$

where

$$\Delta\phi = (\rho_{11} + \sigma_{11})(\rho_{22} + \sigma_{22}) - (\rho_{12} + \sigma_{12})^2$$

and

$$\phi_{z'z'} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{12} & \rho_{22} \end{bmatrix} \quad \phi_{nn} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

Stationary Solutions

In the case where the experimental picture x and desired picture z are stationary, the processing device using the optimum h and k is also a stationary device. Under the stationary condition, h is found from the solution of the set of equations

$$\sum_r \phi_{x'x'}(s-r) h(r) = \phi_{x'z}(s)$$

or

$$\sum_r \phi_{x'x'}(-r) h(r+s) = \phi_{x'z}(s)$$

Since there is only one element in the stationary matrix k , the matrix k is symmetric. Also, since the matrix h is the product of the matrices $\phi_{x'x'}^{-1}$ and $\phi_{x'z}$, it will be symmetric when the matrices $\phi_{x'x'}$ and $\phi_{x'z}$ are symmetric. In like manner the matrix h will be a function of distance when the matrices $\phi_{x'x'}$ and $\phi_{x'z}$ are functions of distance. Further simplification results in the equations for h where the matrices h , $\phi_{x'x'}$ and $\phi_{x'z}$ are functions of only distance. All the h 's at a particular distance are equal and thus require only one equation. That is the set of equations

$$\sum_r \phi_{x'x'}(s-r) h(r) = \phi_{x'z}(s)$$

needs to be solved for only distinct distances in the index s .

In particular, for the auto correlation matrix

$$\phi_{x'x'} = \begin{bmatrix} \phi_{\sqrt{2}} & \phi_1 & \phi_{\sqrt{2}} \\ \phi_1 & \phi_0 & \phi_1 \\ \phi_{\sqrt{2}} & \phi_1 & \phi_{\sqrt{2}} \end{bmatrix}$$

and crosscorrelation matrix

$$\phi_{x'z} = \begin{bmatrix} \omega_1 \\ \omega_1 & \omega_0 & \omega_1 \\ \omega_1 \end{bmatrix}$$

the equations for the linear processor

$$h = \begin{bmatrix} h_1 \\ h_1 & h_0 & h_1 \\ h_1 \end{bmatrix}$$

$$\begin{aligned} \text{are } \varphi_0 h_0 + 4\varphi_1 h_1 &= \omega_0 \\ \varphi_1 h_0 + (\varphi_0 + 2\varphi\sqrt{2} + \varphi_2) h_1 &= \omega_1 \end{aligned}$$

Solution of these equations produce the coefficients

$$\begin{aligned} h_0 &= \frac{\omega_0 (\varphi_0 + 2\varphi\sqrt{2} + \varphi_2) - 4\omega_1 \varphi_1}{\varphi_1 (\varphi_0 + 2\varphi\sqrt{2} + \varphi_2) - 4\varphi_1^2} \\ h_1 &= \frac{\varphi_1 \omega_0 - \varphi_0 \omega_1}{\varphi_0 (\varphi_0 + 2\varphi\sqrt{2} + \varphi_2) - 4\varphi_1^2} \end{aligned}$$

Generalization of this solution indicates that all coefficients at the same distance are equal. The number of simultaneous equations in the solution for the coefficients is equal to the number of distinct distances. This is a particularly practical result in that it indicates that the first step in the linear processing the picture x is to first add all the samples at a particular distance. This produces a computational reduction of about four for rectangular grids and up to twelve for hexagonal grids. The hexagonal matrix of Figure 5 has a computation reduction of 7/2.

$$h = \begin{bmatrix} h_1 & h_1 \\ h_1 & h_0 & h_1 \\ h_1 & h_1 \end{bmatrix}$$

FIGURE 5. A Linear Processor Based on a Hexagonal Matrix

The coefficients for the hexagonal linear process are

$$h_0 = \frac{\omega_0(\varphi_0 + 2\varphi_1 + 2\varphi\sqrt{3} + \varphi_2) - 6\omega_1\varphi_1}{\varphi_0(\varphi_0 + 2\varphi_1 + 2\varphi\sqrt{3} + \varphi_2) - 6\varphi_1^2}$$

$$h_1 = \frac{\omega_1\varphi_0 - \omega_0\varphi_1}{\varphi_0(\varphi_0 + 2\varphi_1 + 2\varphi\sqrt{3} + \varphi_2) - 6\varphi_1^2}$$

Minimum Error

A minimum squared error is produced for every sample in the linear processing

$$y = x' h + \bar{z}$$

where

$$h = \varphi_{x'x'}^{-1} \varphi_{x'z}$$

The use of any other process

$$y = x' (h + h_\delta) + (\bar{z} + k_\delta)$$

produces the auto-correlation of the error picture

$$\varphi_{ee} = e^t e = [x'(h + h_\delta) + (\bar{z} + k_\delta) - z]^t [x'(h + h_\delta) + (\bar{z} + k_\delta) - z]$$

where the diagonal entries are the squared sample errors. After some reduction with the expression $\varphi_{x'x'} h = \varphi_{x'z}$, the error is

$$\begin{aligned}
\varphi_{ee} &= \varphi_{zz} + (\bar{z}^t + k_{\delta}^t) (\bar{z} + k_{\delta}) - (\bar{z}^t + k_{\delta}^t) \bar{z} - \bar{z}^t (\bar{z} + k_{\delta}) \\
&\quad (h^t + h_{\delta}^t) \varphi_{x'x'} (h + h_{\delta}) - (h^t + h_{\delta}^t) \varphi_{x'z} - \varphi_{zx'} (h + h_{\delta}) \\
&\quad + (h^t + h_{\delta}^t) \bar{x}'^t (\bar{z} + k_{\delta}) + (\bar{z}^t + k_{\delta}^t) \bar{x}' (h + h_{\delta}) \\
&= \varphi_{z'z'} - h^t \varphi_{x'x'} h + h_{\delta}^t \varphi_{x'x'} h_{\delta} + k_{\delta}^t k_{\delta}
\end{aligned}$$

Due to the positive definite condition, the diagonal entries of the term $h_{\delta}^t \varphi_{x'x'} h_{\delta}$ and $k_{\delta}^t k_{\delta}$ are always positive except where $h_{\delta} = 0$ and $k_{\delta} = 0$. Thus, for a linear independent picture x' , the mean error is uniquely obtained by the optimum process

$$y = x'h + \bar{z}$$

where

$$h = \varphi_{x'x'}^{-1} \varphi_{x'z} \text{ and } x' = x - \bar{x}.$$

The minimum mean error is the diagonal of the expression

$$\begin{aligned}
\min \varphi_{ee} &= \varphi_{z'z'} - h^t \varphi_{x'x'} h \\
&= \varphi_{z'z'} - \varphi_{zx'} \varphi_{x'x'}^{-1} \varphi_{x'z}.
\end{aligned}$$

In the case of independent additive noise, the minimum error is

$$\begin{aligned} \min \varphi_{ee} &= \varphi_{z'z'} - \varphi_{z'z'} (\varphi_{z'z'} + \varphi_{nn})^{-1} \varphi_{z'z'} \\ &= \varphi_{z'z'} (\varphi_{z'z'} + \varphi_{nn})^{-1} \varphi_{nn} \\ &= \varphi_{nn} (\varphi_{z'z'} + \varphi_{nn})^{-1} \varphi_{z'z'} \\ &= \varphi_{nn} h \end{aligned}$$

The minimum error in the one sample picture of a previous section is

$$e_{\min}^2 = \frac{\rho_{11} \sigma_{11}}{\rho_{11} + \rho_{11}}$$

In the two sample pictures, the minimum error is

$$\overline{e}_{1 \min}^2 = \frac{\rho_{11} (\sigma_{11} \sigma_{22} - \sigma_{12}^2) + \sigma_{11} (\rho_{11} \rho_{22} - \rho_{12}^2)^2}{(\rho_{11} + \sigma_{11}) (\rho_{22} + \sigma_{22}) - (\rho_{12} + \sigma_{12})^2}$$

and

$$\overline{e}_{2 \min}^2 = \frac{\rho_{22} (\sigma_{11} \sigma_{22} - \sigma_{12}^2) + \sigma_{22} (\rho_{11} \rho_{22} - \rho_{12}^2)^2}{(\rho_{11} + \sigma_{11}) (\rho_{22} + \sigma_{22}) - (\rho_{12} + \sigma_{12})^2}$$

In the case of stationary correlation functions in rotation, the minimum error for the two coefficient rectangular grid is

$$\min \varphi_{ee} = \varphi_{z'z'} - \omega_0 h_0 - 4\omega_1 h_1$$

and for the two coefficient hexagonal grid,

$$\min \varphi_{ee} = \varphi_{z'z'} - \omega_0 h_0 - 6\omega_1 h_1$$

Trivial Processing

The uniqueness of the solution to the optimum processing can be used to identify the type of pictures which require no linear processing other than possibly a change in scaling. The linear process which is restricted to a change in scaling the original picture is

$$y = x'd + \bar{z}$$

where d is a diagonal matrix of constants which amplifies each sample independently. In the case where the matrix d represents the optimum linear process,

$$d = \varphi_{x'x'}^{-1} \varphi_{x'z}$$

Since d is the unique solution, the only pictures for which a change in scaling is the optimum process are where there exists a diagonal matrix d such that

$$\varphi_{x'x'} d = \varphi_{x'z}$$

This means that a necessary and sufficient condition is that each column of $\varphi_{x'x'}$ is some multiple of the corresponding column of $\varphi_{x'z}$.

$$[\text{ith column of } \varphi_{x'x'}] d_i = [\text{ith column of } \varphi_{x'z}] \quad \text{every } i$$

The mean error in this case is

$$\overline{e^2}_{\min} = \varphi_{z'z'} - d \varphi_{x'z}$$

As an example, the processing of a picture z which has added to it correlated noise having a correlation function φ_{nn} equal to a multiple k of the correlation function $\varphi_{z'z'}$, of the picture z would be processed by a simple change in scaling. In this case,

$$\varphi_{x'x'} = \varphi_{z'z'} + \varphi_{nn} = (1+k) \varphi_{z'z'}$$

and

$$\varphi_{x'z} = \varphi_{z'z} = \varphi_{z'z'}$$

so that the optimum process is

$$y = x' \frac{1}{(1+k)} + \bar{z}$$

the minimum error is then

$$\overline{e^2}_{\min} = \frac{k}{1+k} \varphi_{z'z'}$$

Orthogonal Pre-processing

Another approach to linear processing results from the construction of an orthogonal matrix Q column-wise composed of the normalized eigenvectors of the correlation matrix $\varphi_{x'x'}$. In particular,

$$\varphi_{x'x'} Q = Q \lambda$$

where λ is the diagonal matrix of eigenvalues λ_i corresponding to the eigenvectors q_i . The normal orthogonal property of the eigenvectors indicates that

$$Q^T Q = I \text{ or } Q^{-1} = Q^T$$

Linear pre-processing of the picture x with the matrix Q produces the picture $x_q = xQ$ which has the auto-correlation function

$$\varphi_{x_q x_q} = \overline{x_q^t x_q} = \overline{Q^T x^t x Q} = Q^T \varphi_{xx} Q = \lambda$$

and a crosscorrelation function

$$\varphi_{x_q z} = \overline{x_q^t z} = Q^t \overline{x^t z} = Q^t \varphi_{xz}$$

Subsequent optimum linear processing of the new picture x_q requires that

$$y = x_q' \lambda^{-1} \varphi_{x_q z} + \bar{z}$$

the mean error is then

$$e_{\min}^2 = \varphi_{z'z'} - \varphi_{zx_q} \lambda^{-1} \varphi_{x_q z}$$

The use of orthogonal processing appears to be of use in pre-processing pictures to obtain the picture x_q , which is composed of samples having zero crosscorrelation and auto-correlation λ_1 . Presently, the experimental use of this type of pre-processing is limited by the necessity to obtain the matrix of eigenvectors Q .

A Stationary Example

A particularly useful experimental example of stationary linear processing is the recovery of a correlated picture from a composite picture of the correlated picture added to an uncorrelated picture. The composite picture could be the result of video transmission of a correlated picture being corrupted by additive noise having a frequency bandwidth corresponding to the sampling rate of the picture. Another example is the recovery of the uncorrelated picture from the composite picture of the sum of the correlated picture and uncorrelated picture. This second situation corresponds to a high-grade picture being corrupted by a fluctuating exposure level resulting in a correlated bias being added to the picture.

Since the pictures are zero-mean and independent, the crosscorrelation $\varphi_{x'z_1}$ between the composite picture x and the correlated picture z_1 is φ_1 , the correlation function of the correlated picture. A similar relationship holds for the independent picture z_2 .

$$\varphi_{xz_1} = \varphi_1 \qquad \varphi_{xz_2} = \varphi_2$$

The matrix equations to be solved are then

$$\varphi_{x'x'} h = \varphi_i \qquad i = 1, 2$$

Since the correlation functions $\varphi_{x'x'}$ and φ_i are stationary and symmetric, the solution h to the equations is also stationary and symmetric.

$$h = \begin{bmatrix} & & & & h_2 \\ & & & & h_2 \\ & h_{\sqrt{2}} & h_1 & h_{\sqrt{2}} & \\ h_2 & h_1 & h_0 & h_1 & h_2 \\ & h_{\sqrt{2}} & h_1 & h_{\sqrt{2}} & \\ & & & & h_2 \end{bmatrix}$$

By the direct application of the method for matrix multiplication the following set of equations is obtained:

$$\begin{array}{rcccccc} (2 + a^2) h_0 + & 4 h_1 + & 2 h_{\sqrt{2}} + & h_2 = & 2 & a^2 \\ h_0 + (13/4 + a^2) h_1 + & 2 h_{\sqrt{2}} + & h_2 = & 1 & 0 & \\ 1/2 h_0 + & 2 h_1 + (5/2 + a^2) h_{\sqrt{2}} + & h_2 = & 1/2 & 0 & \text{or} \\ 1/4 h_0 + & h_1 + & h_{\sqrt{2}} + (2 + a^2) h_2 = & 1/4 & 0 & \end{array}$$

Only one equation is needed for each of the elements of the matrix h . The other equations are identical to these and thus produce no new equations. Since the solution of this set of equations is the unique solution to the optimization of the linear processor, elements of h are zero at greater distance than the extent of the correlation functions $\varphi_{x',x'}$ and $\varphi_{x',z}$ of the experimental and desired pictures. In this case

$$h_d = 0 \quad d > 2$$

Experimental Results

Figure 8 is the composite picture representing the addition of the correlated picture, Figure 6, and the uncorrelated picture, Figure 7. The addition was performed with $a^2 = 1$ so that the correlated picture had a $\varphi_{oo} = 2\sigma^2$, the uncorrelated picture had a $\varphi_{oo} = \sigma^2$, and the composite had a $\varphi_{oo} = 3\sigma^2$. The correlated picture is thus the dominant picture in the composite picture.

The recovered correlated picture is Figure 9 and the recovered uncorrelated picture is Figure 10. As to be expected the dominant correlated picture is recovered with better relative precision than the uncorrelated picture. It is interesting to note the uncorrelated appearance of the recovered uncorrelated picture from the highly correlated composite picture.

The minimum recovery error for linear processing of additive pictures is the diagonal of the correlation function of the error picture.

$$\begin{aligned} \varphi_{ee} &= \varphi_1 (\varphi_1 + \varphi_2)^{-1} \varphi_2 \\ &= \varphi_1 h^{(2)} = \varphi_2 h^{(1)} \end{aligned}$$

where $h^{(2)}$ is the optimum processor used to recover the second picture and $h^{(1)}$ is the optimum processor used to recover the first picture. The recovery error

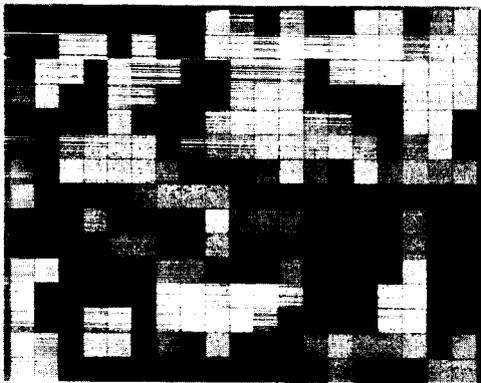


Figure 6. Desired Correlated Picture z_1

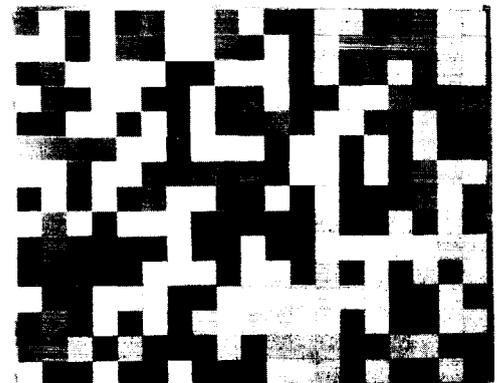


Figure 7. Desired Uncorrelated Picture z_2

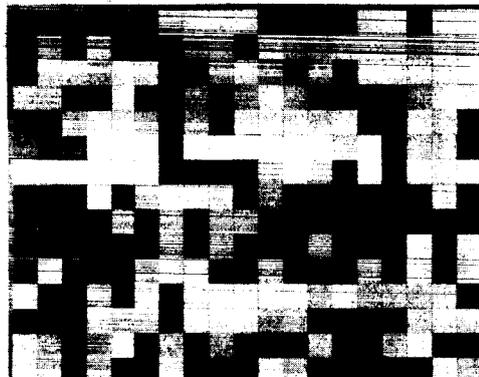


Figure 8. Composite Picture x of the Sum of the Correlated Picture z_1 and
Uncorrelated Picture z_2

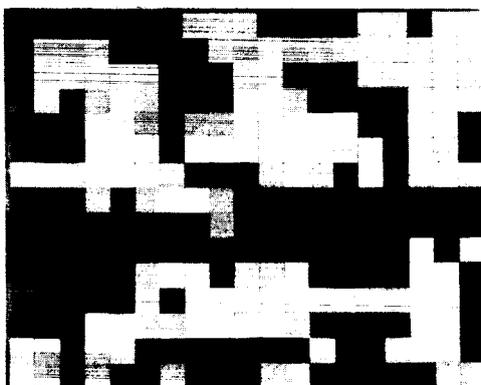


Figure 9. Recovered Correlated Picture y_1 from the Composite Picture

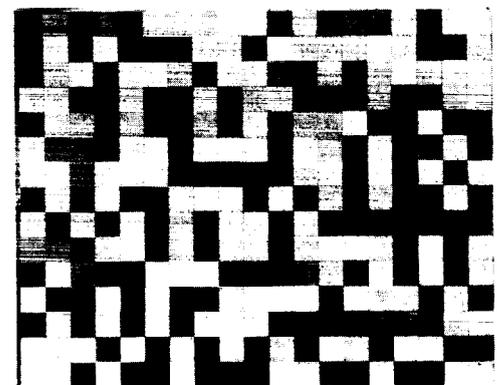


Figure 10. Recovered Uncorrelated Picture y_2 from the Composite Picture

is the same for either picture recovered. The correlated and uncorrelated pictures were both recovered with a theoretical error of $.52\sigma^2$. However, the desired correlated picture had a variance of $2\sigma^2$ while the desired uncorrelated picture had a variance of σ^2 , which indicates a better relative recovery of the correlated picture.

Behavior of the Linear Processor

A general solution of the set of linear equations of the optimum processor can be worked out through some fortunate cancellations and factorizations. For the recovery of the correlated picture, the h's are

$$h_0^{(1)} = \frac{1}{a^2+1} + \frac{a^2-3/4}{a^2+1} h_1^{(1)}$$

$$h_1^{(1)} = \frac{a^2 (a^4 + 9/4a^2 + 11/8)}{a^8 + 31/4a^6 + 131/8 a^4 + 205/16 a^2 + 95/32}$$

$$h_{\sqrt{2}}^{(1)} = \frac{a^2(a^2 + 3)(a^2 - 3/4)}{4(a^8 + 31/4 a^6 + 131/8 a^4 + 205/16 a^2 + 95/32)}$$

$$h_2^{(1)} = \frac{a^2(a^2 + 1/2)(a^2 - 3/4)}{4(a^8 + 31/4 a^6 + 131/8 a^4 + 205/16 a^2 + 95/32)}$$

$$= \frac{a^2 + 1/2}{a^2 + 3} h_{\sqrt{2}}$$

The h's for the recovery of the uncorrelated picture are closely related. They are

$$h_0^{(2)} = \frac{a^2}{1+a^2} + \frac{a^2 - 3/4}{a^2 + 1} h_1$$

$$h_1^{(2)} = -h_1^{(1)}$$

$$h_{\sqrt{2}}^{(2)} = -h_{\sqrt{2}}^{(1)}$$

$$h_2^{(2)} = -h_2^{(1)}$$

As to be expected the sum of the two recovered pictures y_1 and y_2 is equal to the sum x of the correlated picture z_1 and the uncorrelated picture z_2 .

$$y_1 + y_2 = x'h^{(1)} + x'h^{(2)}$$

$$x'\varphi_{x'x'}^{-1}\varphi_1 + x'\varphi_{x'x'}^{-1}\varphi_2$$

$$= x'\varphi_{x'x'}^{-1}(\varphi_1 + \varphi_2)$$

$$= x'$$

$$= z_1 + z_2$$

where $\varphi_{x'x'} = \varphi_1 + \varphi_2$

The behavior of the h's used to recover the correlated picture is plotted in Figure 11. Where the additive corruption is negligible, the processor relies principally on the picture element being processed to produce the processed element. ($h_0 \approx 1$, $h_1 \approx h_{\sqrt{2}} \approx h_2 \approx 0$). At higher levels of corruption the processor relies more on the elements near the element being processed rather than on the element itself.

The behavior of the h's in the recovery of the uncorrelated picture is plotted in Figure 12. It appears that at all levels of corruption, the processor makes use of the elements clustered around the element being processed.

For recovery of either picture the theoretical recovery error $\overline{e^2}$ is computed to be

$$\begin{aligned} \overline{e^2} &= a^2 h_0^2(1) \\ &= \frac{a^2}{a^2 + 1} + \frac{a^2(a^2 - 3/4)}{a^2 + 1} h_1^2(1) \end{aligned}$$

Figure 13 is a plot of this error, $\overline{e^2}$, normalized by the variance of the recovered correlated picture. If h is optimum, the variance of the recovered picture can be calculated by the formula

$$\begin{aligned} \text{diag } \varphi_{yy} &= \text{diag } \left(\varphi_{zz} - \varphi_{ee} \right) \\ &= 2\sigma^2 - a^2 h_0^2(1) \end{aligned}$$

The variance of the uncorrelated picture, $a^2 \sigma^2$, normalized by the variance of the composite picture $(2 + a^2) \sigma^2$ is also plotted in Figure 13 as an indication of the amount of error present in the composite picture x from which the processed picture y is obtained.

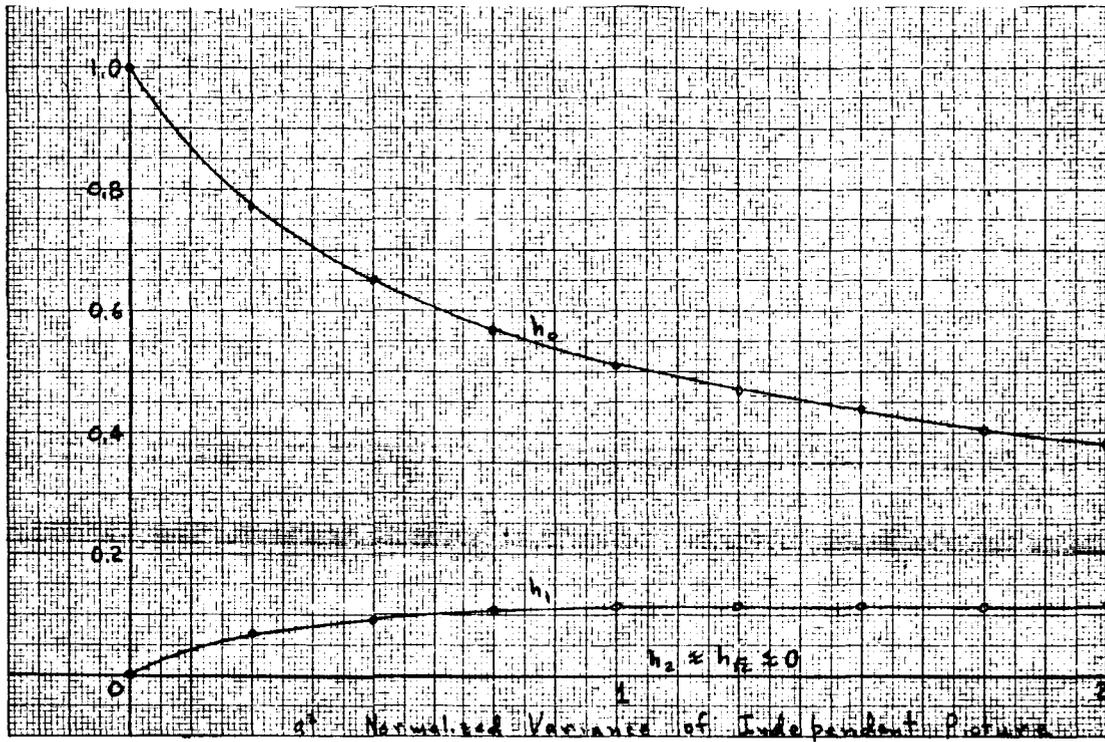


Figure 11. Coefficients of the Optimum Linear Processor for the Recovery of the Correlated Picture

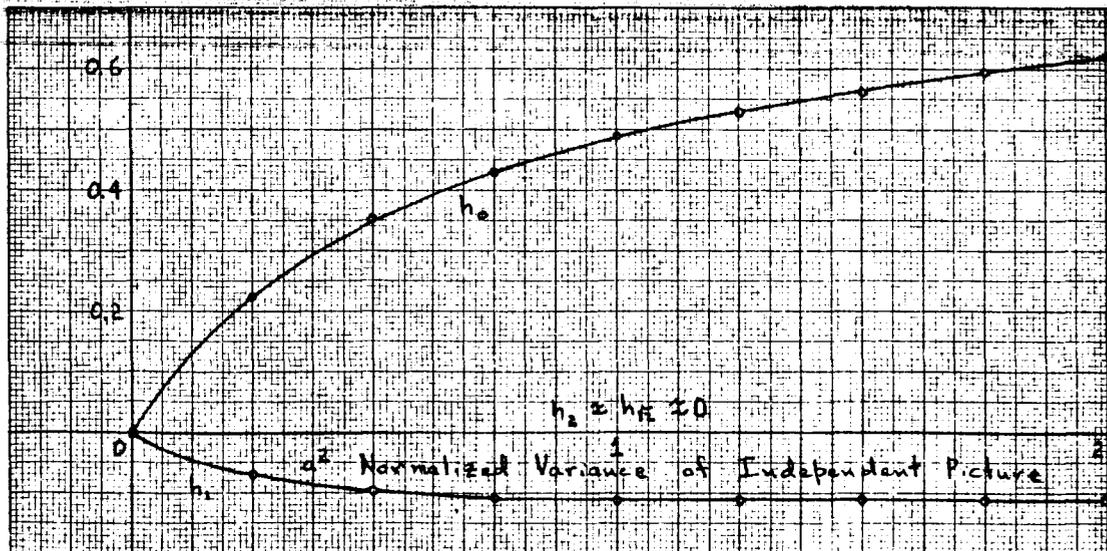


Figure 12. Coefficients of the Optimum Linear Processor for the Recovery of the Uncorrelated Picture

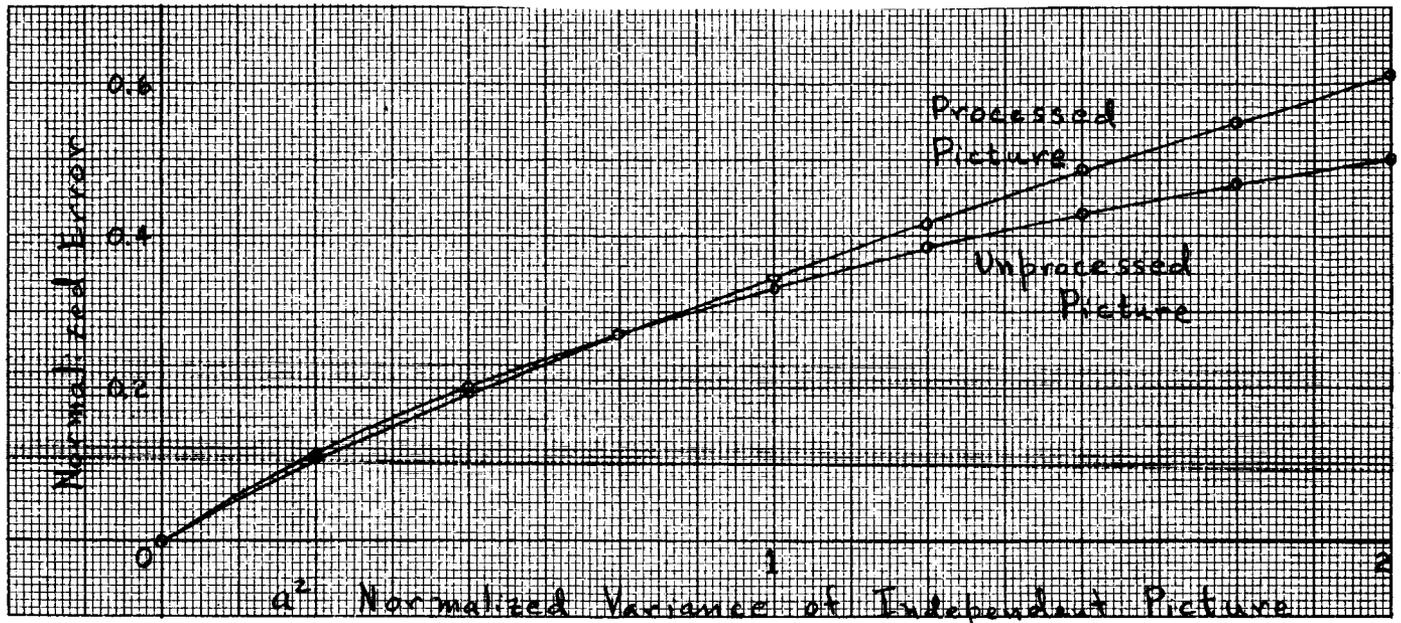


FIGURE 13. Normalized Recovery Error

Processing the Crater Archimedes

Further experimental work was performed on a sampled picture of the crater Archimedes, Figure 14. Figure 15 is a composite picture of the crater and additive noise of half the variance of the original picture. Optimum linear processing then produced Figure 16, which appears to be a decided improvement over the corrupted picture of the crater.

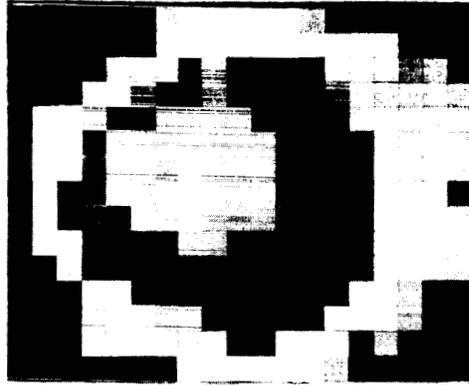


Figure 14. The Crater Archimedes in 285 Samples of Six Shades of Brightness

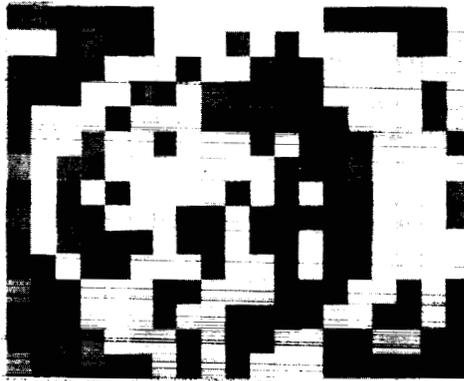


Figure 15. Corruption of the Picture of Archimedes with Additive Independent Noise Having a Variance of Half the Variance of the Original Picture

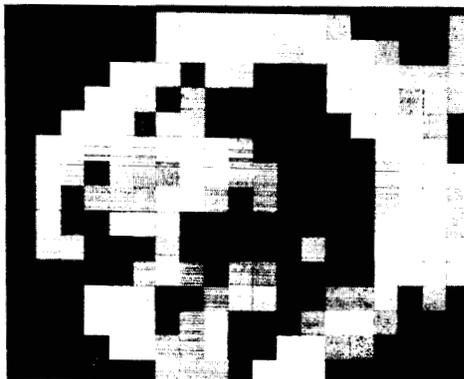


Figure 16. Recovery of the Picture of the Crater Archimedes with an Optimum Linear Processor

116

The picture of the crater had the normalized correlation function

$$\varphi_{z'z'} = \begin{bmatrix} & & & & .08 \\ & & & & .32 & .54 & .32 \\ .08 & .54 & 1 & .54 & .08 \\ & .32 & .54 & .32 \\ & & & & .08 \end{bmatrix}$$

which is plotted in Figure 17 as a function of distance.

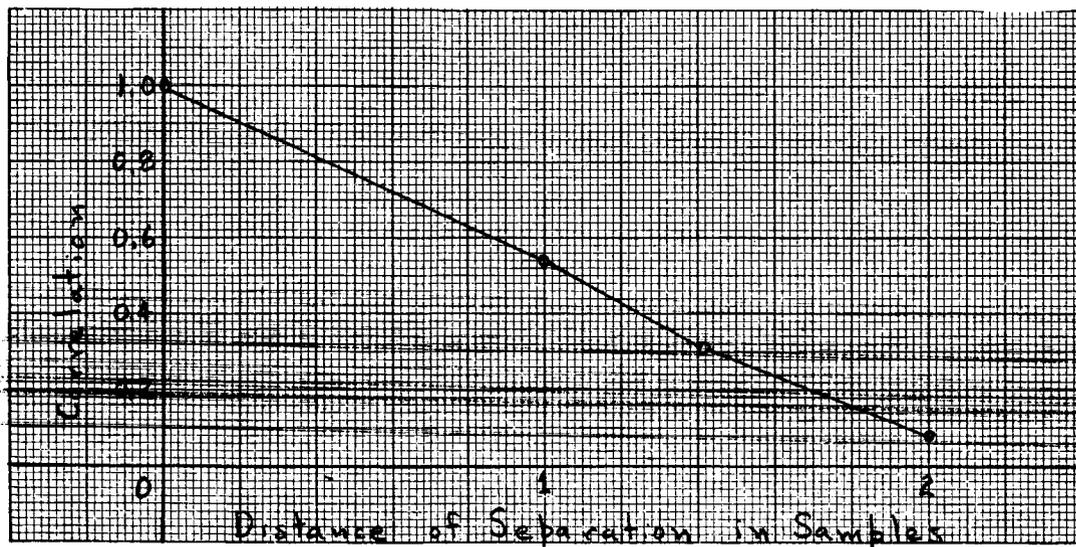


Figure 17. Correlation Function of the Crater Archimedes

Addition of an independent noise picture with a variance of half that of the original picture produces the following set of equations for the determination of the optimum processor.

$$1.50 h_0 + 2.16 h_1 + 1.28 h_{\sqrt{2}} + 0.32 h_2 = 1.00$$

$$0.54 h_0 + 2.22 h_1 + 1.08 h_{\sqrt{2}} + 0.54 h_2 = 0.54$$

$$0.32 h_0 + 1.08 h_1 + 1.66 h_{\sqrt{2}} + 0.64 h_2 = 0.32$$

$$0.08 h_0 + 0.54 h_1 + 0.64 h_{\sqrt{2}} + 1.50 h_2 = 0.08$$

which has the solution

$$h_0 = .47$$

$$h_1 = .12$$

$$h_{\sqrt{2}} = .04$$

$$h_2 = -.03$$

The large value of h_0 indicates a rather poor suppression of the error. However, the experimental results indicate a satisfactory retention of picture features. Perhaps, then, mean-square error is not a good indicator of damage done to a picture by additive noise.

Suppression of Errors

The experiment with artificial pictures produced very little suppression of independent additive errors. This is to be expected when it is considered that the number of processed samples was quite small and the picture was only correlated over a small distance.

An upper bound can be obtained for the amount of suppression of additive independent noise by assuming a highly correlated desired picture. The correlation function of the desired picture z is assumed to be σ^2 times the unit matrix of all 1's. The correlation function of the independent noise is assumed

to be $\sigma^2 a^2$ times the identity matrix of a 1 on the diagonal. All the equations for the coefficients of the processing matrix h are identical so that all the coefficients are equal. The equation for the single coefficient is then

$$(n + a^2) h_0 = 1 \quad \text{or} \quad h_0 = \frac{1}{n + a^2}$$

where there is a total of n coefficients in the matrix h . The variance of the error in the optimum recovery is then

$$\overline{e^2} = \frac{a^2}{n + a^2} \sigma^2$$

which should be normalized by the variance of the recovered picture

$$\begin{aligned} \varphi_{yy} &= \varphi_{zz} - \varphi_{ee} \\ &= \sigma^2 - \frac{a^2}{n + a^2} \sigma^2 \\ &= \frac{n}{n + a^2} \sigma^2 \end{aligned}$$

Thus

$$\frac{\overline{e^2}}{\varphi_{yy}} = \frac{a^2}{n}$$

In contrast, the variance of the error $a^2 \sigma^2$ in the original picture normalized by the variance $(1+a^2) \sigma^2$ of the picture is

$$\frac{\varphi_{nn}}{\varphi_{xx}} = \frac{a^2}{1 + a^2}$$

These two relations are plotted in Figure 18 for a processor operating on five samples. A suppression of the independent noise by a factor of 5 in variance is obtained for small relative levels of noise. As the noise becomes larger the linear processing breaks down and produces no improvement in the relative quality of the picture.

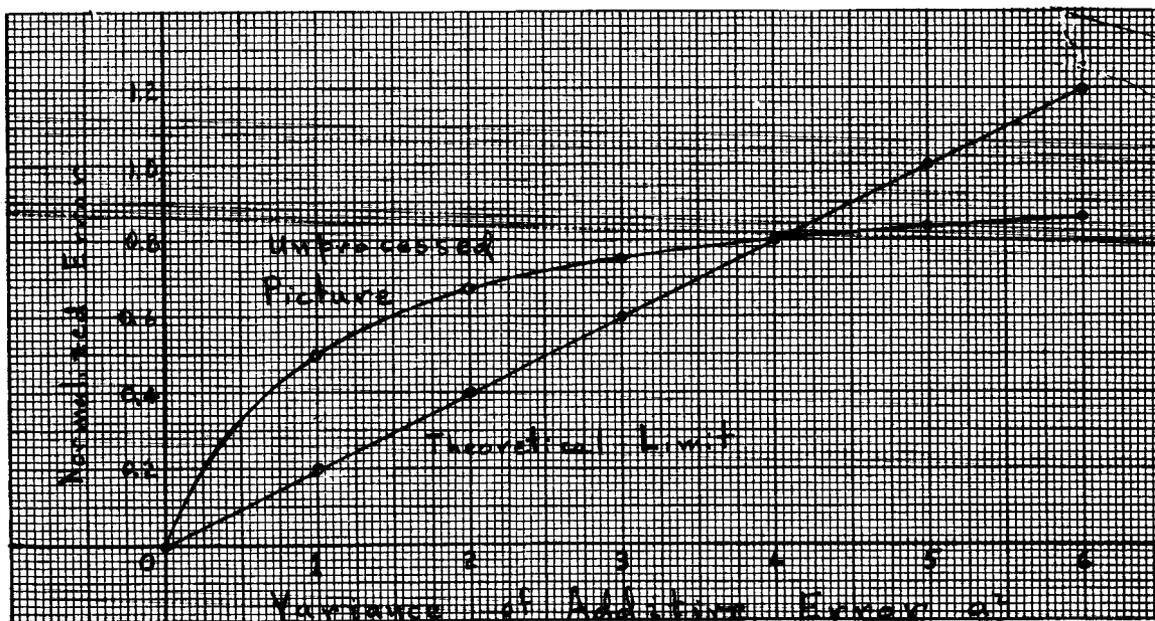


Figure 18. Theoretical Limits to the Suppression of Additive Noise by a Linear Processor Operating on 5 Picture Samples

The upper bound on the suppression of independent noise indicates that the suppression in magnitude of error is inversely proportional to the square root of the number of samples. Thus, in order to obtain a suppression in magnitude by a factor of 10 the processor requires at least 100 samples to

reconstruct one element in the picture. The actual number of samples is presently unknown, as the bound given above is computed under the assumption of a perfectly correlated picture. Whether or not it is possible to theoretically compute the error for a more realistic type of correlation is unknown. The computation will require a trick in the inversion of the correlation matrix.

Minimization with a Constraint

In the optimum processing of pictures it is often desirable to scale the resultant pictures to a particular contrast by having the standard deviation of the resultant picture equal to the standard deviation of the desired picture. Minimization of the mean square error between the resultant picture and the desired picture subject to the constraint that the resultant picture have the same standard deviation as the desired picture results in a linear scaling of the picture produced under an unconstrained optimization.

The standard deviation is given by the diagonal terms of the correlation function of the zero-mean variables.

$$\text{diag. } \varphi_{y'y'} = \text{diag. } \varphi_{z'z'} = \text{constant}$$

The minimization is then of the diagonal elements of the matrix

$$U = \varphi_{ee} + \lambda \varphi_{y'y'}$$

where λ is an undetermined constant and φ_{ee} is the error correlation matrix

$$\varphi_{ee} = \overline{(y - z)^t (y - z)}.$$

The assumption that the output y is obtained by a linear process on the input x requires that

$$y = x'h + k \qquad y' = x'h$$

where $x' = x - \bar{x}$ and $y' = y - \bar{y}$.

Variation of the matrix U produces the result

$$\begin{aligned} \partial U &= \partial \left[\frac{(y - z)^t (y - z) + \lambda y'^t y'}{2} \right] \\ &= 2 \left[(\partial h^t x'^t + \partial k^t) (x'h + k - z) + \lambda \partial h^t x'^t x'h \right] \\ &= 2 \partial h^t \left[\varphi_{x'x'} h(1+\lambda) - \varphi_{x'z} \right] \\ &\quad + 2 \partial k^t [k - \bar{z}] \end{aligned}$$

A sufficient condition for zero variation of the diagonal elements of the matrix U is that

$$\varphi_{x'x'} h(1+\lambda) = \varphi_{x'z}$$

$$\text{and} \quad k = \bar{z}$$

This is the same solution that was obtained for the unconstrained optimization except for a scaling of the output pictures y , producing the desired standard deviation.

Pre-Processing with Prediction

One method of transmitting a picture with a sequence of independent samples, is to sequentially scan the picture, predict the next element to be scanned, measure the prediction error, and transmit the error from which the picture can be reconstructed with a similar prediction.

Theoretically this method will precisely produce a replica of the original picture. In those cases where it is possible to do high quality prediction, considerable reduction can be obtained in the total amount of data which must be transmitted.

Quantization of the transmitted sequence of errors introduces a quantization drift into the reconstruction process. This drift prevents the system from running for any reasonable length of time without having to be reset.

This section describes a method of prediction and reconstruction from quantized errors in such a manner as to avoid the quantization drift in the reconstruction of the data.

The picture is to be reconstructed as in Figure 19 where the next element in the scan is predicted from the previously reconstructed picture and modified by adding the error to this prediction.

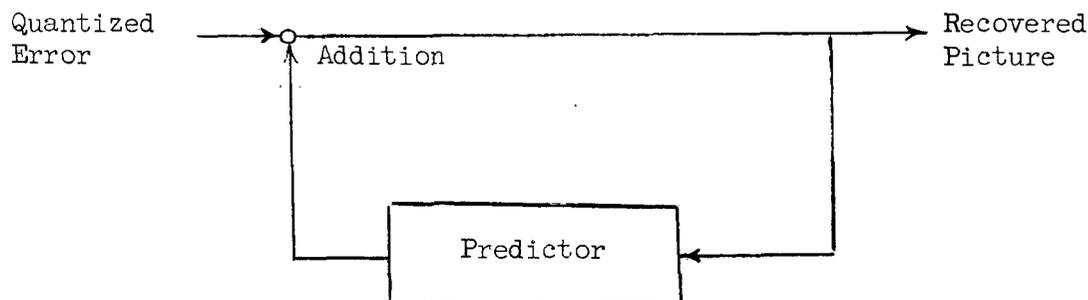


Figure 19. The Reconstruction of a Picture from Quantized Prediction Error

The prediction error is obtained as in Figure 20 by quantizing the error between the actual picture and the predicted picture based upon a reconstruction from the quantized error. The picture is first recovered in the same manner as it is reconstructed in Figure 19. The prediction of the picture is then based upon the reconstructed picture rather than the picture which is to be transmitted. The error is obtained by subtracting the predicted picture from the actual picture.

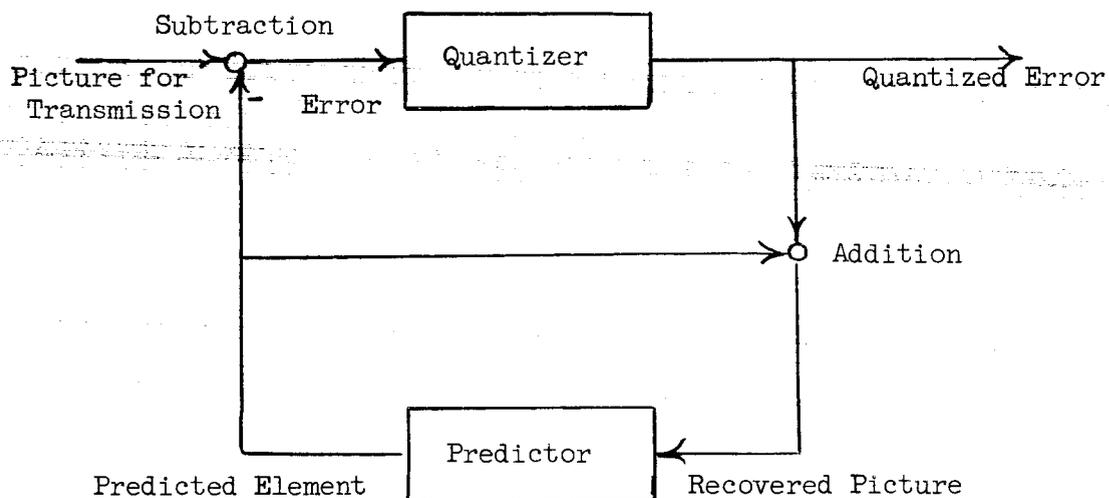


Figure 20, Processing a Picture to Obtain a Quantized Prediction Error

The quantized errors are theoretically a sequence of linearly independent samples. A later section presents a FORTRAN subroutine for transforming a sequence of independent samples into a binary Huffman Code¹⁰ having a maximum entropy per digit. This code essentially minimizes the amount of binary bits which must be transmitted in transmitting a picture. Further use of various block coding techniques as may be found in Reference 22 will produce the necessary reliability in the transmission of the binary data to insure reliable reconstruction of the picture.

where X marks the center of the matrix. The matrix product $\phi_{X'X'}h$ is obtained from the formula

$$\begin{aligned}\phi_{X'X'}h &= \sum_r \phi_{X'X'}(-r) h(r+s) \\ &= \sum_r \phi_{X'X'}(r) h(r+s)\end{aligned}$$

where the predicting processor has the matrix

$$h = \begin{bmatrix} & & & & & h_6 \\ & & & & & \\ & & h_3 & h_2 & h_4 & \\ & & & & & \\ h_5 & h_1 & X & & & \\ & & & & & \\ & & & & & \end{bmatrix}$$

where again X marks the center of the matrix. Equating the matrix product to the crosscorrelation produces the following set of equations for the determination of the coefficients in the matrix of the processor h.

$$\begin{aligned}2 h_1 + 1/2h_2 + h_3 + h_5 &= 1 \\ 1/2h_1 + 2 h_2 + h_3 + h_4 + h_6 &= 1 \\ h_1 + h_2 + 2 h_3 + 1/4h_4 + 1/2h_5 + 1/2h_6 &= 1/2 \\ h_2 + 1/4h_3 + 2 h_4 + 1/2h_6 &= 1/2 \\ h_1 + 1/2h_3 + 2 h_5 &= 1/4 \\ h_2 + 1/2h_3 + 1/2h_4 + 2 h_6 &= 1/4\end{aligned}$$

Solution of this set of equations would then produce the coefficients for the predictor.

An approximate solution can easily be obtained by assuming that $h_4 = 0$ so that $h_1 = h_2$ and $h_5 = h_6$. The third order set of equations results in the solution

$$h = \begin{bmatrix} & & & -1/2 \\ & & -1/6 & 2/3 & 0 \\ -1/2 & 2/3 & X & & \end{bmatrix}$$

The power in the prediction error is obtained by the formula

$$\overline{e^2} = \text{diag} \{ \varphi_{z'z'}, -h^t \varphi_{x'x'} h \}$$

Quantization of Pictures

Experimentally a picture is sampled on a regular grid. It is customary in high quality pictures to use anywhere from 32 to 128 brightness levels in the quantization of the samples. Due to the limitation in experimental equipment, the pictures presented in this report are quantized in 6 brightness levels (black, 20%, 40%, 60%, 80%, and white). The shade of grey g can be computed from the formula

$$g = -\log_{\sqrt{2}} b$$

where b is the brightness level. The grey scale and brightness levels have the relation

$$\begin{array}{rcccccccccccccccc} g & = & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\ b & = & 1 & \frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2\sqrt{2}} & \frac{1}{4} & \frac{1}{4\sqrt{2}} & \frac{1}{8} & \frac{1}{8\sqrt{2}} & \frac{1}{16} & \frac{1}{16\sqrt{2}} & \frac{1}{32} & \frac{1}{32\sqrt{2}} & \frac{1}{64} & \frac{1}{64\sqrt{2}} & \frac{1}{128} \end{array}$$

A geometrical pattern is formed by the relative brightness of the elements of the picture rather than by the precise brightness levels. The tone of the picture is the mean value of the brightness levels. The standard deviation is a measure of the contrast. In numerous geometrical cases, it is desirable to represent a set of pictures with a uniform mean and standard deviation of brightness levels.

The processing of pictures with linear and non-linear processors produces a set of numbers which normally extend beyond the 0-1 range of the brightness levels. Several methods are available for quantizing a set of numbers so that they can be represented in a finite number of brightness levels.

One method of representing a picture with quantized brightness levels is to distribute the quantization levels so that all levels are represented with equal frequency of occurrence.

Another method of representing a picture is to linearly distribute the quantization steps over the range of sample values. This is the method normally used in quantizing the samples of an experimental picture. However, pictures which are the result of numerical processing, tend to have a few excessively small and large values which produce an excessive range. Linearly distributing the quantization steps over this excessive range produces a picture with most of the quantization steps near 50%. These pictures have very little contrast.

The quantization of the pictures in this report is based on a linear system which distributes the four center shades of brightness (20%, 40%, 60%, 80%) between plus and minus one standard deviation in brightness about the mean value of the picture. All values below and above one standard deviation are

made respectively black and white. Figure 21 depicts this type of quantization as applied to uniform and Gaussian distributions. For these two distributions the mean frequency of occurrence of brightness levels is

<u>Brightness Level</u>	<u>Uniform Distribution</u>	<u>Gaussian Distribution</u>
Black	.212	.159
20%	.144	.150
40%	.144	.191
60%	.144	.191
80%	.144	.191
White	.212	.159

The frequency of occurrence of each of the six levels is very close to $1/6 = .167$.

The correlated picture of Figure 3 had the distribution of brightness levels shown in Figure 22. Linear quantization between the standard deviations produces the results of Figure 3 which clearly show the correlation of the picture.

The principal disadvantage of this method of representation is in the comparison of different pictures having different standard deviations. Pictures are presented as though they had the same standard deviations. The addition of the pictures in Figure 6 and 7 to give that of Figure 8 is an example. Figure 6, 7, and 8 have relative standard deviations of $\sqrt{2}$, 1 and $\sqrt{3}$. Their standard deviations are normalized so that a relative comparison of brightness levels cannot be directly made. The method, however, does provide a simple way of representing geometrical patterns with a high degree of contrast in a fairly linear manner.

Figures 23 and 24 are examples of this type of quantization in the representation of the crater Eratosthenes. The white area in the center is an illuminated inner wall which casts the dark shadow extending to the edge of the Figure. The measured correlation function is Figure 25.

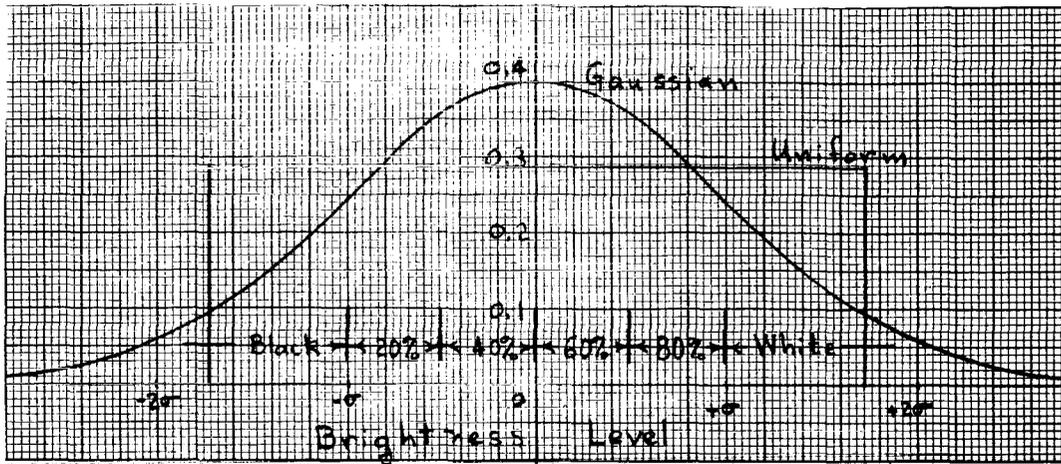


Figure 21. Quantization of a Picture into Six Brightness Levels in Relation to The Gaussian and Uniform Distribution

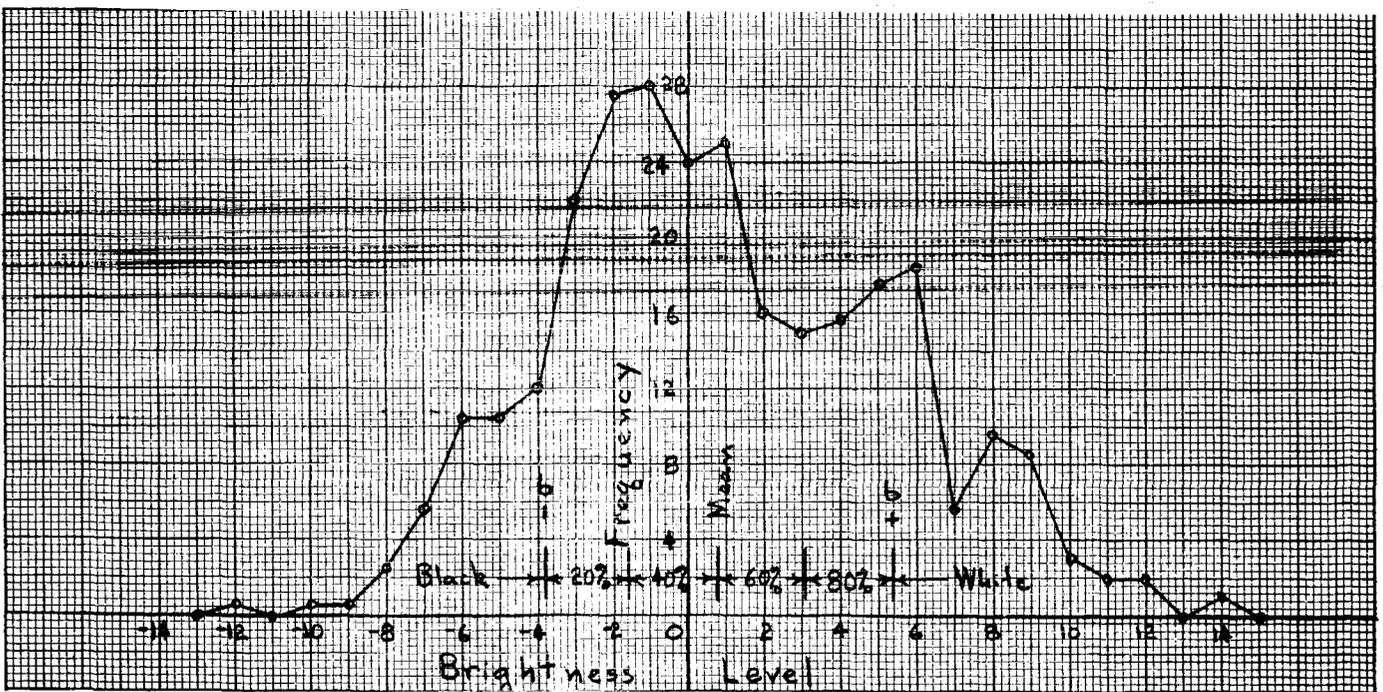


Figure 22. Distribution of Real Numbers of the Correlated Picture of Figure 3 Along with the Quantization Used in its Representation

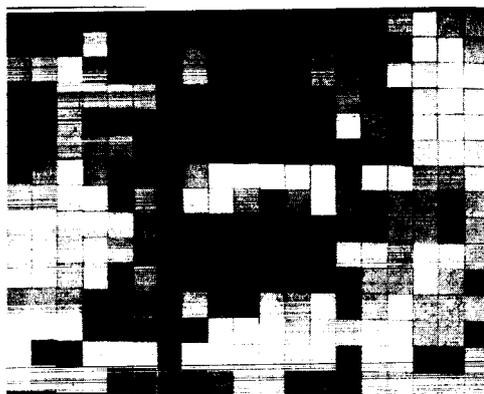


Figure 23. The Crater Eratostenes in 6 Brightness Levels

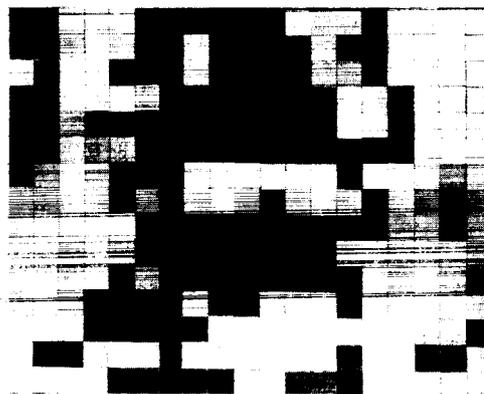


Figure 24. The Crater Eratostenes in 6 Brightness Levels Distributed About the Mean Plus and Minus One Standard Deviation

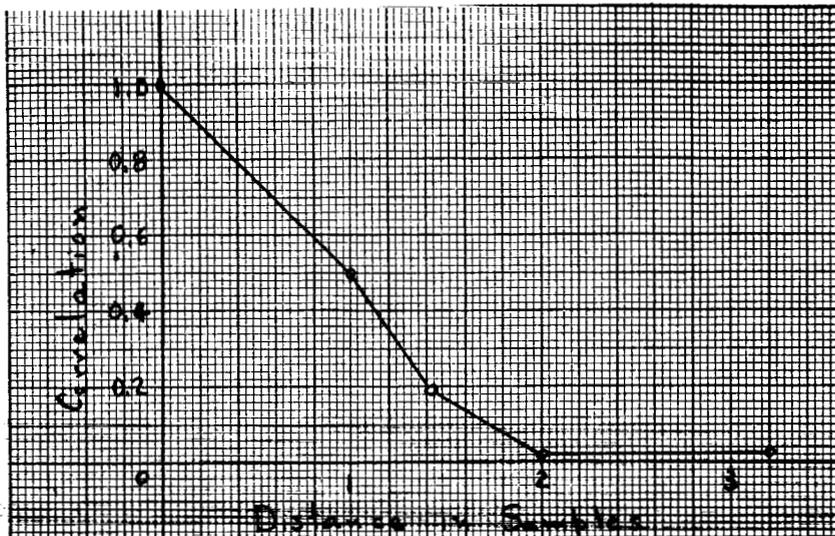


Figure 25, Stationary Correlation Function of the Crater Eratostenes

Machine Programming

The experimental work of this report was performed by hand computation. The amount of computation for a very simple picture is at the limit of practical hand computation.

One of the objectives of this theoretical work has been to cast it in a notation which can be easily programmed and run on a large automatic computing machine. Figure 26 represents the flow charts of the experiment in computing a linear processor for suppressing undesired errors. Figure 27 is the flow chart of a proposed experiment dealing with the removal of redundancy from a picture prior to transmission.

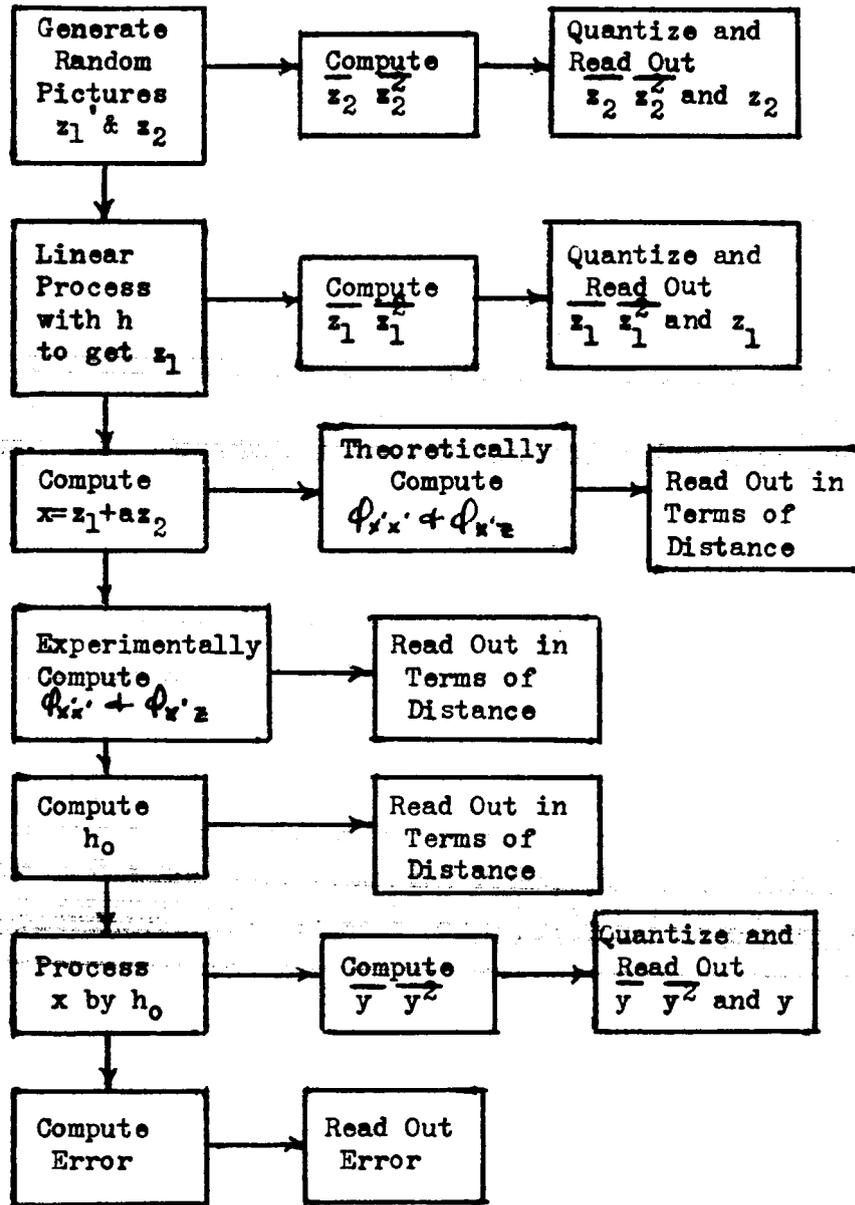


Figure 26. Flow Chart of an Experiment of the Suppression of Undesirable Noise

63

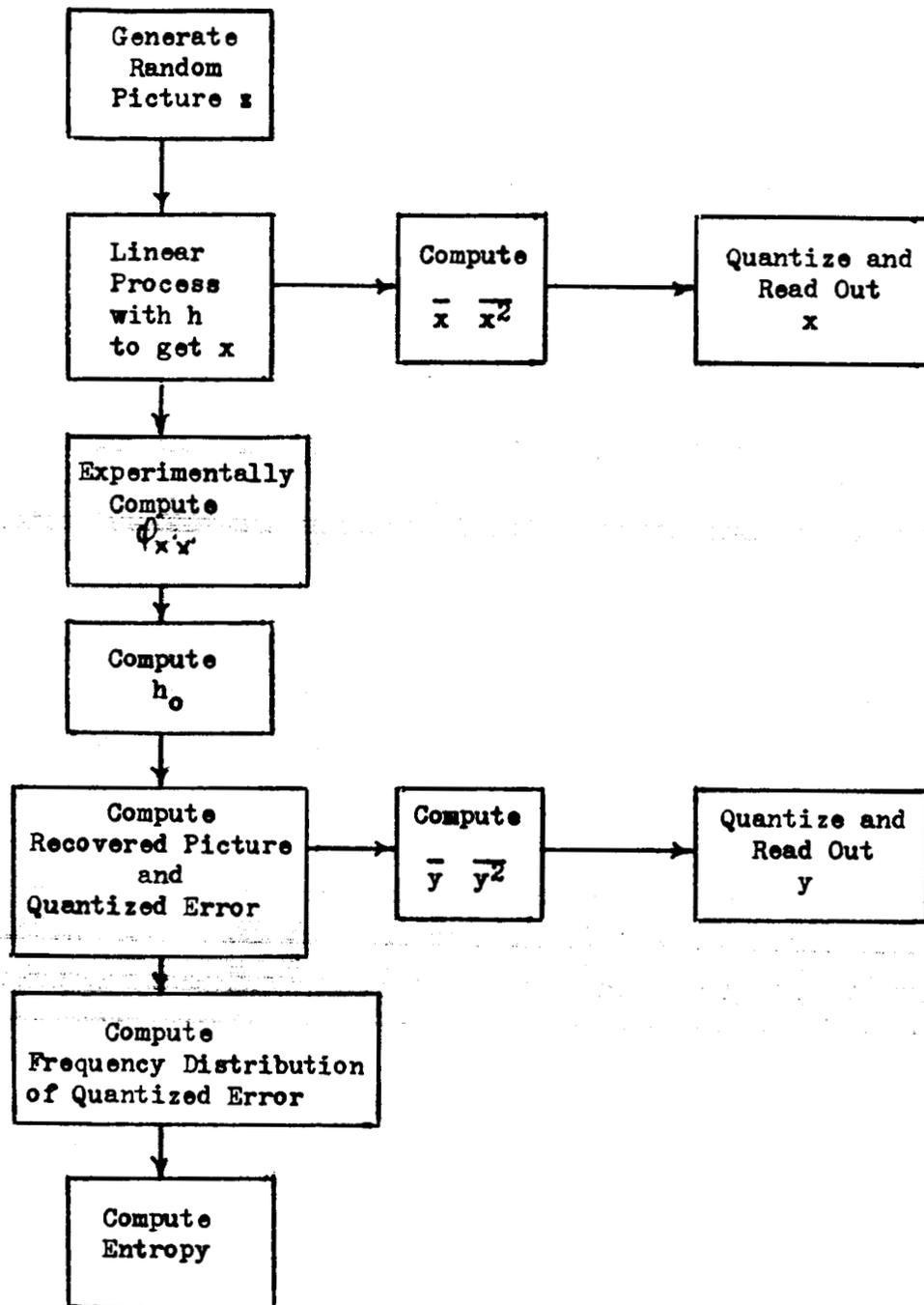


Figure 27. Flow Chart of an Experiment in Removing the Redundancy from a Picture

FORTRAN Subroutines

The following subroutines have been written to facilitate the experimental work in picture processing. The FORTRAN convention contained in Reference 21 for most IBM 709 and 7090 compilers is used.

Crosscorrelation of Two Pictures

The crosscorrelation of the picture X and Y is particularly useful in the design of optimum processors. In the following program these pictures are assumed to be of the same dimension as in Figure 28.

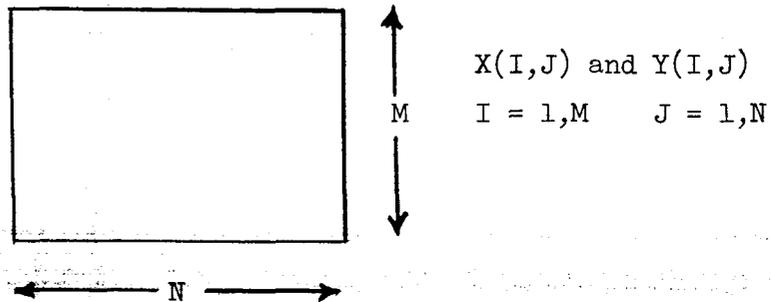


Figure 28, The Number of Elements in a Picture

The correlation to be calculated experimentally has the dimensions of Figure 29. The dimensions should be odd numbers due to the single value of the correlation at the origin, $(M_0, N_0) = \left(\frac{MC + 1}{2}, \frac{NC + L}{2} \right)$.

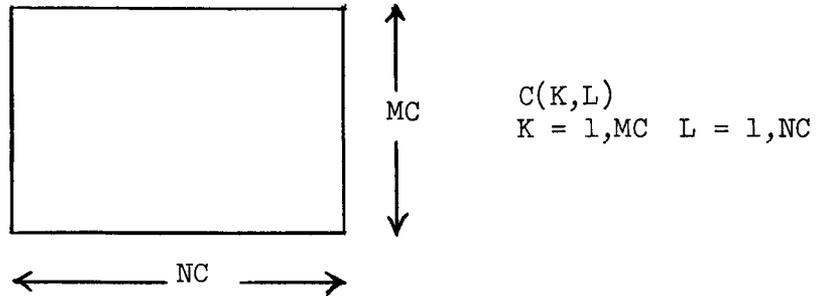


Figure 29, The Number of Elements in the Correlation Function

The crosscorrelation is computed by the formula

$$C(K,L) = \frac{1}{MN} \sum_{I=1}^M \sum_{J=1}^N (X(I,J) - XAVG) * (Y(I + K - MO, J + L - NO) - YAVG)$$

In those cases where the indices of Y are beyond the range of the picture, the average value of Y is used resulting in a zero term being added into the summation.

FORTTRAN Program for Calculating the Crosscorrelation of Two Pictures X and Y

```
SUBROUTINE CRCOR (C, MC, NC, X, XAVG, Y, YAVG, M, N)
DIMENSION X(40,40), Y(40,40), C(40,40)
DO 1 I = 1,M
DO 1 J = 1,N
X(I,J) = X(I,J) - XAVG
Y(I,J) = Y(I,J) - YAVG
T = M * N
MO = (MC+1)/2
NO = (NC+1)/2
DO 4 K = 1,MC
DO 4 L = 1,NC
C(K,L) = 0.0
DO 3 I = 1,M
DO 3 J = 1,N
IY = I+K-MO
JY = J+L-NO
IF (IY) 3,3,2
IF (JY) 3,3,2
IF (IY-M) 2,2,3
IF (JY-N) 2,2,3
2 C(K,L) = C(K,L) + X(I,J) * Y(IY,JY)
3 CONTINUE
4 C(K,L) = C(K,L)/T
RETURN
END
```

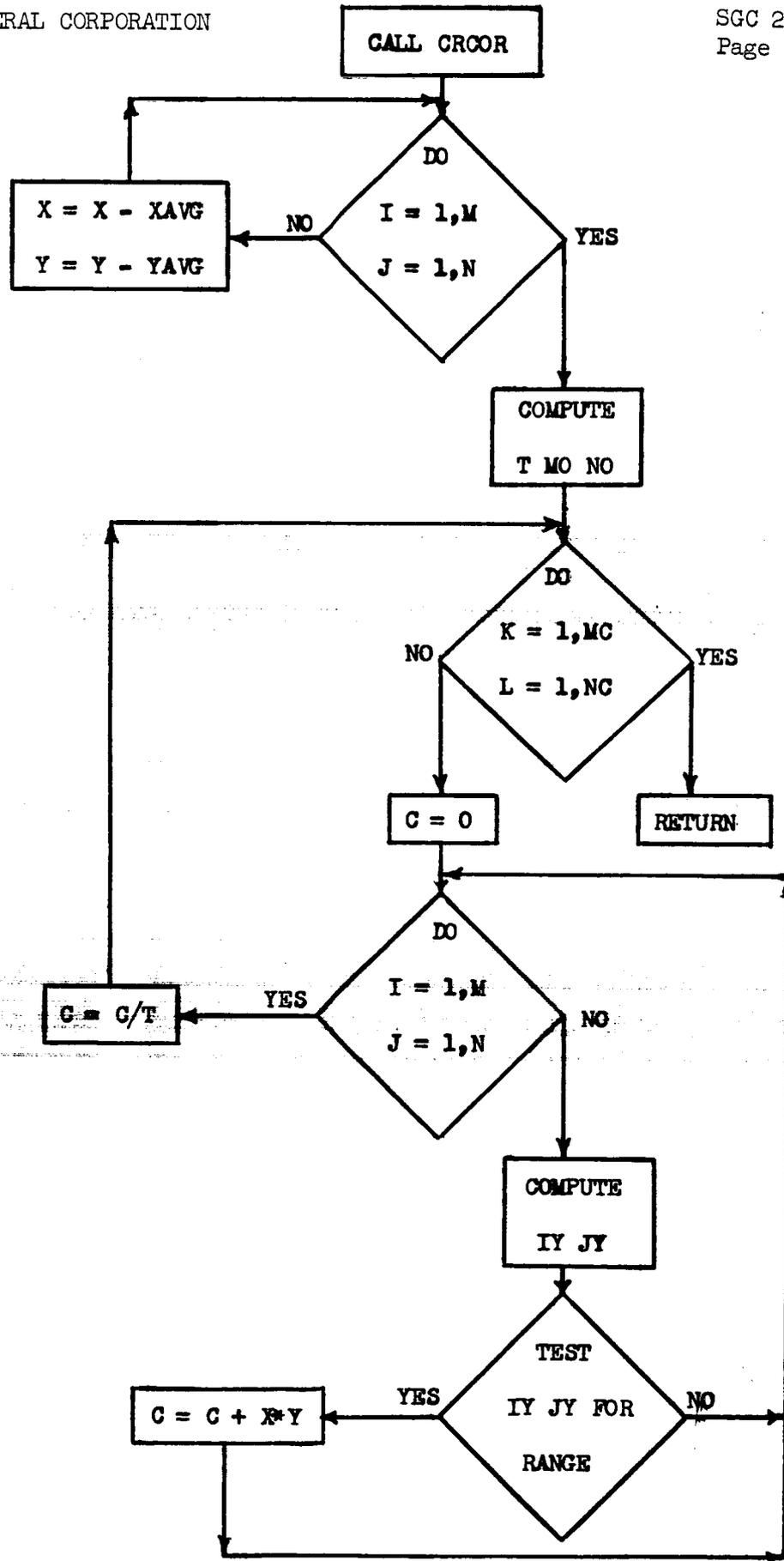


Figure 30. Flow Diagram for the Crosscorrelation of 2 Pictures

Stationary Matrix Multiplication

The matrix product

$$c = hg$$

$$\text{where } c(r) = \sum_s h(s) g(r-s)$$

is computed on the assumption that the dimension of each matrix is $M \times N$ with appropriate subscripts. The center is considered to be the integer part of

$$(M_0, N_0) = \left(\frac{M+1}{2}, \frac{N+1}{2} \right)$$

The computation of the product is then

$$C(I, J) = \sum_{K=1}^{M_H} \sum_{L=1}^{N_H} H(K, L) * G(I - M_{C0} + M_{G0} - K + M_{H0}, J - N_{C0} + N_{G0} - L + N_{H0})$$

where the summation is only over the mutual range of H and G.

FORTTRAN Program for Calculating the Matrix Product

```
SUBROUTINE MATRIX (H, MH, NH, G, MG, NG, C, MC, NC)
DIMENSION H(40,40), G(40,40), C(40,40)
MHO = (MH+1)/2
NHO = (NH+1)/2
MGO = (MG+1)/2
NGO = (NG+1)/2
MCO = (MC+1)/2
NCO = (NC+1)/2
MO = MGO + MHO - MCO
NO = NGO + NHO - NCO
DO 2 I = 1,MC
DO 2 J = 1,NC
C(I,J) = 0.0
DO 2 K = 1,MH
DO 2 L = 1,NH
KG = I - K + MO
LG = J - L + NO
IF (KG) 2,2,1
IF (LG) 2,2,1
IF (KG - MG) 1,1,2
IF (LG - NG) 1,1,2
1 C(I,J) = C(I,J) + H(K,L) * G(KG,LG)
2 CONTINUE
RETURN
END
```

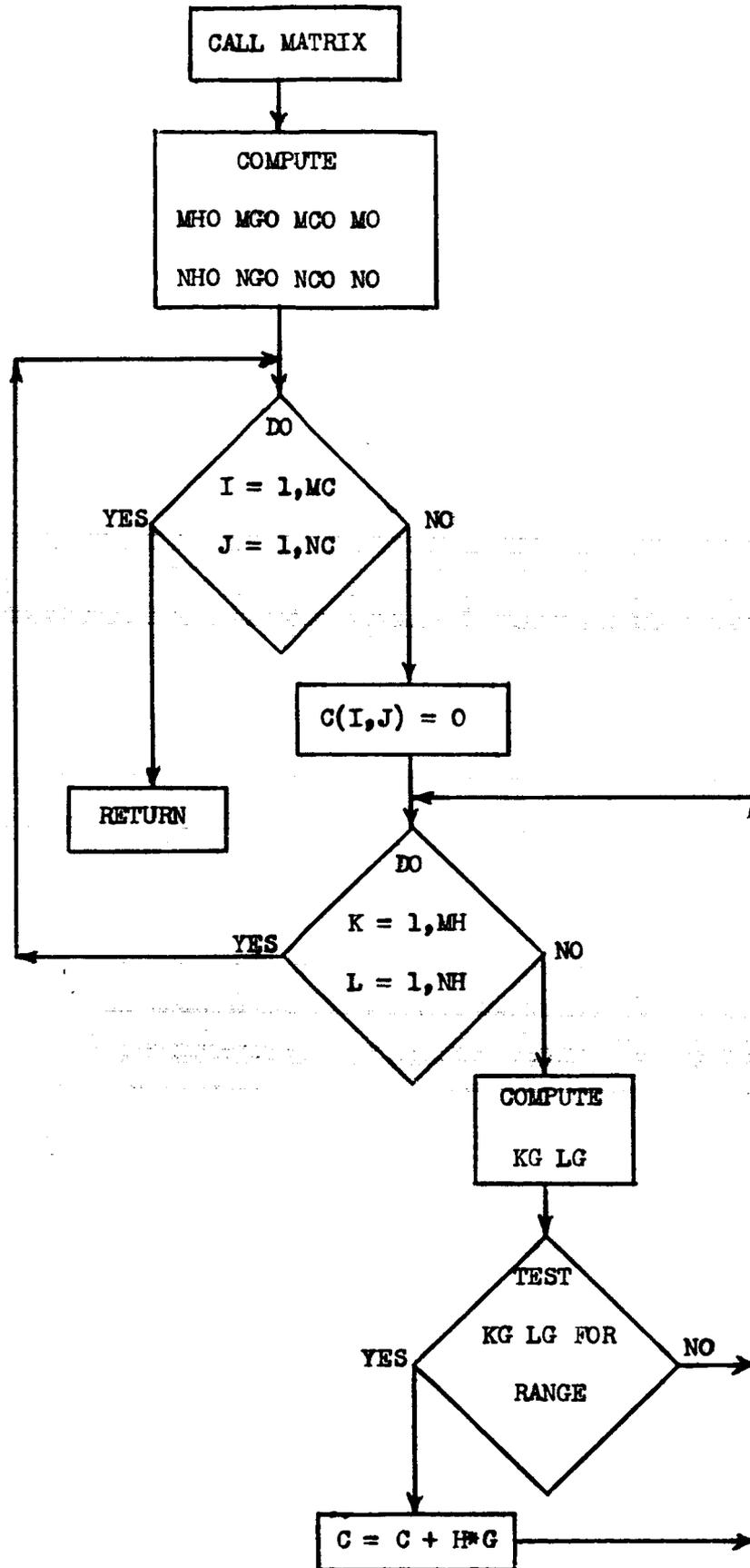


Figure 31. Flow Diagram of Stationary Matrix Multiplication

CONSTRUCTION OF A HUFFMAN CODE¹⁰

The conversion of a set of words into another set of words having maximum entropy per digit is a coding problem for which an exact solution is known in terms of an algorithm.

The frequency of occurrence of the original words are initially listed from the smallest to the largest. The following program assumes that the ordering process has already been performed. The first two words are distinguished by a 0 for the first and a 1 for the second. They are then combined and treated as a single word.

The algorithm proceeds by reordering the new set of words distinguishing the first two words with a 0-1 and combining. Each word is built up as a sequence of 0's and 1's depending upon the algorithm.

An index matrix provides the correspondence between the ordered set of frequencies and the original set of words. The matrix

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

indicates that the first, second, and fourth words are associated with the first frequency. The third and sixth are associated with the second frequency. With this type of index matrix it is then possible to know which of the original words are to be distinguished by 0's and 1's.

The program stores the code words reversed from right to left with a 1 signifying the end of the word. The code words can be obtained from the decimal representation by expressing the representation in binary, reversing the ordering and dropping the terminal 1.

A simpler program may be possible by using the concept of "low order pairing." A set of low order words are paired and moved up into the table without any reordering or moving of the rest of the table.

FORTTRAN PROGRAM FOR CONSTRUCTING A HUFFMAN CODE

```
1 READ 47 N
2 DIMENSION WORD(N) X(N) FREQ(N) INDEX(N,N)
3 READ 48 FREQ
4 DO 10 I = 1,N
5 DO 10 J = 1,N
6 IF (I-J) = 9,7,9
7 INDEX (I,J) = 1
8 GO TO 10
9 INDEX (I,J) = 0
10 CONTINUE
11 DO 12, I = 1,N
12 WORD (I) = 1.0
13 L = N
14 DO 44 M = 2,N
15 DO 20 J = 1,N
16 IF (INDEX(1,J))18,18,17
17 WORD (J) = 2.0 * WORD(J)
18 IF (INDEX(2,J)) 20,20,19
19 WORD (J) = 2.0 * WORD(J) + 1.0
20 CONTINUE
21 IF (L-2) 45,45,22
22 DO 23 J = 1,N
23 X(J) = INDEX(1,J) + INDEX(2,J)
24 F = FREQ (1) + FREQ (2)
25 DO 27 K = 3,L
26 IF (F-FREQ(K)) 27,27,28
27 CONTINUE
28 K = K-3
29 DO 33 I = 1,K
30 IK = I + 2
31 FREQ (I) = FREQ (IK)
32 DO 33 J = 1,N
33 INDEX (I,J) = INDEX (IK,J)
```

FORTRAN PROGRAM FOR CONSTRUCTING A HUFFMAN CODE (CONTINUED)

```
34 K = K+1
35 FREQ (K) = S
36 DO 37 J = 1,N
37 INDEX (K,J) = X(J)
38 K = K+1
39 L = L-1
40 DO 44 I = K,L
41 IK = I+1
42 FREQ (I) = FREQ (IK)
43 DO 44 J = 1,N
44 INDEX (K,J) = INDEX (IK,J)
45 PRINT 49 (WORD (I)/ I = 1,N)
46 STOP
47 FORMAT
48 FORMAT
49 FORMAT
```

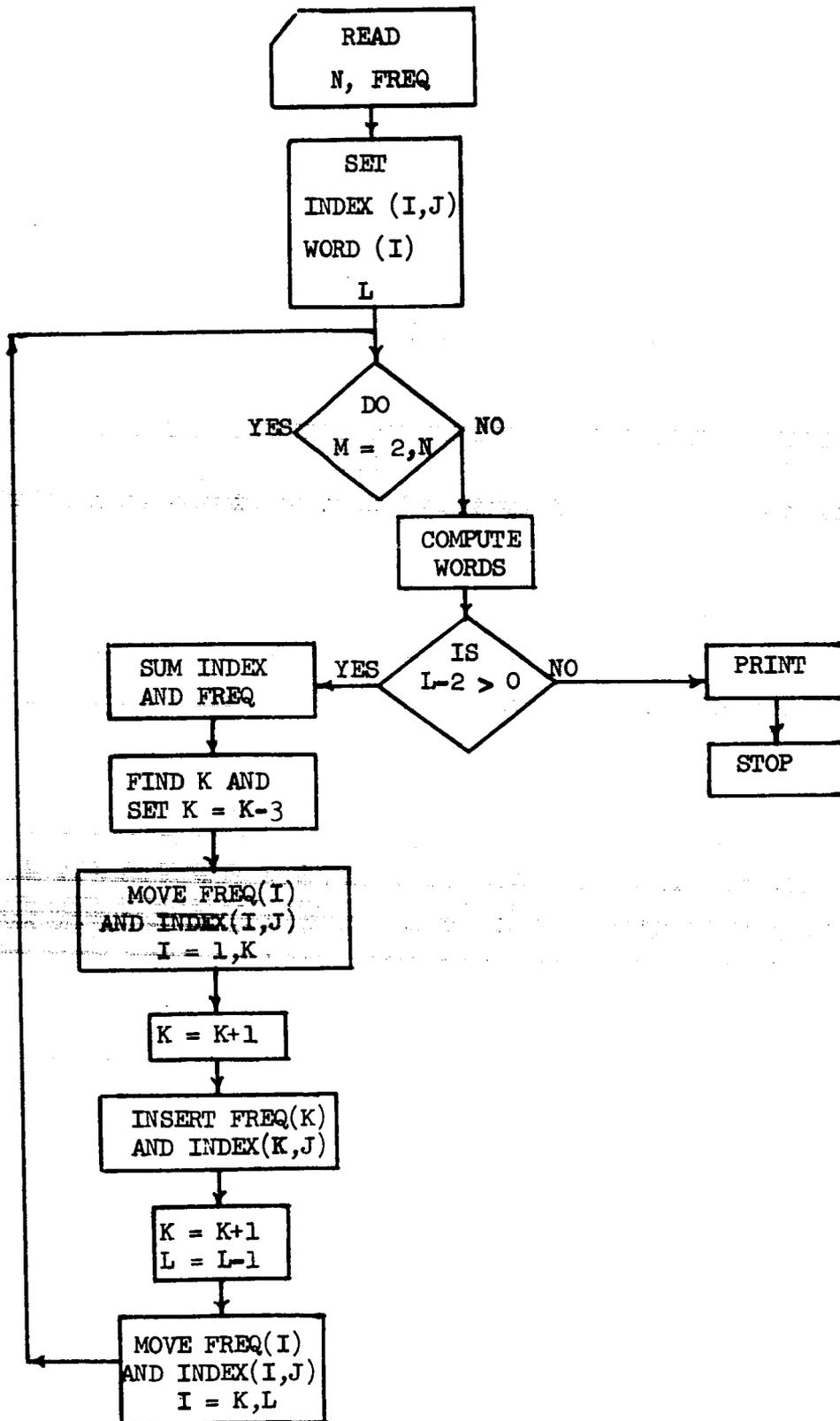


Figure 32, Flow Diagram for the Construction of a Huffman Code

Additional FORTRAN Subroutines

The following subroutines remain yet to be written:

1. Quantization for Display

This subroutine should compute the quantization constants and then quantize the picture, perhaps as it is being read out of the machine. Provision should be made for entering the subroutine with the constants already given as may be the case in fixed scaling.

2. Random Picture Generation

Several random number routines are needed for producing artificial pictures with known properties. Uniform, and Gaussian distributions with zero mean should be the most useful.

3. Theoretical Correlation Computation

This is the computation of the correlation matrix based on the linear processing of a picture having a specified correlation matrix.

4. Optimum Solution

Based on the input correlation function $\phi_{x'x'}$ and the cross-correlation function $\phi_{x'z}$, the optimum processor should be obtained. This is the solution of a set of simultaneous linear equations.

CONVERGENCE OF ITERATIVE SYSTEMS

One method of designing optimal processing systems is through iterative approximations. An approximation to the system is used to compute a better approximation iteratively until the optimal system is obtained. Sakrison⁽²³⁾ has considered an iterative design of a specific optimal system. This section considers some of the basic elements of his analysis of the convergence of a system to an optimal system.

NORMS

In order to consider the convergence of an iterative system to an optimal system it is necessary to have some measure of the error between the approximation X and the optimum system θ . The measure is known as the norm E_t where t denotes the sequence of approximations. A very useful norm is the average squared error in the case of a single variable or the average inner product of the error in the case of a multiple variable system.

$$E_t = \text{Average} \left[(X-\theta) \cdot (X-\theta) \right]$$

Other norms are based on weighting functions of the error such that

$$E_t > 0 \quad X \neq \theta$$

$$E_t = 0 \quad X = \theta$$

DIFFERENCE EQUATIONS

The behavior of the norm is of particular importance in describing the convergence of the approximation to optimal systems. In those cases where the norm can be described by the first order difference equation

$$\Delta E_t \leq a_t + b_t E_t$$

$$\Delta E_t = E_{t+1} - E_t$$

quite a bit can be said about the convergence based on the properties of the coefficients a_t and b_t .

The existence of the difference equation which can be explicitly solved is the most important idea of this approach. Where E_t^i is the solution of the difference equation

$$\Delta E_t^i = a_t + b_t E_t^i$$

and E_t is the actual norm which satisfies the difference relation

$$\Delta E_t \leq a_t + b_t E_t$$

the norm E_t is bounded by the solution E_t^i

$$E_t \leq E_t^i$$

and all t whenever $0 \leq 1 + b_t$

If this were not so, then there would exist a T for which

$$E_{T+1} > E_{T+1}^i$$

and

$$E_T < E_T^i$$

The difference equations produce the relations

$$E_{T+1} - E_T \leq a_T + b_T E_T$$

$$E_{T+1}^i - E_T^i = a_T + b_T E_T^i$$

Subtraction produces the inequality

$$(E_{T+1} - E_{T+1}^i) \leq (1+b_T)(E_T - E_T^i)$$

which is a contradiction under the assumption that $1 + b_T > 0$ since

$$E_{T+1} - E_{T+1}^i > 0$$

$$E_T - E_T^i \leq 0$$

Thus for all iterations t , the norm E_t is less than or equal to the solution E'_t of the difference equation.

$$E_t \leq E'_t$$

SOLUTION OF THE FIRST ORDER DIFFERENCE EQUATION

The first order difference equation

$$\Delta E'_t = a_t + b_t E'_t$$

has the general solution

$$E'_{t+1} = \sum_{i=0}^t a_i \prod_{j=i+1}^t (1+b_j) + E'_0 \prod_{j=0}^t (1+b_j)$$

In the particular case where b_j is the constant b and a_i is the constant a raised to the i th power, the solution has the closed form

$$E'_t = (1+b)^t \frac{1 - \left(\frac{a}{1+b}\right)^{t+1}}{1 - \frac{a}{1+b}} + E'_0 (1+b)^{t+1}$$

where $a_i = a^i$ $b_j = b$

This solution converges to zero approximately as

$$E'_t = \dots (1+b)^t$$

where

$$\frac{a}{1+b} < 1 \quad \text{and} \quad 0 < (1+b) < 1$$

It is easy to see that if $t = 0$, the general solution gives the correct solution

$$E'_1 = a_0 + E'_0 (1+b_0)$$

By induction, the solution for t is assumed correct

$$E'_t = \sum_{i=0}^{t-1} a_i \prod_{j=i+1}^{t-1} (1 + b_j) + E'_0 \prod_{j=0}^{t-1} (1 + b_j)$$

From the difference equation the solution for $t + 1$ is found to be

$$E'_{t+1} = a_t + E'_t (1 + b_t)$$

Substitution of E'_t produces the relation

$$\begin{aligned} E'_{t+1} &= a_t + \sum_{i=0}^{t-1} a_i \prod_{j=i+1}^t (1 + b_j) + E'_0 \prod_{j=0}^t (1 + b_j) \\ &= \sum_{i=0}^t a_i \prod_{j=i+1}^t (1 + b_j) + E'_0 \prod_{j=0}^t (1 + b_j) \end{aligned}$$

Which is the relation for the general solution of E'_t .

CONVERGENCE TO ZERO

Another property of the first order difference equation, which makes it useful in the analysis of iterative convergence, is that the solution E'_t converges to zero whenever the coefficients satisfy the constraints

$$\begin{aligned} 0 \leq 1 + b_j \leq 1 & \qquad \sum_{j=0}^{\infty} b_j = -\infty \\ a_i \geq 0 & \qquad \sum_{i=0}^{\infty} a_i = A < \infty \end{aligned}$$

The logarithm of the second term of the general solution produces sequence of relations

$$\begin{aligned} \lim_{t \rightarrow \infty} \ln \left[E_0 \prod_{j=0}^t (1 + b_j) \right] &= \lim_{t \rightarrow \infty} \left\{ \ln E_0 + \sum_{j=0}^t \ln(1 + b_j) \right\} \\ &\leq \ln E_0 + \lim_{t \rightarrow \infty} \sum_{j=0}^t b_j \\ &= -\infty \end{aligned}$$

This implies that

$$\lim_{t \rightarrow \infty} E_0 \prod_{j=0}^t (1 + b_j) = 0$$

Considerably more mathematical rigor is needed to show that the first term of the general solution also converges to zero. A sufficient proof is one which shows that for every $\epsilon > 0$ there exist a T such that

$$\left| \sum_{i=0}^t a_i \prod_{j=i+1}^t (1 + b_j) \right| < \epsilon$$

for every t greater than T .

The proof needs several steps. First, since $0 \leq 1 + b_j \leq 1$,

$\prod_{j=i+1}^t (1 + b_j) \leq 1$ and since the $a_i \geq 0$ there exists some I such that

$$\sum_{i=I+1}^t a_i < \frac{\epsilon}{2}$$

for every t .

These inequalities imply that

$$\left| \sum_{i=I+1}^t a_i \prod_{j=i+1}^t (1 + b_j) \right| \leq \sum_{i=I+1}^t a_i < \frac{\epsilon}{2}$$

The second part of the proof results from knowing that the product converges to zero. That is, there exists some T such that for every $t > T$

$$\left| \prod_{j=I+1}^t (1 + b_j) \right| < \frac{\epsilon}{2A}$$

This inequality produces the relation that

$$\begin{aligned} \left| \sum_{i=0}^I a_i \prod_{j=i+1}^t (1 + b_j) \right| &\leq \left| \sum_{i=0}^I a_i \prod_{j=i+1}^t (1 + b_j) \right| \\ &< \frac{\epsilon}{2A} \sum_{i=0}^I a_i \\ &\leq \frac{\epsilon}{2} \end{aligned}$$

Thus the total sum from 0 to I and $I + 1$ to t is less than ϵ whenever $t > T$. This implies that the second term converges to zero. The rate at which it converges to zero depends on the particular coefficients of the difference equation.

PROPERTIES OF THE ITERATION

An example of the use of the first order difference equation is in the calculation of a norm equal to the inner product of the error between the system parameters X and the optimum parameters θ

$$E_n = \text{Average} \left[(X_n - \theta) \cdot (X_n - \theta) \right]$$

The iterative procedure for the determination of the process is an iterative correction

$$X_{n+1} = X_n + \epsilon_n$$

The error is then

$$X_{n+1} - \theta = X_n - \theta + \epsilon_n$$

so that the norm is

$$E_{n+1} = E_n + \text{Average} \left\{ \epsilon_n \cdot \epsilon_n + 2(X_n - \theta) \cdot \epsilon_n \right\}$$

This is then a difference equation of the norm.

$$\Delta E_n = \text{Average} \left\{ \epsilon_n \cdot \epsilon_n + 2(X_n - \theta) \cdot \epsilon_n \right\}$$

In those cases where there exists appropriate a_n and b_n such that

$$\text{Average} \left\{ \epsilon_n \cdot \epsilon_n + 2(X_n - \theta) \cdot \epsilon_n \right\} \leq a_n + b_n E_n$$

The behavior of the norm E_n can be analyzed quite effectively with the solution of the difference equation.

References

1. N. Abramson, D. Braverman, "Learning to Recognize Patterns in a Random Environment," IRE Transactions on Information Theory, Vol. IT-8, No. 5, Sept., 1962, S 58-63, International Symposium on Information Theory.
2. R. C. Amara, "The Linear Least Squares Synthesis of Continuous and Sampled Data Multivariable Systems," Stanford Electronics Labs., TR40, July 28, 1958.
3. Armar Bose, "A Theory of Nonlinear Systems," Research Laboratory of Electronics, MIT, TR 309, May 15, 1956.
4. D. J. Braverman, "Machine Learning and Automatic Pattern Recognition," Stanford Electronics Labs., TR 2003-1, February 17, 1961.
5. Murray Eden, "Handwriting and Pattern Recognition," IRE Transactions on Information Theory, Vol. IT-8, No. 2, February 1962, pp 160-166.
6. D. Gabor, et al., "A Universal Nonlinear Filter, Predictor and Simulator which Optimizes itself by a Learning Process," Proc. IEE, Vol. 108, pt B, pp 422-438, July 1961.
7. C. W. Harrison, "Experiments with Linear Prediction in Television," BSTJ., Vol. 31, pp 764-783, July, 1952.
8. H. Highleyman, "Linear Decision Functions, with Application to Pattern Recognition," Proc. IRE, Vol. 50, No. 6, Part 1, June 1962, pp 1501-1514.
9. Ming-Kuei Hu, "Visual Pattern Recognition by Moment Invariants," IRE Transactions on Information Theory, Vol. IT-8, No. 2, February 1962, pp 179-187.
10. D. A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes" Proc. IRE, Sept. 1952, pp 1098-1101.
11. T. Kailath, "Correlation Detection of Signals Perturbed by a Random Channel," Trans. IRE-PGIT-6, pp 361-366, June 1960.
12. R. E. Kalman, "New Methods and Results in Linear Prediction and Filtering Theory," RIAS, TR 61-1, 1961.
13. E. R. Kretzmer, "Statistics of Television Signals," BSTJ., Vol. 31, pp 751-763, July 1952.
14. E. R. Kretzmer, "Reduced Alphabet Representation of Television Signals," 1956 IRE National Convention Record, pt 4, pp 140-147.

References (Continued)

15. P. M. Lewis II, "The Characteristic Selection Problem in Recognition Systems," IRE Transactions on Information Theory, Vol. IT-8, No. 2, February 1962, pp 171-178.
16. C. S. Lorens, "Optimum Filtering," JPL, EP 633, May 15, 1959.
17. C. S. Lorens, "The Doppler Method of Satellite Tracking," 3rd Natl. Con. Mil. Elec., 1959 Proceedings, pp 255-263.
18. C. S. Lorens, "Aspects of Theoretical and Experimental Map-Matching," Space-General Corporation, RM 17, December 15, 1961.
19. C. S. Lorens, A. M. Boehmer, J. Gallagher, "Theoretical and Experimental Processing of Lunar Television Pictures," Bi-Monthly Progress Report No. 1, 2, 3, SGC 203R-1, 2, 3, May, July, Sept. 23, 1962.
20. J. K. Lubbock, "A Self-Optimizing Nonlinear Filter," Proc. IEEE, Vol. 108, pt B, pp 439-440, July 1961.
21. Daniel D. McCracken, A Guide to FORTRAN Programming, John Wiley and Sons, Inc., New York, 1961.
22. W. W. Peterson, Error Correcting Codes, MIT Press and Wiley, 1961.
23. D. A. Sakrison, "Application of Stochastic Approximation Methods to System Optimization," MIT+ RLE TR 391, July 10, 1962, and ScD Thesis Dept. EE, MIT, June 1961.
24. W. F. Schreiber, "The Measurement of Third-Order Probability Distributions of Television Signals," IRE Transactions on Information Theory, Vol. IT-2, pp 94-105, Sept., 1956.
25. George S. Sebestyen, "Recognition of Membership in Classes," Trans. IRE-PGIT-7, pp 44-50, January 1961.
26. Norbert Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series, MIT Technology Press and John Wiley, 1949.
27. Norbert Wiener, Nonlinear Problems in Random Theory, MIT Technology Press and John Wiley, 1958.
28. D. C. Youla, "On the Factorization of Rational Matrices," Polytechnic Institute of Brooklyn, Microwave Research Institute, RR PIBMRI 855-60, Sept. 20, 1960, and Trans. IRE-PGIT-7, pp 172-189, July 1961.