INTRODUCTION TO
COMPUTER IMAGE PROCESSING

JOHANNES G. MOIK

PRICES SUBJECT TO CHANGE

OCTOBER 1973

GSFC —— GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND

For information concerning availability
of this document contact:

Technical Information Division, Code 250
Goddard Space Flight Center
Greenbelt, Maryland 20771

(Telephone 301-982-4488)

INTRODUCTION TO COMPUTER IMAGE PROCESSING

Johannes G. Moik
Computer Sciences Corporation*

October 1973

GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland

INTRODUCTION TO COMPUTER IMAGE PROCESSING

Johannes G. Moik
Computer Sciences Corporation

ABSTRACT

This report presents theoretical backgrounds and digital techniques for a class of image processing problems. It is intended to provide knowledge for the use of the VICAR and SMIP systems in operation at the Laboratory for Meteorology and Earth Sciences.

Generally two kinds of image processing can be distinguished. One kind transforms images into new images being different from the original in some desirable way. Image restoration and image enhancement belong to this group and the objective is to compensate for distortions introduced by the image forming process. Since image formation can be described or approximated by linear processes, two-dimensional linear system theory and linear transformations provide the theoretical framework for image restoration and image enhancement.

The result of the other kind of image processing is not an image but may take the form of a description or parametrization. Curve detection, object extraction, area determination and classification are problems in this class of image processing where linear transformations, statistics, graph theory and heuristics provide useful mathematical methods.

This report is mainly devoted to the first kind of image processing. Image formation in the context of linear system theory, image evaluation, noise characteristics, mathematical operations on images and their implementation are discussed in chapter 2. Various techniques for image restoration and image enhancement are presented in chapters 3 and 4 respectively. Chapter 5 describes methods for object extraction and the problem of pictorial pattern recognition and classification is briefly discussed in chapter 6.

CONTENTS

CONTENTS (Continued)

ILLUSTRATIONS

INTRODUCTION TO COMPUTER IMAGE PROCESSING

1. INTRODUCTION

Image processing in a broad sense deals with the manipulation of data which are inherently two-dimensional in nature. The purpose of image processing is to aid man to extract information from images. This implies that man is usually not capable of extracting all of the information in an image. The image may be blurred by some defect such as defocus or image motion. In addition, images always contain noise due to sensor characteristics, and, in some cases the principal degradation may be noise alone.

The human visual system is a remarkable optical processing system with some very unique capabilities. It has, however, also limitations. The need for image processing is based upon the assumption that the human visual system is not necessarily efficient at performing all tasks related to the extraction of information from an image. The validity of this assumption has been supported by experiments which have shown that for a range of image defects and noise levels the human visual system is not an efficient information extraction system [1].

Image processing must always use some kind of a priori knowledge. Without a priori knowledge about the characteristics of object space and the imaging system there would be no basis for judging whether a picture is a good representation of an object and therefore there is a reason to process the image. Thus, some form of a priori knowledge must be applied to a degraded image in order to extract information from it. One kind of a priori knowledge is concerned with intelligence information. Processing done on a degraded image may be different if the image is known to come from a certain source rather than being unknown. For example, different processing methods may be applied to high resolution images of the earth versus images of distant astronomical objects or X-ray pictures of the human body.

Another kind of a priori knowledge of great importance to image processing is concerned with the physical process of the formation of an image. This includes knowledge of object characteristics, transmission medium characteristics, properties of sensor and systems which recorded the image and possibly the characteristics of the scanner used to scan a photograph. For example, the correction of photometric and geometric distortions which occur in imaging with a vidicon camera requires knowledge of the characteristics of the vidicon tube. All this information is used to reduce the number of variables involved in processing. In fact, one kind of image processing which is often referred to as a priori processing deals with the clever design of imaging systems to minimize

degradations. This discussion, however, will concentrate on processing of given images without having control over how the images are formed.

The description and processing of images can be facilitated when the image forming processes have some kind of mathematical structure upon which a characterization can be based. This will help in the development of a theory of image processing since portions of this field as yet remain more an art rather than a science. As with any emerging scientific field, heuristic techniques play a major role in the solution of many problems. It is not the purpose of this paper to present heuristic approaches which would often provide encouraging results for a given problem, but to present theoretical backgrounds which could possibly provide a common basis of approach to a class of image processing problems.

For example, the basis for most of the design technology in the field of signal processing is the theory of linear systems. That linear system theory is as advanced as it is, stems from the fact that the defining properties of linear systems guarantee that they can be analyzed. The analysis, based on the principle of superposition leads directly to the concepts of scanning, sampling, filtering, convolution, stochastic estimation, etc. Equally important is the idea that the mathematical structure of the information being processed be compatible with the structure of the physical processes to which it is exposed.

In image processing where electrical technology and optics are a dominating influence, linear models have been traditionally used. This is natural since image processing is based on those branches of classical physics which employ linear mathematics as their foundation namely electric measurements, electronics, signal theory and communication theory. In optical image processing, the laws of image formation and degradation are derived from linear diffraction theory. In the section on image formation it will be shown that the structure of linear systems is compatible with the structure of a large class of images themselves.

For the purpose of this discussion, a system is mathematically defined to be a mapping of a set of input functions into a set of output functions. For imaging systems, the inputs and outputs can be real-valued functions or complex-valued functions of a two-dimensional independent spatial variable (x, y). The systems considered are characterized by many-one mappings. A convenient representation of a system is a mathematical operator, $\delta\{\ \}$, which operates on input functions to produce output functions. Thus, if the function $f(x_1, y_1)$ represents the input to a system and $g(x_2, y_2)$ represents the corresponding output, then by the definition of $\delta\{\ \}$, the two functions are related through

$$g(x_2, y_2) = \delta\{f(x_1, y_1)\} \tag{1}$$

2

The discussion of this paper will be restricted to the class of linear systems. A system is said to be linear if the following superposition property is obeyed for all input functions s and t and all complex constants a and b:

$$\delta\left\{a\,s(x_1,y_1)+b\,t(x_1,y_1)\right\} = a\delta\left\{s(x_1,y_1)\right\} + b\delta\left\{t(x_1,y_1)\right\} \qquad (2)$$

The great advantage afforded by linearity is the ability to express the response of the system to an arbitrary input in terms of the responses to certain elementary functions into which the input has been decomposed. Each of these produces a known response and by virtue of linearity the total response can be found as a corresponding linear combination of the responses to the elementary stimuli.

A simple and convenient decomposition of the input is offered by a property of the δ function which states that

$$f(x,y) = \int_{-\infty}^{\infty}\!\!\int f(\xi,\eta)\,\delta(x-\xi,y-\eta)\,d\xi\,d\eta \qquad (3)$$

This equation may be regarded as expressing f as a linear combination of weighted and displaced δ functions. The elementary functions of the decomposition are the δ functions; they may physically represent an idealized point source of light. A possible definition of the two-dimensional δ function is

$$\delta(x,y) = \lim_{N\to\infty} N^2\,e^{-N^2(x^2+y^2)} \qquad (4)$$

To find the response of a system to the input f, substitute (3) in (1):

$$g(x_2,y_2) = \delta\left\{\int_{-\infty}^{\infty}\!\!\int f(\xi,\eta)\,\delta(x_1-\xi,y_1-\eta)\,d\xi\,d\eta\right\} \qquad (5)$$

Regarding the number $f(\xi,\eta)$ as simply a weighting factor applied to the elementary function $\delta(x_1-\xi,y_1-\eta)$, the property (2) is invoked for linear systems

3

to allow $\delta\{\ \}$ to operate on the individual elementary functions. This yields

$$g(x_2, y_2) = \int_{-\infty}^{\infty}\!\!\int f(\xi, \eta)\, \delta\left\{\delta(x_1 - \xi, y_1 - \eta)\right\} d\xi\, d\eta \qquad (6)$$

Now let the symbol $h(x_2, y_2; \xi, \eta)$ denote the response of the system at point (x_2, y_2) of the output space to a δ function input at coordinates (ξ, η) of the input space. The function h is called the impulse response of the system. The system input and output can now be related by the simple equation

$$g(x, y) = \int_{-\infty}^{\infty}\!\!\int f(\xi, \eta)\, h(x, y; \xi, \eta)\, d\xi\, d\eta \qquad (8)$$

This fundamental expression demonstrates the important fact that a linear system is completely characterized by its response to unit impulses. For the case of a linear imaging system, this result has the interesting physical interpretation that the effects of imaging elements can be fully described by the images of point sources located throughout the object field [2].

The techniques of image processing find applications in many areas, notably: image restoration, image enhancement, object extraction, pictorial pattern recognition and the efficient encoding of pictures for transmission and storage. This paper will only discuss the various methods for image restoration, image enhancement and object extraction. The common questions underlying these areas are:

1) How can images be represented and described?

2) How are images formed?

3) What mathematical operations can be used on images?

4) How can these mathematical operations be implemented?

5) How can image quality be evaluated?

6) What is the effect of noise on images?

In the next chapter these questions will be elaborated upon.

4

2. IMAGE PROCESSING CONCEPTS

2.1 IMAGE REPRESENTATION AND DESCRIPTION

Representation is an important question in the transmission, storage and processing of any information. A physical image, as a carrier of information, is represented in the form of light energy. This natural representation can be used by creating a signal proportional to the intensity of that energy. This is common in television and digital image processing whereas the photographical process does not use the representation by light intensity.

Mathematically an image can be described by a function of two real spatial variables. The value of such a picture function $f(x,y)$ at a given point is called grey value. In a black and white picture this value is determined by one parameter, the light intensity at the given point. Color pictures are described by several parameters. Any physical picture is of finite extent, hence, one can suppose that any picture function is zero outside a rectangle R of prespecified size. Since the amount of light reaching the observer from a physical picture is finite and nonnegative one can assume that there is an upper bound M to the possible brightness of any physical picture. Thus, any picture function is bounded and nonnegative and will be described by

$$0 \leq f(x,y) \leq M \qquad \text{for all} \qquad (x,y) \in R$$

$$(9)$$

$$R = \left\{ (x,y) \mid 0 \leq x \leq x_M , \quad 0 \leq y \leq y_N \right\}$$

The raster-scan operation of scanning instruments and image digitizers imposes a sequential row structure on images. The orientation of the coordinate system used in this paper is shown in Figure 1 where the x-axis is in the direction of increasing line numbers.

For color pictures the definition of a picture function must be extended to include the color information. This can be done by taking advantage of the fact that the space of all possible colors can be spanned by a three-dimensional vector space. This fact, which rests on physiological rather than physical evidence, allows to define a picture function as a vector-valued function $f(x,y) = \left(f_1(x,y), f_2(x,y), f_3(x,y) \right)$ where f_1, f_2, f_3 give the coordinates of the color of the picture at the point (x,y). Since the parametrization of the color space is not unique, the basis vectors must be specified to define a picture function completely. One basis is the set of the three primary colors, red, green and

5
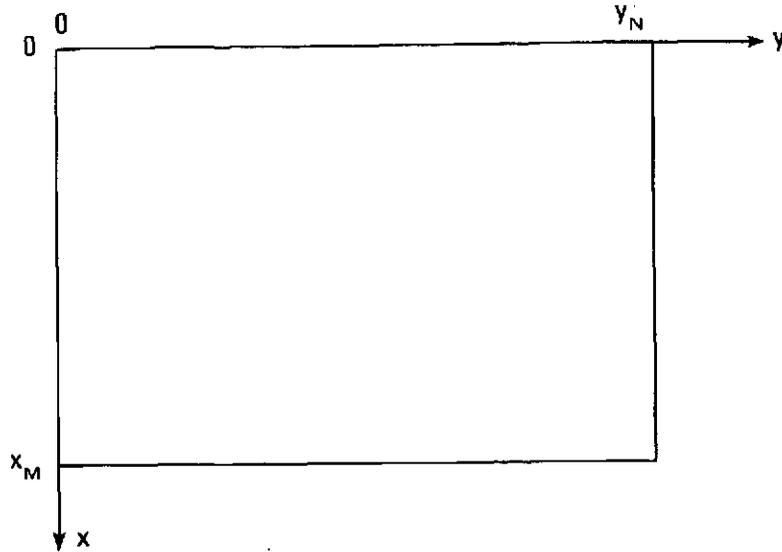
Figure 1. Image Domain

blue. Physically, using this basis corresponds to taking three scalar pictures of a single scene using, successively, red, green and blue filters. Consequently, the image is represented by three intensity matrices $f_R(x, y)$, $f_G(x, y)$ and $f_B(x, y)$, where the subscripts denote red, green and blue primary components.

Normalized color coordinates of a picture point can be defined by

$$f_r = \frac{f_R}{f_R + f_G + f_B}$$

$$f_g = \frac{f_G}{f_R + f_G + f_B}$$

$$f_b = \frac{f_B}{f_R + f_G + f_B}$$

These numbers are independent of x and y and are called chromaticity coordinates or trichromatic coefficients. They sum to unity and two of them can be taken to define the normalized color coordinates at a point in the picture.

6

Figure 2 shows a representation of the 2-dimensional normalized color space for f_r and f_g. Normalizing colors amounts to changing the basis of the 3-dimensional color space from the original red, green and blue coordinates to normalized red, normalized green and total intensity.
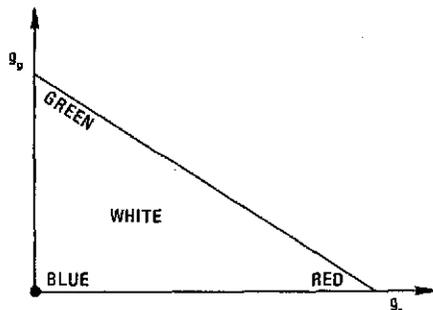
Figure 2. Normalized Color Space

The amount of information that a color picture can convey to the observer is much greater than from a black and white picture. Experiments have indicated that although the human eye can differentiate only two or three dozen brightness levels it is able to separate thousands of various colors.

2.2 IMAGE FORMATION

Images are formed of light reflected from objects. The output of an image forming system is always degraded to some extent. Blurring is caused by the diffraction of light through a finite aperture, by aberrations of the system and by motions of object or imaging systems. Therefore, a practical image will only approximate the original image given in (9). The information content of the recorded image is limited by the resolution of the imaging system and by the presence of noise. The degradation of the original image is thus determined by the transfer characteristics of the imaging system. Because most image forming methods involve linear mechanisms, a practical image can be regarded as an additive superposition of points in the original image. This fact is expressed in (8) where $g(x, y)$ represents the image intensity recorded by the imaging system, $f(\xi, \eta)$ is the original object intensity function and $h(x, y; \xi, \eta)$ is the response in the image coordinates (x, y) to a unit impulse at (ξ, η) in the object coordinates. The impulse response $h(x, y; \xi, \eta)$ is also called point-spread function, obviously it must be nonnegative. The relation (8) holds when the illumination is incoherent and the input and output functions represent intensity distributions [2].

7

Generally, the response $h(x, y; \xi, \eta)$ in the image space varies with the position (ξ, η) of the input impulse and is called space-variant or shift-variant point-spread function. If h is a function only of the differences in the coordinates, then $h(x - \xi, y - \eta)$ characterizes a shift-invariant or space-invariant system and the superposition integral (8) becomes a convolution integral (10).

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi, \eta) \, h(x - \xi, y - \eta) \, d\xi \, d\eta \qquad (10)$$

In any practical imaging system, the observed image $g(x, y)$ will always be corrupted by noise of some kind. For example, if $g(x, y)$ has been recorded on film, there is grain noise or when $g(x, y)$ is sampled for digital processing, there is quantization noise. Thus, beside the noise free problem (8) there is the corresponding problem (11) when noise is added to the system output.

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi, \eta) \, h(x - \xi, y - \eta) \, d\xi \, d\eta + n(x, y) \qquad (11)$$

The assumption that noise is additive is subject to criticism. Some noise sources like granularity in photographic recording can be more accurately modeled as multiplicative [3]. However, the additive assumption is common to almost all work in image processing because it makes the problem mathematically tractable. Some cases in which multiplicative noise can be handled are discussed in [4].

Degradations due to motion blurring and geometrical distortion may also limit the resolution in the recorded image. Examples are images taken from aircraft and spacecraft. Image formation in this case can also be modeled by the linear space-variant system (8). These systems can be analyzed by decomposition into geometric coordinate distortions and a space-invariant operation [5, 6].

For practical systems the point spread function $h(x, y)$ decreases to zero or to negligible values for large $|x|$ and $|y|$. Therefore, it can be assumed that

$h(x, y) = 0$ for $|x| > x_L$, $|y| > y_L$ and (10) can be written

$$g(x, y) = \int_{x-x_L}^{x+x_L} \int_{y-y_L}^{y+y_L} f(\xi, \eta)\, h(x - \xi,\, y - \eta)\, d\xi\, d\eta \qquad (10a)$$

Considering the finite extent of the picture function $f(x, y)$, the convolution can be computed for $x_L \leq x \leq x_M - x_L$, $y_L \leq y \leq y_N - y_L$.

2.3 MATHEMATICAL OPERATIONS ON IMAGES

Image processing involves the transformation of an image from one form into another. Two distinct kinds of processing are generally possible. One uses a form of transformation which results in a new image being different from the original in some desirable way. The result of the other kind is not an image but may take the form of a decision, description or a parameterization.

The first kind of image processing can be thought of as image coding, image restoration and image enhancement. It involves the transformation of an image by restoration operators and enhancement operators. Restoration operators compensate for image degradation and permit the reconstruction of the original image. Degradation and restoration are thus inverse operations. Since restoration attempts to compensate for distortions caused in part by linear mechanisms, linear processing methods are used extensively. Enhancement operators change the image further rather than restore it in order to emphasize certain features for the human observer. This selective accentuation and suppression reduces the information content of an image and it is important to decide what to retain. The mathematical theory of integral transformations is a suitable framework for describing many aspects of restoration, enhancement and object detection.

The second kind of image processing deals with structural operators or feature extractors and syntactic schemata. It is a mapping from pictures to descriptions of pictures. In examining a picture one is often interested only in extracting from it a description of what it depicts. This is the problem of pictorial pattern recognition and scene analysis. The desired description may be merely a classification of the picture into one of a small set of prespecified classes. The description may also involve properties of, and relationships among, objects that appear in the picture. To obtain such a description, it is usually necessary to explicitly locate the objects in the picture and to measure their properties and interrelationships. Picture descriptions in terms of objects, properties and

relationships can be expressed using special picture languages [7]. Mathematical methods useful for this class of image processing include statistics, graph theory, theory of formal languages and automaton theory.

The integral transform $F(u, v)$ of a function $f(x, y)$ is defined by

$$F(u, v) = \int_a^b \int_c^d f(x, y) \, K(u, v, x, y) \, dx \, dy \tag{12}$$

where $K(u, v, x, y)$ is called the kernel of the transformation. The limits of integration may be finite or infinite. If the kernel K and the transform F are known, then the process of determining a solution f of the integral equation is called inversion. The properties of the original function f and the kernel K over the regions of integration determine the existence of the integral transform F and the existence of an inversion formulae giving f as an integral transform of F:

$$f(x, y) = \int_\alpha^\beta \int_\gamma^\delta F(u, v) \, K^{-1}(u, v, x, y) \, du \, dv \tag{13}$$

Mathematical conditions can be found in [16]. If this formalism is used to describe two-dimensional image formation, then f represents the original image, the kernel represents the action of the imaging system and the transform gives the resultant image. If the image forming process can be modeled by a linear, space-invariant system, the transformation can be described by a linear homogeneous operator. In this case a large body of mathematical theory becomes available to image processing.

Since any homogeneous linear operator is equivalent to a convolution integral, those linear integral transforms for which there are convolution and inversion theorems are most useful in image processing. Convolution and inversion theorems are known for several kernels, such as the Fourier, Hankel, Laplace, Mellin, Hadamard and Haar-Walsh transforms. Another useful property of a kernel is separability. A kernel is said to be separable if it can be written as a product of two functions

$$K(u, v, x, y) = K_1(x, u) \cdot K_2(y, v) \tag{14}$$

10

Separability allows two-dimensional transforms to be reduced to a sequence of one dimensional transforms which facilitates the computation considerably. The transformation kernel is called separabel symmetric if

$$K(u, v, x, y) = K_1(x, u) \cdot K_1(y, v) \tag{15}$$

Since the statistical intensity variations of most images are nearly the same in the vertical and horizontal directions only separable symmetric kernels need to be considered [17].

The image forming model (10) can be mathematically described by an operator S. If $\mathcal{T}_{\xi,\eta}$ denotes a translation operator

$$\left[\mathcal{T}_{\xi,\eta}(f)\right](x, y) = f(x - \xi, y - \eta) \tag{16}$$

then S satisfies the following criteria:

a) S is linear, i.e. it satisfies the superposition property (2)

b) S is homogeneous or position invariant, i.e. it commutes with every translation operator $\mathcal{T}_{\xi,\eta}$

$$S\left(\mathcal{T}_{\xi,\eta}(f)\right) = \mathcal{T}_{\xi,\eta}\left(S(f)\right) \tag{17}$$

Equation (10) can also be written as

$$g(x, y) = f(x, y) * h(x, y) \tag{18}$$

where * denotes the two-dimensional convolution operation.

Two useful operations in image processing which can be derived from convolution are crosscorrelation and autocorrelation. They can be used to measure the goodness of match of two patterns. Let $\bar{h}$ denote the result of reflecting h in the origin, $\bar{h}(x, y) = h(-x, -y)$ for all x, y. The convolution $f * \bar{h}$ is called the

crosscorrelation of f and h and is denoted by f × h. Thus,

$$f \times h = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\xi, \eta) \, f(\xi + x, \, \eta + y) \, d\xi \, d\eta \tag{19}$$

The crosscorrelation of f with itself, f × f, is called the autocorrelation of f [7]. Symmetry and antisymmetry properties of convolution and correlations are useful in image analysis.

Among the various possible transforms the Fourier transform has been the most useful in enhancement and object detection for several reasons. It can be readily interpreted in terms of spatial frequency, has a separabel kernel and can be effeciently implemented by digital as well as electro-optical and optical means [8,9].

2.3.1 Fourier Analysis in Two Dimensions

This section will summarize definition, some existence conditions and properties of the two-dimensional Fourier-transform. No attempt at great mathematical rigor is made, but rather an operational approach will be adopted.

The Fourier-transform of a function g of two independent variables will be represented by $\mathcal{F}\{g\}$ and is defined by

$$\mathcal{F}\{g\} = G(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \, e^{-2\pi i (ux+vy)} \, dx \, dy \tag{20}$$

The Fourier-transform is a complex-valued function of two independent variables, u and v, which are referred to as spatial frequencies. Similarly, the inverse Fourier-transform of a function G (u, v) will be represented by $\mathcal{F}^{-1}\{G\}$ and is defined as

$$\mathcal{F}^{-1}\{G\} = g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(u, v) \, e^{2\pi i (ux+vy)} \, du \, dv \tag{21}$$

12

Before discussing the properties of the Fourier-transform and its inverse, the question of the existence of the transform pair (20) and (21) shall be briefly elaborated upon. For certain functions, these integrals may not exist in the usual mathematical sense. The following set of sufficient conditions is commonly quoted [10]:

1. g must be absolutely integrable over the entire picture plane.

2. Any discontinuities in g are finite.

3. g must have only a finite number of discontinuities and a finite number of maxima and minima in any finite rectangle.

Although some functions (e.g., periodic functions) do not have a Fourier-transform, as may be verified by reference to the conditions above, it is possible to find a meaningful transform, provided the functions can be defined as the limit of a sequence of functions that are transformable. By transforming each member function of the defining sequence, a corresponding sequence of transforms is generated and the limit of this sequence is called the generalized Fourier-transform of the original function [10]. To illustrate the calculation of a generalized transform, consider the Dirac δ function which violates existence condition 2. Each member function of the defining sequence (4) satisfies the existence requirements and has a Fourier-transform given by

$$\mathcal{F}\left\{N^2 \; e^{-N^2(x^2+y^2)}\right\} \;=\; e^{-(u^2+v^2)/N^2}$$

Accordingly the generalized Fourier-transform of $\delta(x, y)$ is

$$\mathcal{F}\left\{\delta(x, y)\right\} \;=\; \lim_{N \to \infty} e^{-(u^2+v^2)/N^2} = \; 1 \tag{22}$$

The question of the existence of a Fourier-transform will be of no further concern, for practical difficulties are rarely if ever caused by problems involving the existence of the defining integrals.

Some properties of the transform that will be useful are [2, 10]

Linearity:

$$\mathcal{F}\left\{a f_1(x, y) + b f_2(x, y)\right\} \;=\; a F_1(u, v) + b F_2(u, v) \tag{23}$$

13

Scaling:

$$\mathcal{F}\left\{f(ax, by)\right\} = \frac{1}{|ab|} F\left(\frac{u}{a}, \frac{v}{b}\right) \tag{24}$$

Shifting:

$$\mathcal{F}\left\{f(x-a, y-b)\right\} = F(u,v) \, e^{-i(ua+vb)} \tag{25}$$

That is, the Fourier-transform is not a shift-invariant operation. Translation of a function in the space domain introduces a linear phase shift in the frequency domain.

Convolution:

$$\mathcal{F}\left\{f(x,y) * h(x,y)\right\} = F(u,v) \, H(u,v) \tag{26}$$

The convolution of two functions in the space domain is equal to the inverse Fourier-transform of the product of their Fourier-transforms.

Autocorrelation:

$$\mathcal{F}\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi, \eta) \, f^*(\xi-x, \eta-y) \, d\xi \, d\eta\right\} = |F(u,v)|^2 \tag{27}$$

The Fourier-transform of the autocorrelation function is the power spectrum of a signal.

Parseval's Theorem:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) \, g^*(x,y) \, dx \, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v) \, G^*(u,v) \, du \, dv \tag{28}$$

14

This theorem is generally interpretable as a statement of conservation of energy.

Gradients:

$$\mathcal{F}\left\{\nabla^2 f(x, y)\right\} = -(u^2 + v^2) F(u, v) \tag{29}$$

Inversion:

$$\mathcal{F}^{-1}\left\{\mathcal{F}\left\{f(x, y)\right\}\right\} = f(x, y) \tag{30}$$

Conjugate symmetry:

If $f(x, y)$ is real, then

$$\mathcal{F}\left\{f(x, y)\right\} = F(u, v) = F^*(-u, -v) \tag{31}$$

The two-dimensional Fourier-transform is just one way of decomposing a function $g(x, y)$ into a set of coefficients of orthogonal functions. Any orthogonal transformation will provide a similar decomposition. The advantage of the Fourier-transform is that the orthogonal functions are trigonometric functions which are the eigenvector solutions of linear space-invariant systems.

In the following, the intuitive consequences of the definition of the Fourier-transform pair shall be considered. The transform of a picture function (21) defines the weighting coefficients of its expansion in a sum of complex exponentials. To understand the transform, a feeling for the pictorial appearance of the exponential

$$e^{2\pi i (ux+vy)}$$

for a given value of u and v is needed. The locus of points in the x - y plane for which this complex function is real (and has thus zero phase) is given by

$$y = -\frac{u}{v} x + \frac{n}{v} \qquad \text{n integer}$$

15

which is a set of parallel lines whose slope and spacing is determined by u and v. There are, thus, two geometrical aspects associated with a given point (u, v) in the spatial frequency plane: an orientation ⊕ and a spacing L.

It is interesting to visualize the implications of the Fourier-transform F (u, v) of a picture f (x, y) being nonzero only at the spatial frequencies (u, v). Therefore, and by consequence of (31) only two terms will contribute to the decomposition (21). The picture will look like

$$F(u, v) \ e^{2\pi i(ux+vy)} + F(-u, -v) \ e^{2\pi i(-ux-vy)}$$

which is real and is like a sinusoidally undulating surface whose crests are a set of parallel lines. Hence, each symmetric pair (u, v) and (-u, -v) of spatial frequencies contributes to the generalized sum (21) one picture consisting of parallel stripes of sinusoidally varying intensity. The greater the magnitude of the transform is at (u, v), the more important is this contribution.

Suppose now, a picture function f (x, y) consisting of vertically oriented dark and light stripes with abruptly changing transitions. The Fourier-transform then will be nonzero only in the u axis, since the stripes are vertical in the picture.

In general, edges in a picture introduce spatial frequencies along a line in the complex frequency plane orthogonal to the edge. Intuitively, high spatial frequencies correspond to sharp edges, low spatial frequencies to regions of approximately uniform grey level. The orientation of a spatial frequency corresponds to the orientation of an edge in the picture.

2.3.2 Image Transformations

The transformations mentioned above belong to the class of unitary transformations. Until recently few systems took advantage of the power of the theory of unitary transformations. With the development of fast algorithms which vastly decrease the computational requirements, great interest has been generated in digital spectral decomposition [8, 9]. The theory of unitary transformations can provide powerful processing methods. One of the most appealing justifications for the use of unitary transforms lies in the properties of its eigenvalues and eigenvectors which define optimal solutions to various systems in the theory of conventional as well as stochastic systems.

In fact, consider the linear shift-invariant system (10) with impulse response h and denoted by the operator $\mathcal{L}$. If the input function is of the form

$f(x, y) = e^{i(ux+vy)}$, the corresponding response is given by

$$g(x, y) = e^{i(ux+vy)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\xi, \eta) \, e^{-i(u\xi+v\eta)} \, d\xi \, d\eta$$

The integral is called the system transfer function

$$H(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\xi, \eta) \, e^{-i(u\xi+v\eta)} \, d\xi \, d\eta \tag{32}$$

$H(u, v)$ is the Fourier-transform of $h(\xi, \eta)$ and

$$\mathcal{L}\left\{ e^{i(ux+vy)} \right\} = H(u, v) \, e^{i(ux+vy)} \tag{33}$$

This shows that $e^{i(ux+vy)}$ is an eigenfunction of $\mathcal{L}$ for any u and v and $H(u, v)$ are the eigenvalues of $\mathcal{L}$.

In the area of stochastic processes the decomposition of a nonperiodic random process into a series of orthogonal functions is given by the Karhunen-Loéve theorem [18]. Each element of a visual image can be considered as a random variable which is proportional to the intensity of the scene at the element location. The entire pattern can then be considered as a random process and the pattern statistics can be described by the covariance matrix C determined over all patterns. If C is known, the Karhunen-Loéve transformation provides maximum compression of the image information resulting in the minimum set of features necessary to obtain class separation.

The transformation is found by determining the eigenvectors of C defined by

$$C\Phi = \lambda\Phi$$

The set of eigenvectors Φ determines a new pattern space where each element is uncorrelated. The picture information is generally distributed over many

17

elements in the original space and only over a few in the transformed space if the correlation between elements is high.

While the eigenvector solution to specific systems provides optimum orthogonal decomposition, calculation of eigenvalues and eigenvectors is a very difficult if not numerically impossible task. Consequently suboptimum decompositions are desirable if they can be efficiently generated by digital processes. In section 2.4 the discrete Fourier-transform as the most important example for the implementation of a fast decomposition algorithm will be discussed.

2.4 IMPLEMENTATION OF IMAGE PROCESSING OPERATIONS

There are basically two ways to implement image processing operations, either by digital or by optical means. The first way is referred to as digital image processing, the second as optical or electron-optical image processing [46,47]. The question which arises is what types of operations can better be done on a digital computer than by optics and vice versa. In this section, digital computer and coherent image processing techniques are briefly reviewed and compared to indicate their relative merits and faults.

The main uses of coherent optical systems in image processing have been Fourier transformation and linear filtering. These operations are possible because of the Fourier transforming property of a lens [2]. If a film transparency with amplitude transmission $g(x,y)$ is placed in the front focal plane of a lens and is illuminated with collimated monochromatic light, then the amplitude of the light in the back focal plane of the lens will be the Fourier transform of $g(x,y)$. Linear filtering can be performed by putting a transparency with amplitude transmission $H(u,v)$ in the back focal plane of the lens and using a second lens whose front focal plane coincides with the back focal plane of the first lens. Then at the back focal plane of this second lens, the light amplitude will be the convolution of the original image with the inverse Fourier transform of the filter amplitude transmission.

Digital image processing uses a digital computer to implement the various operations on images. This implies that the image has to be in digital form. Digital images are obtained either directly as from some imaging systems aboard spacecrafts or they are digitized by a film scanner. This is the usual case for biomedical images. The digital computer is used because of its great flexibility in the implementation of image processing operations, and its accuracy. However, conventional digital computers are sequential, they can perform only one (at most a few) arithmetical operation(s) at a time. Thus, image processing with digital computers is usually very time consuming. The convolution of a picture of size N with a mask of size M, for example, takes $N^2 \cdot M^2$ multiplications

which easily becomes prohibitive. The development of fast algorithms, e.g. the fast Fourier-transform, improved the situation considerably. Great savings could be realized in the processing time required if it were possible to perform the identical operations for each picture point in parallel. For this purpose one could use a parallel computer having many processing units that operate under central control like the ILLIAC IV [11,12]. A problem is to adopt the algorithms for parallel processing [13].

A more specifically picture-oriented approach to parallel computation involves special-purpose digital hardware that can perform operations simultaneously on each element of an array of numbers [14]. If an entire array can be shifted in any of the four principal directions and two arrays can simultaneously be added or multiplied elementwise, digital pictures could be convolved very rapidly. Only $(N + M - 1)^2$ shifts and multiplications would be required instead of $N^2 \cdot M^2$ multiplications by the sequential method.

The sequential digital computer with its ability to be programmed for any processing concept is rightly a very useful experimental tool. Even on such a computer it is usually possible to perform simple, logical and shifting operations simultaneously on each binary digit of a computer word. The PAX II picture processing system uses this approach [15].

2.4.1 Comparison of Coherent Optical and Digital Image Processing

1. Flexibility

Coherent optical systems are essentially limited to linear operations on the amplitude transmission variations of a film transparency. On the other hand, digital computers can be used to do linear operations on amplitude, intensity or density. More importantly, digital computers can handle nonlinear operations, removal of noise, corrections for deficiencies of the imaging system and can produce numerical output. Thus, as to flexibility, the digital computer is superior to optical systems, it will probably always be used for preprocessing of images.

2. Capacity and Speed

In a coherent optical system, the film is used as storage, resulting in an enormous capacity. The data on the film can be operated on in parallel, so that the speed is limited, in principle, only by the speed of light. A digital computer has a limited memory, but a great amount of auxiliary memory can be attached to it. Films can also serve as storage if a film scanner is available. However, conventional digital computers operate sequentially on the data, they can perform only one (at most a few) operation(s) at a time. Therefore, operation on

19

large images takes a long time to bring the data into the central processor and to process them.

3. Accuracy

In a coherent optical system there are various sources of errors, such as: imperfect optical components, film grain noise and nonlinearity, thickness variations of film emulsions, errors in spatial filters and imperfect alignment of the optical system. These errors are difficult to control and an accuracy of only 3 to 5 percent can be expected in a coherent optical system. In digital processing, there are inherent errors due to sampling and quantization. These errors, however, can be made arbitrarily small by increasing sampling rate and quantization levels. In practice, the accuracy of digital computer image processing is limited by the resolution of the imaging system or film scanner.

In summary, the main advantages of a coherent optical system are its storage capacity and processing speed and the main advantages of a digital computer are its flexibility and accuracy. Coherent optical systems are suitable for performing linear operations such as Fourier transforms and linear filtering on large images. When accuracy or nonlinear operations are required, the digital computer has to be used. A future image processing laboratory could benefit from a combined use of coherent optical systems and digital computers. Such an image processing center would include a general-purpose digital computer and various special-purpose digital and optical processors. It is also extremely important that man-machine interaction be made flexible and convenient so that at each critical stage of the processing, the human operator can examine the results obtained and decide on what to do next. The data transfer between a digital computer and a coherent optical system is conceptually possible, however, much analysis and experimental work need to be done.

2.4.2 Digital Image Processing

When pictures are processed by a digital computer, they are usually represented as discrete arrays of numbers, i.e., as matrices, rather than as functions. Any M by N matrix (g_{ij}) with real, nonnegative elements can be thought of as defining a digital picture function g. It can be shown that any picture function is indistinguishable from an M by N digital picture function for sufficiently large values of M and N. In digital image processing a picture can take on only a finite set of values, its grey values are quantized. It can also be shown that any picture function is indistinguishable from a quantized picture function, provided that sufficiently many levels are allowed [7]. This guaranties that the mathematical operations discussed in section 2.3 can be performed on digital pictures.

20

The question how finely a picture must be sampled in order to preserve all its information is answered by the sampling theorem. If the samples are taken sufficiently close to each other, the sampled data are an accurate representation of the original function in the sense that this function can be reconstructed with some accuracy by interpolation. For the class of bandlimited functions the reconstruction can be accomplished exactly, providing that the interval between samples is not greater than a certain limit. A function $g(x, y)$ is called bandlimited if its Fourier-transform $G(u, v)$ is zero whenever either $|u|$ or $|v|$ is greater than some number W. The sampling theorem applies to the class of bandlimited functions and states that the sampling interval be not greater than $1/2W$.

In digital image processing linear operations such as convolution, transformation, and filtering can be succinctly expressed in vector space notation. The advantages are a compact notation and the ability to use results derived for one-dimensional problems.

A discrete representation of the convolution integral (10a) can be found by numerical approximation. The simplest formula would be the trapezoidal rule, which applied to (10a) gives

$$g(m\Delta x, n\Delta y) \approx \Delta x \, \Delta y \sum_{i=m-\frac{L}{2}}^{m+\frac{L}{2}} \sum_{j=n-\frac{L}{2}}^{n+\frac{L}{2}} f(i\Delta x, j\Delta y) \, H\left[(m-i)\,\Delta x, (n-j)\,\Delta y\right]$$

(34)

$$m = \frac{L}{2}, \frac{L}{2}+1, \cdots, M - \frac{L}{2} - 1$$

$$n = \frac{L}{2}, \frac{L}{2}+1, \cdots, N - \frac{L}{2} - 1$$

where $x = x_M/M$, $y = y_M/N$, $L = 2x_L/\Delta x = 2y_L/\Delta y$. Hence, the processed picture is of smaller dimension than the original picture. Translation and including

all quadrature factors in $\tilde{h}$ yields

$$g(m, n) = \sum_{i=m}^{m+L-1} \sum_{j=n}^{n+L-1} f(i, j)\, \tilde{h}(m - i + L, n - j + L) \qquad (35)$$

$$m = 1, 2, \cdots, M'$$

$$n = 1, 2, \cdots, N'$$

where $M' = M - L$ and $N' = N - L$.

Let now $\underline{f}$ denote a column vector of dimension $M \cdot N$ with the elements of the digital picture array $f(i, j)$ ordered columnwise. With a similar interpretation of $\underline{g}$ and $\underline{n}$ (35) can be written in vector form as

$$\underline{g} = B\underline{f} + \underline{n} \qquad (36)$$

where B is an $M' N' \times M N$ matrix partitioned as

$$B = \begin{pmatrix} B_{1,1} & B_{1,2} & \cdots & B_{1,L} & 0 & 0 & 0 \\ 0 & B_{2,2} & \cdots & B_{2,L} & B_{2,L+1} & 0 & 0 \\ 0 & 0 & & & & 0 \\ \vdots & & & & & 0 \\ 0 & & & 0 & B_{N', N-L+1} & B_{N'N} \end{pmatrix} \qquad (37)$$

22

The $M' \times M$ submatrix B_{pq} is given by

$$
B_{pq} = \begin{pmatrix}
\tilde{h}(K,p\text{-}q+L) & \tilde{h}(K\text{-}1,p\text{-}q+L) & \cdots & \tilde{h}(1,p-q+L) & 0 & \cdots & 0 \\
0 & \tilde{h}(K,p\text{-}q+L) & & & & & \\
& & 0 & & & & \\
& & & & & & 0 \\
0 & 0 & & \tilde{h}(L,p\text{-}q+L)\cdots\tilde{h}(1,p\text{-}q+L) & & &
\end{pmatrix}
\tag{38}
$$

for

$$1 \leq p \leq N'$$

$$p \leq q \leq L + p - 1$$

The structure of B is given by

$$
B_{pq} = B_{p+1,\,q+1}
\tag{39}
$$

and consequently, the rows of B are shifted versions of the first row. The convolution operation in vector form requires MNL^2 operations if the zero multiplications of B are avoided. The block diagram of the image forming process expressed in vector form is shown in Figure 3. The degradation is modeled by
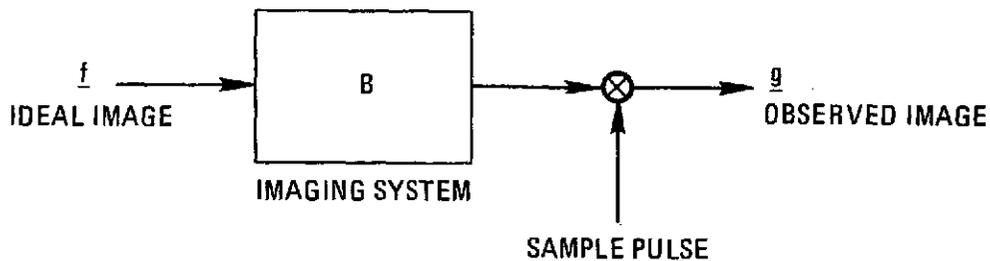
Figure 3. Digital Image Formation

23

the matrix B, the observable vector $\underline{g}$ is of smaller dimension than the vector $\underline{f}$ representing the ideal image. Hence, the solution of (35) is generally not unique.

2.4.3 Discrete Convolution Theorem

The discrete convolution theorem is of great importance in digital filtering. It makes possible the convolution of two digital images by the product of their discrete Fourier transforms. Hence, the discrete convolution theorem is the basis for the application of the fast Fourier transform (FFT). The important aspect to be noted in connection with the discrete convolution theorem is that while the convolution takes a number of operations proportional to $M \times N \times L^2$ by the direct method, the FFT method requires a number of operations proportional to $MN(2 \log M + 2 \log N + 1)$. Operation means, throughout this report, a multiplication followed by an addition. The Fourier transform method consists in transforming the functions f and h, multiplying the transforms and transforming the product back.

For the discrete convolution theorem to hold, it is necessary that the discrete functions f and h be periodic. These periodic functions are defined in the manner of Cooley [34]:

$$f_p(x, y) \;=\; \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} f(x + jP, \, y + kP) \qquad (40)$$

$$h_p(x, y) \;=\; \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} h(x + jP, \, y + kP) \qquad (41)$$

$$g_p(x, y) \;=\; \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} g(x + jP, \, y + kP) \qquad (42)$$

The subscript p on a function denotes the periodic function formed by the superposition of the nonperiodic function shifted by all multiples of a fundamental period. The period P is the smallest power of 2 greater than or equal to Q, where $Q = \max(M, N)$. Since the original functions are zero outside their regions of definition, a choice of P suitably large will lead to no overlap of the functions when the shifted replications are summed. Thus, the $M \times N$ image

24

array $f(m, n)$ and the $L \times L$ impulse response array $h(k, l)$ are embedded in the upper left corner of $P \times P$ arrays of zeros.

The circular convolution of the periodic functions is defined as

$$g_p(m, n) = \sum_{i=1}^{P} \sum_{j=1}^{P} f_p(i, j) \, \tilde{h}_p(m - i + 1, n - j + 1) \tag{43}$$

$$m = 1, 2, \cdots, P$$

$$n = 1, 2, \cdots, P$$

The circular convolution is equivalent to an ordinary convolution for a proper choice of P because the original functions were assumed to be identically zero beyond their regions of definition. The periodic replications contain these zero values and cancel the periodic "wrap-around".

Equation (43) can be written in vector form as

$$\underline{g}_p = B_p \underline{f}_p \tag{44}$$

where $\underline{g}_p$ is a column vector of size P^2 whose elements are $g_p(m, n)$, $\underline{f}_p$ is a column vector of dimension P^2 and B_p is a matrix of size $P^2 \times P^2$ which can be partitioned into $P \times P$ submatrices $B_{s,t}$.

$$\begin{pmatrix}
B_{1,1} & 0 & 0 & 0 & B_{1,P-L+2} & B_{1,P} \\
B_{2,1} & B_{2,2} & 0 & 0 & 0 & B_{2,P} \\
& & & & & \\
B_{L,1} & B_{L,2} & & & & B_{L-1,P} \\
0 & B_{L+1,2} & & & & 0 \\
0 & 0 & & & & \\
& & & & & \\
& & & & & 0 \\
0 & & B_{P,P-L+1} & B_{P,P-L+2} & & B_{P,P}
\end{pmatrix} \tag{45}$$

25

Each row and column of B_p contains L non-zero submatrices. The structure of B_p is given by

$$B_{s,t} = B_{s+1,t+1} \tag{46}$$

with

$$s = s \bmod P$$

$$t = t \bmod P$$

It should be noted that indices are to be interpreted modulo P. Therefore, a row can be obtained by a circular right shift of the row above it. Furthermore, the first row is a circular right shift of the last row. Such a matrix is called a circulant matrix. The circular behavior is a direct consequence of the fact that $h_p(i,j)$ is a periodic function.

It can be shown that the Fourier transform basis vectors are eigenvectors of the circulant matrix B_p [35]. Now (44) can be written in terms of the diagonal representation of B_p

$$\underline{g}_p = W D_h W^{-1} \underline{f}_p \tag{47}$$

where D_h is the diagonal matrix of eigenvalues of B_p and W is a matrix whose columns are the eigenvectors of B_p. By rearranging this equation into

$$W^{-1} \underline{g}_p = D_h W^{-1} \underline{f}_p \tag{48}$$

it can be seen that $W^{-1} \underline{g}_p$ and $W^{-1} \underline{f}_p$ are the Fourier transforms of $\underline{g}$ and $\underline{f}$ respectively and the diagonal matrix D_h is the Fourier transform of $h_p(m,n)$. Consequently (48) is a term-by-term product of Fourier transform and is the frequency domain representation of the circular convolution (44). For a proper choice of P the circular convolution is equivalent to the aperiodic convolution. This proves that the discrete convolution can be computed by the fast Fourier transform algorithm.

26

2.4.4 Discrete Fourier Transform

The discrete Fourier transform of a periodic digital picture function f_p (40) is given by

$$F_p(m,n) = \frac{1}{N} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f_p(j,k) \, e^{-2\pi i \left[\frac{jm}{M} + \frac{kn}{N} \right]} \tag{49}$$

$$m = 0, 1, \cdots, M-1$$

$$n = 0, 1, \cdots, N-1$$

The discrete inverse transform is given by

$$f_p(j,k) = \frac{1}{M} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} F_p(m,n) \, e^{2\pi i \left[\frac{jm}{M} + \frac{kn}{N} \right]} \tag{50}$$

$$j = 0, 1, \cdots, M-1$$

$$k = 0, 1, \cdots, N-1$$

For this Fourier series representation to be valid the image must be considered periodic horizontally and vertically as shown in Figure 4. The right hand side and the left hand side, as well as the top and the bottom of the image are adjacent. Spatial frequencies along the coordinate axes of the transform plane arise from these transitions. Although these are false spatial frequencies from the stand-point of being necessary for representing the image within the image boundary, they do not impair reconstruction. On the contrary, these spatial frequencies are required to reconstruct the shape boundaries of the image.

2.5 EVALUATION OF IMAGE QUALITY

Image quality is an important concern in image processing. Its meaning can be clarified by understanding what is to be measured when dealing with images and
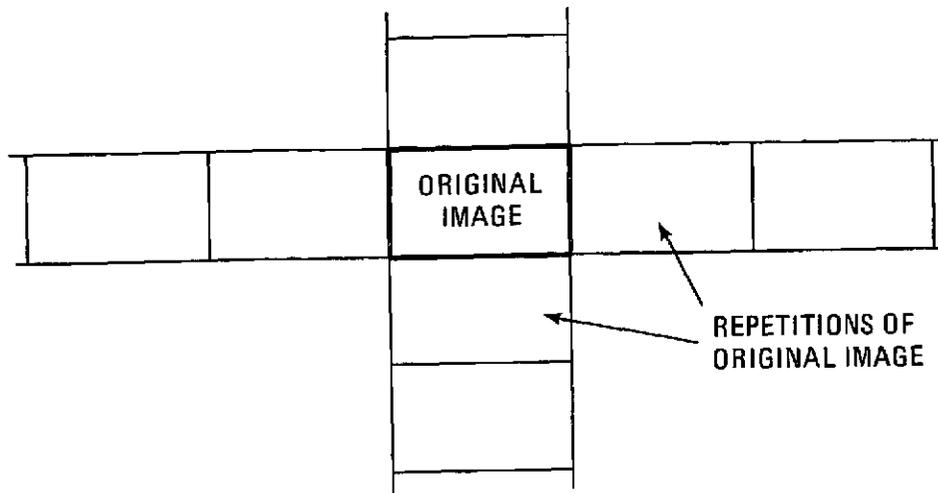
Figure 4. Fourier Series Representation of an Image

by strengthening the bridge between the physical and visual aspects of many image processing issues. The availability of sophisticated digital methods supports the need for precision. There is, however, the realization that the lack of standards for reading images into and writing images out of digital form can bias the effectiveness of a process and can make uncertain the comparison of results obtained at different installations.

Whenever a picture is converted from one form to another, e.g., scanned or displayed, the question arises how faithfully the information contained in the input picture is conserved in the output picture. It is difficult to find subjective distortion measures and a part of the problem stems from the fact that physical and subjective distortions are necessarily different. One of the difficulties in specifying image quality is that the intended use of the images is not well defined and it is unlikely that a wide variety of tasks can be covered by a single quality criterion. Consideration of the purpose for which an image was recorded and of the interaction of images with the human observer will help to evaluate image quality.

Where the goal is extraction of information and where the image is to be processed prior to viewing, the information content of the image is a true evaluation criterion [19]. Since any real imaging system will have limitations imposed by resolution and noise, there is a fundamental limit to the information contained in an image. The human visual system is incapable of efficient extraction of information which is contained in a degraded image. The improvement achieved by processing can be evaluated by comparing the ability of the human observer to extract information from the image before and after processing. This criterion

28

depends on the specific visual task which is to be performed. When the specific task is not defined, a satisfactory evaluation of image quality is difficult to achieve.

Some authors have indicated that a signal-to-noise ratio concept should be useful in the prediction of image quality. An attempt to unify image quality criteria for resolution and detection used statistical decision theory [20].

2.6 NOISE CHARACTERISTICS AND REMOVAL

The practical limit to all quantitative and photointerpretive measurements on a properly encoded image is the presence of noise. Enhancement processes such as filtering to improve image resolution can sharpen features only at the expense of over-all signal-to-noise ratio. Therefore, one of the most important initial steps in digital image processing is the suppression of noise, so that subsequent restoration, enhancement and extraction operations can be performed on maximum signal-to-noise ratio imagery to achieve optimal results.

Many noise sources exist in imaging systems ranging from random, wideband and thermal noises to highly structured periodic noises. The precise separation of any noise from the data of a single frame must be based on quantifiable characteristics of the noise signal that distinguish it uniquely from the other image components. In space it is generally not possible to obtain multiple imagery in order to use frame-averaging techniques. In most real situations only statistical information about the various image components is available and, thus, their separation is approximate. The essence of noise removal is to isolate and remove the identifiable and characterizable noise components as vigorously as possible, so as to do a minimum of damage to the actual image data. In most cases, the errors introduced to the real signal by the removal process, while small, vary from point to point and can only be measured if detailed knowledge about the scene being photographed is available. Certainly the efficacy of noise removal is data dependent.

The main types of structured noise appearing in images are periodic, long-line and spike noises. Periodic noise arises from the coupling of periodic signals related to the raster-scan and data-sampling drives into the imaging electronics. For typical spacecraft systems, these periodic noises exhibit phase coherence over times that are long compared to the frame time of the camera. For this reason, the periodic noise appears as a two-dimensional pattern exhibiting periodicity along the scan lines and perpendicular to them [49]. A useful method for characterizing this periodicity is in terms of a Fourier decomposition. A first order removal of periodic noise components can be achieved by filtering

in the two-dimensional Fourier domain. The SMIP system provides the program FOURIER2 for computing the two-dimensional Fourier-transform of an image, the program FOURPIC to display the amplitude of the Fourier spectrum and the program FREQFILT for two-dimensional filtering with various filters in the frequency domain.

Long-line or streak noise is produced by a variety of mechanisms such as gain variations, data outages, and tape recorder dropouts. This type of noise becomes apparent as horizontal streaks especially after removal of periodic noise. It is characteristic for images from the ERTS-1 multispectral scanner. The characteristic that distinguishes streak noise from the actual scene is its correlation along the scan-line direction and the lack of correlation in the perpendicular direction. This distinction is not complete since linear features are present in some natural scenes and noise removal based on this characteristic may result in major damage to the true signal in regions that contain scene components resembling the noise.

A technique to correct for streak noise is to compare the local average intensity value of lines adjacent and parallel to the streaks with the average value of the streak itself and to apply a gain factor to account for any differences [50]. A multiplicative rather than an additive correction is applied because the physical origin of the noise is multiplicative (magnetic tape dropouts). This correction is particularly data dependent in its effect and although providing a global improvement it may introduce artifacts in the detail. The program ERTSFIX in the VICAR and SMIPS libraries performs streak corrections on ERTS-MSS images.

Spike noise is caused by bit errors in data transmission or the occurrence of temporally sharp disturbances in the analog electronics. It produces isolated picture elements that deviate significantly from the surrounding data. Spike noise can be removed with a simple technique. Each picture element is examined and if it is significantly above or below each of its neighbors, it is replaced by the average neighboring intensity. Using the digital computer to isolate and remove various structured noise components from the raw images can considerably improve the signal to noise ratio. This preprocessing produces already an enhancement that allows analysis of detail closer to the resolution limits of the imaging system.

3. IMAGE RESTORATION

The extraction of features or objects from an image may not be possible if it is severely distorted. Restoration compensates for the distortions and attempts reconstruction of the original image. Degradations may be caused by the various factors discussed in section 2.2. In terms of the model for image formation expressed by (11) the restoration task can be defined as follows: given g, utilize the a priori information about h, f and n to make a good estimate $\hat{f}(x, y)$ of f. The various restoration schemes differ from each other in the assumed a priori information as well as in the criterion by which the goodness of the estimate is judged.

Frequently, the a priori information is little more than a reasonable assumption. All that is usually known about the desired picture is that $f(x, y)$ is nonnegative, bounded and zero outside some region. Likewise, a priori information about the noise is usually very meager and often it is assumed that n has a known constant spectral density. The impulse response h can be known, either from theory or calibration of the imaging system, or it may be unknown. It is assumed that the restored image is to be viewed by a human observer. This requires that the restoration should redisplay the information in the degraded image in such a way that it enables a human observer to identify the original objects with as much detail as possible. Because of the restrictions discussed in section 2.5 on image quality, no absolute criterion by which the goodness of the estimate is to be judged can be specified. Each restoration scheme assumes some intuitively reasonable criterion of goodness. This confesses the ignorance of the optimum tradeoff between resolution and noise.

Restorative techniques can be classified into 1) inverse filtering, 2) optimal filtering, 3) constrained deconvolution, 4) restoration with correction tables and interpolating functions and 5) other methods. In the discussion of these methods it is assumed that the impulse response of the degrading system is known.

3.1 INVERSE FILTERING

Inverse filtering attempts perfect restoration without regard to noise. Fourier-transformation of the noise-free model (10) gives

$$G(u, v) = H(u, v) \cdot F(u, v) \tag{51}$$

31

in the spatial frequency domain. Formally, the restored image $\hat{f}$ is obtained by reinverting

$$\hat{f}(x, y) \;=\; \mathcal{F}^{-1}\left\{\frac{G(u, v)}{H(u, v)}\right\} \tag{52}$$

provided H does not vanish at any point (u, v). The method fails if H has zeros at spatial frequencies within the range of interest. The second flaw is that H decreases rapidly for large values of u and v whereas noise has a fairly uniform spectral distribution. From (52) it is apparent that the restoration enhances high-frequency noise. Modifications have been suggested to overcome these drawbacks [19]. A simple and efficient method is to limit the amplitudes of H^{-1} at higher frequencies in order to limit enhancement of high-frequency noise or to replace H^{-1} by zero in the range of (u, v) over which the noise is larger than the signal.

3.2 OPTIMAL FILTERING

A way to avoid the arbitrariness of the inverse filtering approach is to minimize the discrepancy between f and $\hat{f}$. The measure of this discrepancy should ideally correspond to that of the human visual system. Because this is not known in detail an alternative is to find the optimum restoration for a simpler objectively defined criterion such as the minimum mean-square-error (MSE) criterion for which the optimum restoration can be computed. From the point of view of image restoration for a human observer minimizing the MSE is an inadequate criterion. It is well known that the eye demands much more faithful reproduction of regions where the intensity changes rapidly than of regions with little change. It is also known that the sensitivity of the eye to a given error in intensity depends strongly on the intensity. The minimum MSE criterion, on the other hand, weighs an error independently of the intensity at which it occurs. It has been shown, however, that the method is capable of producing very good restorations [21, 22].

In the following, a brief outline of the mathematical derivation of the optimum restoration for digital images will be given. Picture and noise are considered to be members of the random processes $\{f\}$ and $\{n\}$, respectively. Given (36), a restoring matrix W has to be determined such that the estimate

$$\underline{\hat{f}} \;=\; W\underline{g} \tag{53}$$

32

minimizes the error measure

$$\epsilon = E\left\{(\hat{\underline{f}} - \underline{f})^T (\hat{\underline{f}} - \underline{f})\right\} \tag{54}$$

where E denotes the expected value. If it is assumed that $\underline{f}$ and $\underline{n}$ are uncorrelated stationary Gaussian random vectors with zero mean, the error measure becomes

$$\epsilon = tr\left[(WB - I) K_f(WB - I)^T + WK_n W^T\right] \tag{55}$$

Here tr denotes the trace of a matrix, I is the identity matrix and K_f and K_n are the covariance matrices $E(ff^T)$ and $E(nn^T)$ respectively. The minimum for ϵ in (55) occurs for

$$W = K_f B^T (BK_f B^T + K_n)^{-1} \tag{56}$$

Note that for $K_n = 0$ the optimum filter becomes the inverse filter. Figure 5 shows a block diagram of Wiener filtering.

As to the assumptions, the Gaussian hypothesis cannot be valid because the {f} and {n} processes are ensembles of positive functions. Thus, the optimum restoration filter derived does not minimize the MSE. However, of all estimates obtainable by spatially invariant filtering of $\underline{g}$, the linear estimator (56) gives the least MSE. Generally Wiener filtering represents optimum processing when a meansquared criterion is assumed to be optimal and the processing is performed independent of the human visual system.
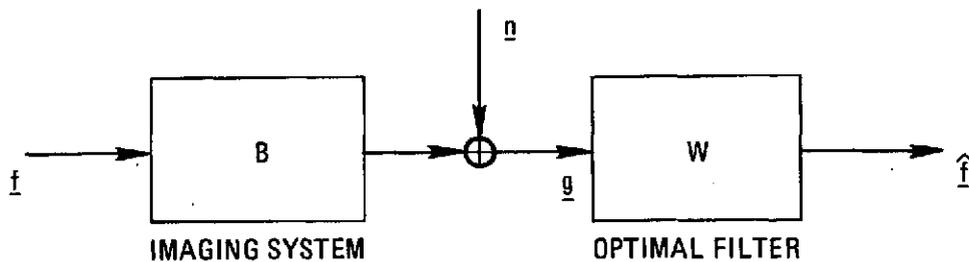
Figure 5. Optimal Filtering

3.3 CONSTRAINED DECONVOLUTION

This method allows the experimenter to search for the optimum restoration by using a digital computer in an interactive manner [23]. The experimenter specifies certain parameters and tolerances, on the basis of which a restoration is produced. If the result is not satisfactory, the experimenter may try a new set of specifications based on his previous observations and on his subjective evaluation of the obtained restoration. The use of introspective evaluation in a feedback loop can be easily accomplished in the SMIP system [24].

The method is based on the vector form (36) of the convolution relation. The effect of noise and a priori information is introduced by specifying lower and upper limits for each component of these vectors. A nice feature of this method of restoration is that the constraints prevent the restored image from having negative values. None of the other methods can guarantee this in the presence of noise.

3.4 CORRECTION TABLES

Rectification of geometric distortion and photometric errors which are spatially and temporally stable involves changing coordinates and brightness values throughout the image. Corrections can be tabulated for selected points and suitable interpolation formulas used at intermediate points. Displacements for straightening the image coordinate raster can be obtained using a two-dimensional test grid. Photometric errors can be corrected by using the results for uniformly illuminated test fields. In most cases the correction involves nonlinear, position dependent operations [25]. This restoration technique applies to degradations not described by the linear position-invariant image forming model which is the basis for the other restoration methods. The program GEOM in the VICAR system provides a capability of correcting for distortion in the location of picture elements.

3.5 OTHER METHODS

In purely mathematical terms, the restoration of a degraded photograph is equivalent to the solution of a Fredholm integral equation of the first kind. This kind of equation arises in many physical problems and various methods have been proposed for numerically solving such equations [26, 27].

4. IMAGE ENHANCEMENT

An optimally restored picture may not be the most efficient form for visual data which are to undergo further processing and interpretation. Nonrestorative pre-processing may be necessary to aid either human or automatic photointerpretation. The goal of image enhancement is to improve the quality of an image for human viewing without recourse to knowledge of degrading phenomena. Thus, enhancement operators change the image rather than restore it. Both mathematical and heuristic techniques are utilized in this process with the emphasis on the "viewing" of the image for extraction of information that may not have been so readily apparent in the original. Since the most common method for evaluating enhancement results is subjective human evaluation, understanding the evaluation process is also important for each enhancement application. Image quality is, however, difficult to evaluate as was discussed in section 2.5.

The problem is to model the human viewing process such that an optimum display can be computed for the human as a communication sink. Although work in this area is limited, some pleasing results have been obtained by certain combinations of logarithmic, linear and exponential processes. The problem is that no viable fidelity criterion for the human eye has been developed. If such a criterion would be known, then the power of information and communication theory could be directed toward optimal use of the channel between the image and man.

The heuristic techniques try to attenuate or discard irrelevant material and at the same time to emphasize or clarify features and objects of interest. Attenuation is achieved by smoothing or integrating operators, but these also tend to blur detail. Emphasis is usually accomplished by sharpening or differentiating operators, or by feature or contrast enhancers, but these also tend to accentuate noise. Smoothing without obliteration of the relevant and sharpening without amplification of the irrelevant are the desired ends [28].

Most such techniques are linear and position-invariant and fit well into linear systems theory. Therefore, enhancement of regions and contours can be described in spatial or spectral terms. Pictures are convolved with suitable spatial masks or their spectra are multiplied by frequency filters. While restoration deals with entire spectra and means of obtaining well-behaved inverse operators, enhancement deals with parts of spectra.

A human interpreter using adequately enhanced images can in many cases perform a successful classification of pictorial data without usage of time consuming mathematical-statistical classification procedures. The various enhancement techniques can be organized as follows:

1. Spatial filtering

2. Frequency filtering

3. Contrast and contour enhancement

4. Histogram equalization

5. Color enhancement

6. Other enhancement techniques

Spatial filtering and frequency filtering are linear techniques based on the convolution integral (10) and the convolution theorem. The other methods involve linear and nonlinear operations and heuristics.

4.1 SPATIAL FILTERING

4.1.1 Smoothing of Regions

Smoothing can be used in order to suppress noise that is present in a picture. Primitive smoothing operations use convolution (spatial masking). If the digital picture is represented by an $M \times N$ matrix

$$F = \left\{ f_{ij} \mid 0 \leq i \leq M, \ 0 \leq j \leq N \right\} \tag{57}$$

and the mask by the matrix

$$W = \left\{ w_{ij} \mid 0 \leq |i| \leq m, \ 0 \leq |j| \leq n \right\} \tag{58}$$

with $m \leq M$, $n \leq N$, then the transformed image G is produced by convolving W with F. The elements of G are weighted local averages or finite differences depending on the algebraic signs of the mask weights:

$$g_{ij} = \sum_{p=-m}^{m} \sum_{q=-n}^{n} w_{pq} f_{i+p, j+q} \qquad m \leq i \leq M - m, \ n \leq j \leq N - n \tag{59}$$

For border elements with subscripts $1 \leq i < m$ or $M - m < i \leq M$ and $1 \leq j < n$ or $N - n < j \leq N$ special provisions have to be made.

When mask entries are identical, each point of the transformed image is a simple average of original values in its neighborhood. This is just the convolution of the picture with a function that has the value $1/A$ inside the neighborhood and 0 outside, where A is the area of the neighborhood.

When the mask entries are unequal, the average is weighted. The weighting pattern is to reflect specific noise and correlation characteristics of the image. Weights that decrease monotically from the central position are most commonly used. The larger the mask size, the more homogeneous is the smoothed image, but the poorer the rendition of fine detail and significant transitions. The mask size has to be smaller than the smallest detail to retain. It is desirable to introduce procedures in which the decision to smooth or not to smooth and the nature of the smoothing vary from one point to another. The simplest class of such operations combines averaging with thresholding. If the average grey value in some neighborhood of a point exceeds its grey value by more than a threshold t_1 then the grey value is replaced by the average, if it exceeds a threshold t_2 do not smooth [7]. This procedure can be used to clean up "pepper and salt" noise (isolated dark points in light regions and vice versa). Good results have been obtained for even small (3 × 3) neighborhoods. Smoothing can be accomplished with the FILTER program in the SMIP system.

4.1.2 Enhancement of Edges (Sharpening)

Enhancement of edges is an important technique in image processing since borders contain a significant portion of the useful pictorial information. The most effective non-restorative sharpening methods use some type of spatial differentiation. If sharpening is required only in some particular direction, the directional derivative can be used. Sharpening in every direction can be accomplished by taking the derivative in the gradient direction [28]. This maximal directional derivative is equal to the squares of the derivatives in any pair of orthogomal directions. For a digital picture (57) the crudest approximation would be

$$(\nabla f)_{ij} = \sqrt{(f_{i+1,j} - f_{ij})^2 + (f_{i,j+1} - f_{ij})^2} \qquad (60)$$

Another useful combination of derivatives is the Laplacian

$$\nabla^2 f(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

For a digital picture the approximation using an 8-element neighborhood is

$$(\nabla^2 f)_{ij} = 8 f_{ij} - \left[f_{i-1,j} + f_{i-1,j} + f_{i,j-1} + f_{i+1,j-1} \right.$$

$$\left. + f_{i+1,j} + f_{i+1,j+1} + f_{i,j+1} + f_{i-1,j+1} \right] \tag{61}$$

yielding the mask

$$W_L = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix} \tag{62}$$

The Laplacian is, thus, approximated by convolving the picture with a mask having a positive peak that is surrounded by a negative annular valley, with the values chosen so that the integral of the mask is zero. The convolution of such a mask with the picture is zero in regions where the picture is constant or linear but not at edges across which the second derivative is nonzero.

The weakness of these techniques is their overemphasis of accidental fluctuations. The spatial mask for sharpening must be larger than the largest differtial feature to be retained. Edge enhancement with the SMIP system can be accomplished with the FILTER program or with LAPLACE for the mask W_L.

4.2 FREQUENCY FILTERING

Two-dimensional spectral analysis can be used to systematically describe the effects of spatial mask in terms of resolution and noise. Since an image can be represented by a series of orthogonal functions, it may be easier to deal with the coefficients in the series than with the original picture. The Fourier-transform is the most important functional representation used. Each mask is

dual to a filter in the frequency domain and inversely, hence masks can be classified according to the action of the corresponding frequency filters. A qualitative classification into 1) low pass, 2) high pass, 3) low emphasis, 4) high emphasis and 5) band pass filter is convenient. Masks should be symmetric about the central element if isotropy is desired and asymmetric if directional information is to be emphasized.

The function of the low pass filter is to retain low frequency information and reject high frequencies. Thus, low pass filtering is related to smoothing. The problem is to select the cutoff frequency so that the filter is narrow enough to remove noise and wide enough to preserve detail. The purpose of the high pass filter is to remove low frequencies corresponding to background and to enhance edges. Thus, high pass filtering is related to sharpening. A bandpass filter retains frequencies between two limits and is equivalent to a low pass, high pass pair. High and low emphasis filters try to compensate for spectral losses and to stretch contrast, but not at the risk of disregarding useful information. The simplest low emphasis filter is an equal weight mask in the spatial domain. In the frequency domain this mask corresponds to a sinx/x filter. One of the most useful image enhancement filters is the high frequency emphasis filter. This filter passes a certain amount of low frequency information and emphasizes the high frequency information producing an image with lower contrast but enhanced edge information. The program FREQFILT in the SMIP system can be used to apply a variety of filters to the Fourier transform of an image.

4.3 CONTRAST AND CONTOUR ENHANCEMENT

4.3.1 Grey Scale Transformations

Selective contrast enhancement can often be obtained by transformation of the grey scale or by grey scale requantization. Generally, quantization schemes involve a fixed number of grey levels apportioned uniformly over the range of greyvalues. This method can create false contours and may make the quantized approximation of the original picture unacceptable, since real objects in the picture may be concealed [7].

If the grey scale statistics of the image can be determined, it is possible to quantize optimally by using finely spaced quantization levels in the most informative part of the scale. This increases the average accuracy of the quantization without increasing the number of levels.

The choice of quantization levels can be made to depend on the nature of the pictures being analyzed. The eye is relatively poor at estimating the grey levels

immediately adjacent to a sharp edge on a picture, so that coarse quantization can be used near such edges. If too few quantization levels are used in a smooth region of the picture, the transitions from level to level will show up conspicuously as false contours. The SMIP system provides commands for a dynamic selection of quantization levels.

Another type of contrast enhancement is effected by linear or nonlinear transformations of the grey scale. Portions of the grey scale may be stretched or compressed. Contrast enhancement by linear stretching over the brightness range has its limit if there is already a large difference in brightness between light and dark portions of the picture. If, however, a picture is examined in each local region for fine variations, and only those brightness values near the fine structure are stretched, then the details can be brought out without the saturation caused by indiscriminate contrast stretching [25]. The program STRETCH may be used to perform linear and nonlinear transformations of the grey scale.

4.3.2 Enhancements of Contours and Curves

Contour information is usually extracted after edge enhancement. Connected regions can be assigned continuous connected borders which are tracked and encoded. It may be desirable to smooth the borders by removing points which cause unnecessary perturbations or sharp variations in curvature. Curves can be approximated by interpolating straight-line segments between sample points. A type of polygonal approximation is known as chain encoding [29]. A polygonal contour of minimal length can be fitted to a closed contour by using nonlinear programming [30].

Contours are usually extracted by thresholding the original picture or transformations of it or by using derivatives. The resultant contours are often thick or discontinuous. The enhancement problem is to provide continuation and closure by filling gaps and eliminating perturbations and by thinning. Thinning processes are to reduce connected elongated objects to line-like representations which preserve connectivity. Many algorithms have been suggested [31,32].

4.4 HISTOGRAM EQUALIZATION

This enhancement technique uses as many grey levels as possible for the display of an image by uniformly distributing the grey values. The technique is equivalent to a nonlinear position-invariant operator on the intensity scale. Equalization is useful for images with heavily biased histograms toward an end of the grey scale range. It makes subtle changes more evident in regions for which such changes occur most frequently, while losing subtle intensity changes in

other regions [33]. The program SPREAD in the SMIP system performs histogram equalization on an image.

4.5 COLOR ENHANCEMENT

The human eye is able to separate more colors than brightness levels. Therefore, image processing techniques should include color images. One specific technique, called pseudo color, involves the use of color in the presentation of intensity images. The objective is to aid human interpretation by representing a black-and-white image in color and taking advantage of the human visual system to perceive color differences. This technique actually generates color where none exists.

For a typical black-and-white image scene the average observer can distinguish simultaneously only about 15 to 20 grey-scale steps from black to white. The use of black-and-white images has the effect of restricting the operation of the visual system to the vertical axis of its color perception space (Figure 6). Hence, the ability of the visual system to distinguish many hues and saturations at each brightness level is unused. Clearly, the visual system using simultaneous brightness and chromatic variation can distinguish many more levels of information than one using the grey scale alone. It can also more readily recognize patterns of constant density when these are replaced by a given color through a pseudo-color process [48].

The enhancement of images by pseudo-color should be carefully distinguished from false-color enhancement schemes. False-color, like true-color, utilizes multispectral information from the original scene, but the wavelength-band is
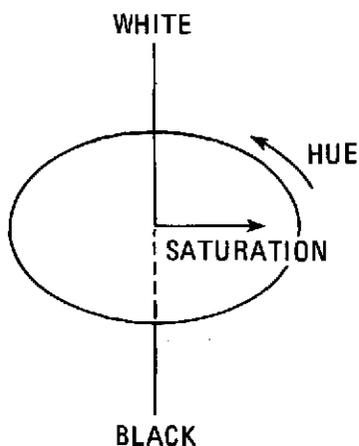
Figure 6. Color Perception Space

41

not necessarily restricted to the visible spectrum. In pseudo-color, the situation is fundamentally different. The energy from any point in the object scene is spectrally nonselective, it differs from other points only in intensity. The pseudo-color process transforms the recorded black-and-white image into color in a manner that controls the relationship between the color in the final image and the corresponding intensity in the original. Proper choice of this relationship permits full use of the abilities of the human visual system to use hue and saturation in addition to brightness for the purpose of discrimination. Such a technique is described in [45]. The program FOTO in the SMIP system transforms a picture into a form suitable for generating a pseudo-color image on the EIS machine.

Another enhancement technique, using false color, can be used to estimate the gradient in each component of the vector-valued picture function. Summing the magnitudes of the estimates will enhance boundaries between regions of different colors even if regions had the same brightness in the black and white pictures. Extraction algorithms can also be extended to vector-valued picture functions.

4.6 OTHER ENHANCEMENT TECHNIQUES

Subtraction of images is a useful method for amplifying differences between two pictures. Subtraction of an unprocessed picture from a filtered picture, for example, can be used to evaluate the filter itself when changes produced by the filter are rather subtle. After registration, subtraction can be applied to detect temporal changes in images of the same scene.

Another technique which has been found useful is forming the ratios of pictures. This process largely eliminates the brightness components of the original pictures and produces a color display whose color variations are more indicative of material variations than the simple pseudo color displays.

The program PICFUNC in the SMIPS-library allows the function

$$z = \frac{Ax + By + C}{Dx + Ey + F} (Gx + H)^I + J \tag{63}$$

to be performed on two input pictures whose digital intensity values are x and y respectively and z are the elements of the result picture.

The color pictures on the following three pages show the effect of enhancement operations on an ERTS-1 MSS5 image of the Great Salt Lake area. In the first picture the original image (upper left) has relatively low contrast and was linearly contrast stretched to the full range for visual inspection (upper right). High frequency detail is already enhanced by this simple stretching operation. A nonlinear stretch with a cuberoot function was performed to enhance detail in the lower range of the intensity scale (lower right). Comparison of the stretched images with the original reveals considerable detail not visible in the original. The picture with the histogram equalized (lower left) shows also increased contrast over most of the range. A uniform stretching operation was used in order to have a uniform color presentation throughout a series of pictures.

The second color picture shows results of spatial filtering the same ERTS-1 MSS5 image. The picture pair in the left half represents the highpass and lowpass components after filtering with a 5 × 5 filter with unit weights. The lowpass cutoff frequency is $f_c = 1.75$ km^{-1}, therefore, higher frequency components are removed from the lowpass image. The highpass image is obtained by subtracting the lowpass signal from the original and adding a constant. The image pair in the right half shows the highpass and lowpass components after filtering with a 51 × 51 filter with unit weights. Here all frequency components above $f_c = 0.172$ km^{-1} are removed from the lowpass image which shows only the global background variation. In comparison, the highpass filtered image reveals fine edge details. Filtering was performed with the program FASTFIL2 in the VICARS library.

The third picture shows at the top false color images of bands 4, 5 and 7 and the ratios 6/4, 7/5 and 7/6 of the same ERTS scene. Below the ratios 6/4 and 7/5 are shown in pseudo color. The program PICFUNC was used to calculate the ratios.

The black and white image represents the two dimensional Fourier transform of a 512 by 512 area starting at line 685 and column 100 of the original ERTS-1 MSS5 image. The spatial frequencies along the coordinate axes arise from the imager boundaries. The two other pronounced frequency directions are orthogonal respectively to the shore lines and to the feature which crosses the lake.

These pictures demonstrate the flexibility of VICARS and SMIPS for image manipulation. The individually stretched images were combined with CONCAT, pseudo color enhanced with FOTO for the E.I.S. film writer. For the two other pictures pseudo color processing was performed with COLFILT. The blue, green and red images were inserted into background pictures with OPTINS, converted to film on the OPTRONICS film writer and photographically superimposed. The entire analysis, from the original pictorial data to the final color picture can be performed without any programming knowledge using only the operation provided by the system.
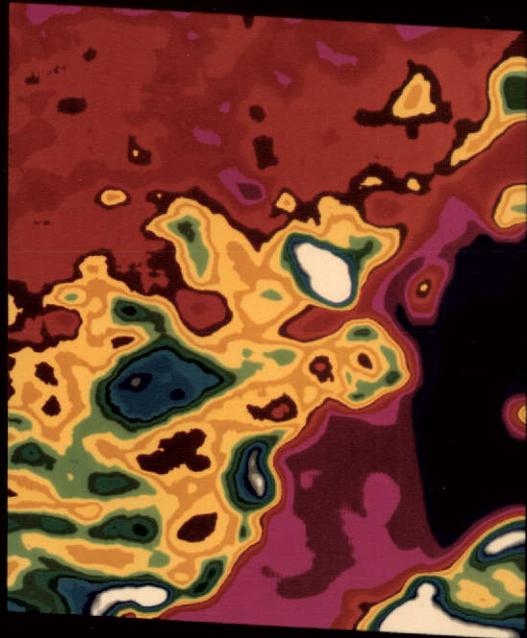
44
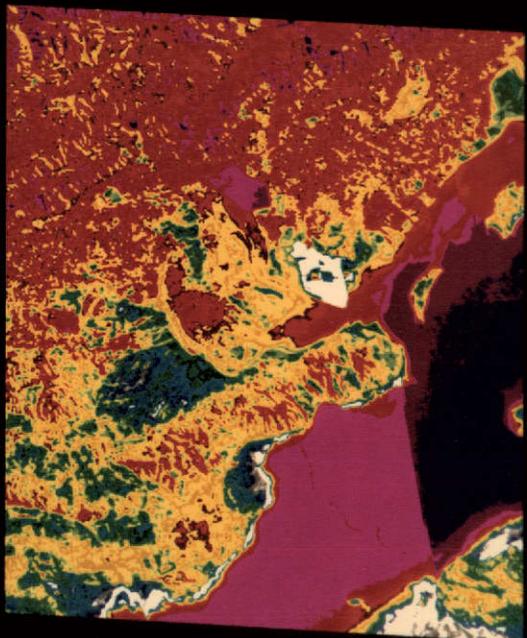
ERTS-1 1069-17415 MSS5 RAW DATA
BELOW: HISTOGRAM EQUALIZED

LINEAR STRETCH
CUBEROOT STRETCH
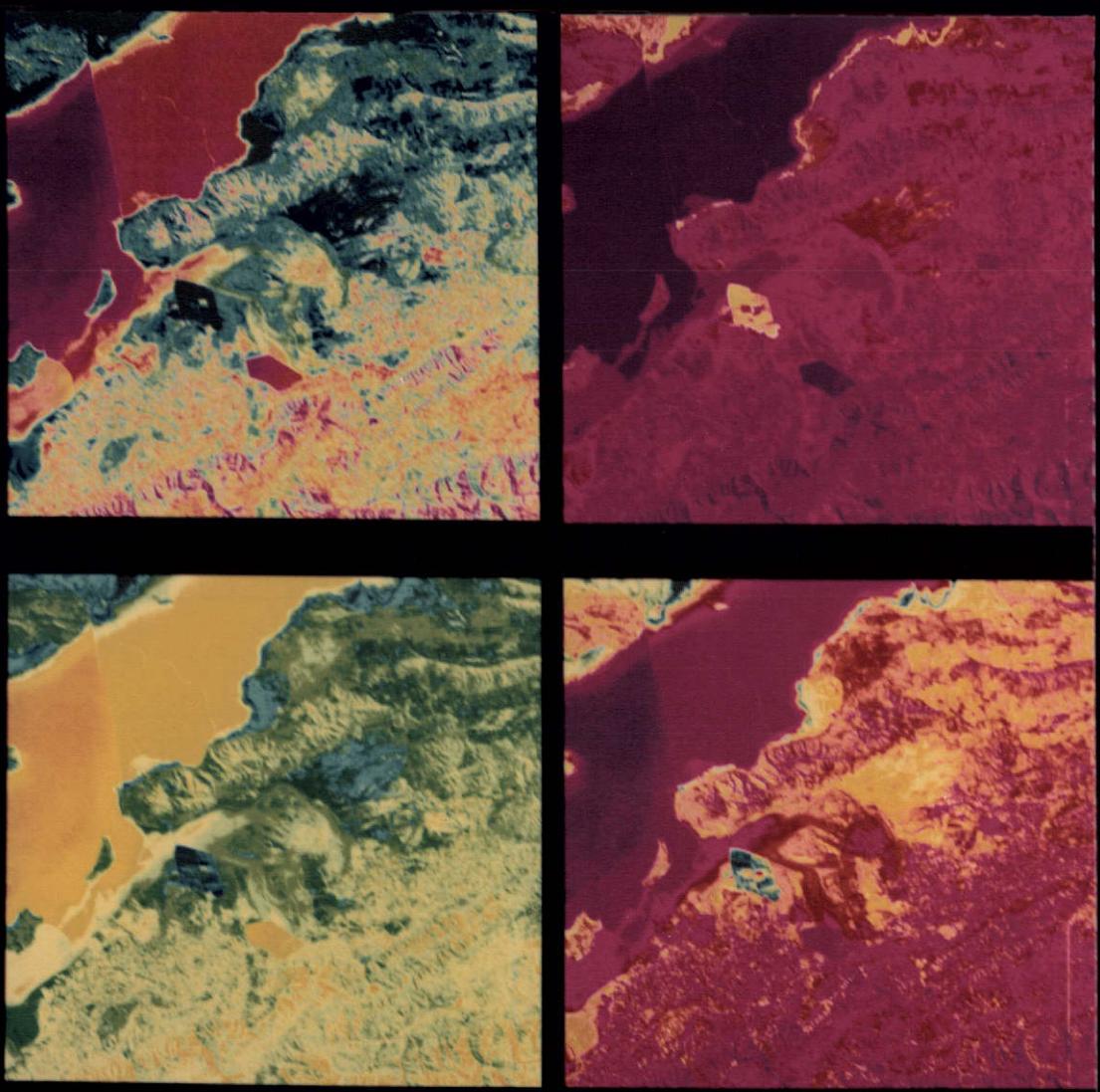
255 247 22A 214 200 1A6 172 15A 144 130 116 102 AA 74 60 4A 32 1A 1 0

46
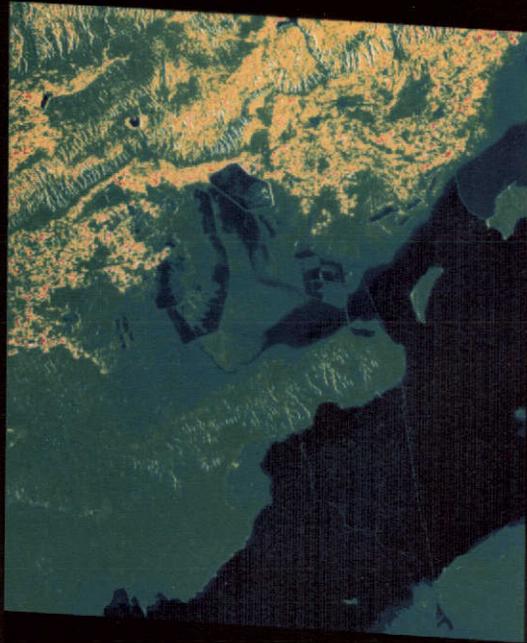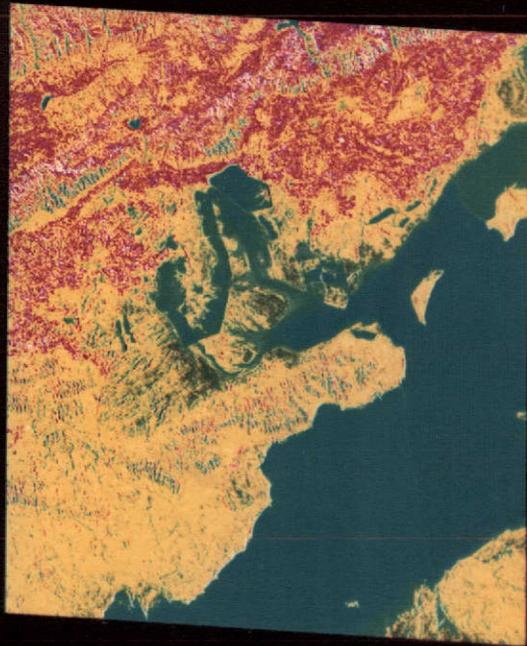
ERTS-1 1069-17414 MSS5 FOURIER TRANS.
LOG-AMPLITUDE 512X512 SL=685 SS=100
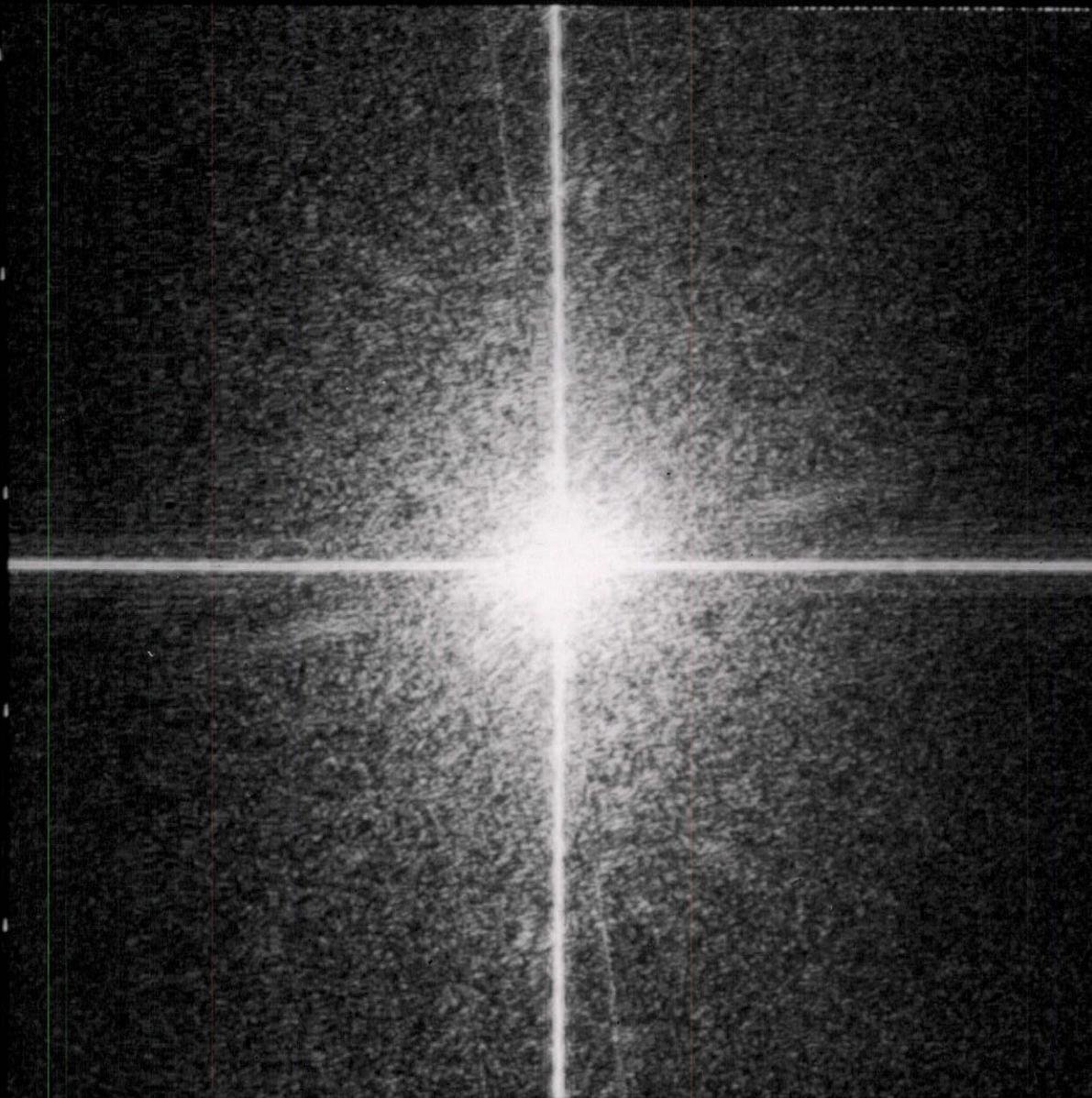
5. OBJECT EXTRACTION

The role of object extraction in image processing is to find a description of a picture in terms of appropriate picture subsets (objects) and to specify properties of these subsets. Specifying a subset of a picture is equivalent to specifying its characteristic function, i.e., the function whose value is 1 at the points of the subset and 0 elsewhere. Two types of processing, detection and articulation (segmentation), are commonly subsumed under object extraction. Detection is concerned with making decisions about the presence or absence of specific objects and estimating their position in the image. Articulation is concerned with the assignment of boundaries to objects in the picture.

5.1 OBJECT DETECTION

Detection refers to locating objects which are prespecified in terms of either local spatial distribution of grey values (template) or a transform of such a distribution (filter). Detection involves finding optimum matches between templates and images. The templates or filters can be selected intuitively, by random generation and systematic evaluation [36], by adaptive learning from sets of training images [37,38] or by design according to the principles of statistical communication theory [18,39]. The main mathematical tools for object detection are correlation, optimum frequency filtering and other transformations such as the Hadamard transform [40].

A picture or some transform is compared with elements from a library of representative features and objects or their appropriate transforms. For each feature a figure of merit is computed at each picture point which indicates the degree of match for that feature. These picture transformations can be thresholded and converted to a map locating instances of the features in the picture.

5.1.1 Cross Correlation

One measure of how well a portion of a picture matches a template can be defined as

$$M(m, n) = \sum_i \sum_j |g(i, j) - t(i - m, j - n)| \qquad (64)$$

where $g(i,j)$ is the digital picture, $t(i,j)$ is the template and i, j are such that $i - m$ and $j - n$ are in the domain of definition of the template. This definition

amounts to computing $M(m,n)$ for all template translations (m,n) and noting those translations for which $M(m,n)$ is small.

Another measure of similarity is based on the Euclidean distance between two vectors. This leads to the cross-correlation between two functions g and t, defined by

$$R(m,n) = \frac{\sum_i \sum_j g(i,j)\, t(i+m,j+n)}{\left(\sum_i \sum_j g^2(i,j)\right)^{\frac{1}{2}} \left(\sum_i \sum_j t^2(i+m,j+n)\right)^{\frac{1}{2}}} \qquad (65)$$

where the sum is taken over all i and j within the domain of the translated template. The picture and the template are declared similar when the cross-correlation is large, the peak height being a measure of degree of match. However, numerical results are sensitive to changes in orientation, mean grey value, contrast and noise making the detection of the peak difficult.

The optimal spatial mask for a feature is dependent not only on its spatial pattern, but also on its internal spatial correlations [41]. The simple correlation measure, however, ignores the spatial relationship of points within each image. A statistical correlation measure can be defined as

$$R_s(m,n) = \frac{\sum_i \sum_j p(i,j)\, t(i+m,j+n)}{\left(\sum_i \sum_j p^2(i,j)\right)^{\frac{1}{2}} \left(\sum_i \sum_j t^2(i+m,j+n)\right)^{\frac{1}{2}}} \qquad (66)$$

where $p(i,j)$ is obtained by convolving the sampled image $g(i,j)$ with the filter function $D(i,j)$. Thus,

$$p = g * D \qquad (67)$$

It can be shown [41], that for a noise free Markow process image (an image with a covariance matrix of the form $\left[\rho^{|i-j|}\right]$ where ρ is the correlation

49

coefficient between row elements of g), the optimal spatial filter is given by

$$
D = \begin{pmatrix} \rho^2 & -\rho(1 + \rho^2) & \rho^2 \\ -\rho(1 + \rho^2) & (1 + \rho^2)^2 & -\rho(1 + \rho^2) \\ \rho^2 & -\rho(1 + \rho^2) & \rho^2 \end{pmatrix} \tag{68}
$$

For a random pattern ($\rho = 0$) the statistical correlation measure reduces to the simple cross-correlation. For a highly correlated image ($\rho \approx 1$) the mask becomes

$$
D = \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix} \tag{69}
$$

This operator is the discrete approximation of the mixed fourth derivative $\partial^4/\partial x^2 \, \partial y^2$. The linear preprocessing of the image prior to the application of the basic correlation measure has been shown to provide a considerable improvement in the detectability of specific patterns or image misregistrations.

5.1.2 Matched Filtering

Various filtering operations have been described for image restoration and image enhancement. A filter that offers classification and correlation is the matched filter. The matched filter maximizes the signal-to-noise ratio in a detection process when the noise perturbation is considered additive [18]. If the pattern $f(x, y)$ is to be detected in the image

$$
g(x, y) = f(x, y) + n(x, y) \tag{70}
$$

50

then the matched filter is given by

$$H(u, v) = \frac{F^*(u, v)}{N(u, v)} \qquad (71)$$

where $F^*(u,v)$ is the conjugate of the Fourier transform of $f(x,y)$ and $N(u,v)$ is the power spectrum of the noise process $n(x,y)$. This concept can be generalized to gradient-matched filters which detect on the basis of edges rather than energy [42]. The matched filter is translation but not rotation invariant, the peak in the output plane determines the location of the object to be detected.

5.2 ARTICULATION OF IMAGES

Articulation is concerned with the optimal assignment of boundaries to objects. The nontriviality of this task is apparent from inspection of multilevel digital pictures (see Figure 7). The problem is to find a description which refers to various subsets (objects) of the picture and specifies properties of these subsets. An image processing system must be capable of singling out the appropriate picture subsets. This process is also called segmentation. There is no universal method of segmenting a picture. Many different types of subsets can be objects, depending on the type of description that is required [8].

5.2.1 Thresholding

The basic method used for singling out a subset of a picture is to obtain the subset by thresholding. Object extraction by thresholding can be used on original pictures or on transformed (preprocessed) pictures. Specifically, for any picture g and any picture operation T that transforms pictures into pictures the object can be taken as the set of points on which $t_1 \leq T(g) \leq t_2$, where t_1 and t_2 are real constants.

When thresholding is applied to the picture itself, the presumption is that regions of interest display different, fairly constant grey values. In many cases this may be a very natural approach; e.g., in earth observations, clouds, water and terrain will have distinguished grey values.

When thresholding is applied to a spatially differentiated picture (see Section 4.1) contouring occurs. This method can be used to single out subsets that cannot be characterized as having a prespecified range of grey levels, but that contrast with their surroundings and will yield outlines of such subsets.

START LINE = 160   START SAMPLE = 200   LINE SPACING   1   SAMPLE SPACING   1   NGRY = 13

FROM   0   TO   11   DISPLAYED   AS
FROM   12   TO   13   DISPLAYED   AS
FROM   14   TO   15   DISPLAYED   AS
FROM   16   TO   17   DISPLAYED   AS
FROM   18   TO   19   DISPLAYED   AS
FROM   20   TO   21   DISPLAYED   AS
FROM   22   TO   23   DISPLAYED   AS
FROM   24   TO   25   DISPLAYED   AS
FROM   26   TO   27   DISPLAYED   AS
FROM   28   TO   29   DISPLAYED   AS
FROM   30   TO   31   DISPLAYED   AS
FROM   32   TO   33   DISPLAYED   AS
GREATER   THAN   34   DISPLAYED   AS

Figure 7. Compressed ERTS-1 Image

52

Although thresholding is very effective on many types of images, it has short-comings: 1) false contouring and smoothing can occur, 2) there is no guarantee that the resultant contours are closed curves and 3) the results are very sensitive to the selection of thresholds. It is, however, often possible to select good thresholds by examining the frequency distribution of grey levels in g. Especially for multimodal histograms it is possible to segment the picture by choosing the thresholds between the mode peaks. This is extremely facilitated in an interactive image processing system like SMIPS, where it is possible to display the histogram on the screen and to select immediately the thresholds.

Figure 7 shows a part of the ERTS-1 MSS5 scene from the first color picture for which the original 128 value greyscale has been compressed to 13 levels. The result of thresholding this picture at the points indicated in the histogram in Figure 8 is shown in Figure 9. Simple thresholding has considerably enhanced this particular feature.

5.2.2 Region Analysis

Once a figure has been extracted from a picture, it becomes possible to perform operations in which the points of the figure are involved. Region analysis attempts to simplify a digital picture by partitioning it into a set of disjoint regions. Each region is composed of picture points having the same grey value and connected to each other. Region analysis involves the concepts of connectivity, geometry and shape [8,43] and will not be discussed in this report.

5.2.3 Contour Following

A recurring theme in image analysis is that a picture may be simplified by representing objects in the picture by their contours. The use of gradient operators for border detection has been discussed. Another technique is called contour following, it involves tracing out the boundary between a figure and its background.

Two conditions must be fulfilled if a contour following algorithm is to be successfully applied. First, since a picture generally has many grey levels, there must be some way of defining the figure whose contour is to be followed. In simple cases the figure can be extracted from the background by thresholding. This is a reduction to a binary picture. When this cannot be done, it is sometimes possible to combine a gradient operator with a contour follower [43]. The second prerequisite for successful contour following is that the figure has no spurious gaps in it. This problem can sometimes be overcome by first smoothing the picture in order to fill small gaps. Contour following is intrinsically a serial problem. An error made at any step makes it more likely that succeeding steps will also be in error. Therefore, the applicability of contour following seems restricted to pictures with low noise levels.

53

FREQUENCY DISTRIBUTION   SL= 100   SS= 100   NL= 490   NS= 409   HISTUP= 255   HISTLN=   0

AVERAGE GRAY LEVEL = 20.2933   STANDARD DEVIATION = 16.2653 NUMBER OF POINTS = 160000

FREQUENCY HISTOGRAM - NORMALIZED TO MAXIMUM = 100   (LEVELS WITH ZERO FREQUENCY OMITTED)

GRAY	FREQ	PERCENT	0	10	20	30	40	50	60	70	80	90	100

Figure 8. Histogram

START LINE = 160   START SAMPLE = 260   LINE SPACING   1   SAMPLE SPACING   1   NGRY = 7

```
          FROM    0    TO   12   DISPLAYED   AS
          FROM   13    TO   18   DISPLAYED   AS   .
          FROM   19    TO   19   DISPLAYED   AS   H
          FROM   20    TO   20   DISPLAYED   AS   O
          FROM   21    TO   21   DISPLAYED   AS   X
          FROM   22    TO   22   DISPLAYED   AS   A
          GREATER   THAN   22   DISPLAYED   AS   T
```

```
          2222222222222222222222 2222222222222222223333333333333333333333333333
          66666666677777777778888888889999999900000000011111111122222222
          01234567890123456789012345678901234567890123456789012345678901234567
```

```
160
161   ..............................................T................
162   ..............................................................
163   ..........................................+HADPXH...
164   ..........................................HHA................
165   ..............................................................
166   ..............................................................
167   ..............................................................
168   ..............................................................
169   ...............................H.......C+.....
170   ..............................................HHHMH..........
171   ...........................HOHOHXOOCCXXXO.............
172   ..........................YAAHH...............
173   .......................OTR....................
174   ......................IT....................
175   ....................XTO....................
176   ...................OO....................
177   .................CH....................
178   .................HX....................
179   ..............X....................
180   .............OA....................
181   ...H........AO....................
182   ...........MX....................
183   ..........HT....................
184   .........XO....................
185   ........X....................
186   .......A....................
187   .......OO....................
188   ......MX....................
189   ......O....................
190   .....oX....................
191   ....O....................
192   ...................
193   ...HX....................
194   ..XO....................
195   .M....................
196   ....................
197   ..O....................
198   ..O....................
199   .M....................
200   ...O....................
201   ...HA....................
202   ...H....................
203   ....................
204   ..O....................
205   .M....................
```

Figure 9. Feature Extracted from Figure 7

55

6. PICTORIAL PATTERN RECOGNITION

The huge masses of data in the form of pictures that are being continuously collected present a vital need for the automatic analysis of pictorial patterns. Presently the ability of machines to perform perceptual tasks is, however, very limited. The study of a more modest problem, the assignment of patterns to one or several prespecified classes (classification) has led to an abstract mathematical model that provides the theoretical basis for classifier design. This subject of mathematical pattern recognition is covered in [43,44] which contain many references to previous work. The purpose of this section is only to introduce the terminology and to formulate the problem of automatic pattern classification.

Mathematical pattern recognition is the study of mathematical techniques which support human experience in the classification of patterns. Mathematically, the problem is to find a function that maps a set of pictures into a set of classes. It is usually convenient to do this in three steps, preprocessing, feature extraction and classification. Figure 10 illustrates the concept.

REAL WORLD — PATTERN SPACE — FEATURE SPACE — CLASSIFICATION SPACE

SENSOR → PRE-PROCESSING → FEATURE SELECTION → CLASSIFICATION →
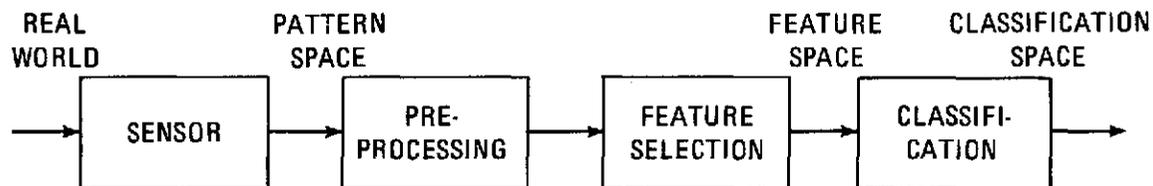
Figure 10

The physical world is sensed by some sensor system which describes a representation of that world by R scalar values where R is typically quite large. Some preprocessing is then performed on the raw pictures, e.g., radiometric and geometric corrections, noise cleaning, etc. Because R is large, it is desirable to reduce the dimensionality of the pattern space while still maintaining the discriminatory power for classification purposes which is inherent in the pictures. Therefore, in a feature space of dimensionality N (much smaller than R) classification rules can be computed in reasonable amounts of time. The classification space in which one of K classes can be selected is of dimensionality K.

It should be emphasized that in many cases image enhancement and classification by a human interpreter can solve the same problem at considerably lower cost. Automatic classification without feature selection and preprocessing is time

consuming and dangerous because noise may blur decision boundaries resulting in misclassification.

The essence of pattern recognition resides in the selection of few, good variables. Therefore, feature selection is probably the most important aspect of pattern recognition. Two distinct reasons for the need of feature selection should be observed. The pattern space is represented by the sensor data and sensors are often designed by other than classification considerations. Thus, it does not seem unreasonable to conjecture that there may be combinations of the dimensions of the pattern space affording meaningful classification power which otherwise would not be exploited. The second need for feature selection lies in the requirement for a space in which classification algorithms can be efficiently implemented. In the high dimensional pattern space even the simplest classification algorithms are quite time consuming on large scale digital computers. In addition, the selected features in the reduced space may cluster better than in the pattern space and will allow simpler decision rules. With improper feature extraction the classification algorithms will necessarily be less efficient and classification errors will increase.

In many cases features seem to have been selected because of their mathematical tractability or ease of implementation rather than because of their suitability for the given classification task. The soundest approach to feature selection is to use knowledge about the structure of the patterns and the definition of the classes as a guide in choosing the features and preprocessing operations. Feature selection is much more problem dependent than classification.

The problem of classification is basically one of separating the feature space into regions, one for each category. Classification has evolved to a higher state of refinement than feature selection. Two areas of classification can be distinguished. In the context of supervised learning prototypes are known as to their correct classification. The classification problem is finding separating surfaces which correctly classify the known prototypes and which afford some degree of confidence in correctly classifying unknown patterns.

In many data analysis applications such classification information is not available. Thus, the subject of nonsupervised learning attempts to apply recognition techniques to unclassified data. Results may be descriptions of the number of classes or clusters the data fall into. Once clusters are defined, the supervised learning techniques become meaningful. Clustering can be defined as the nonsupervised classification of objects which amounts to the process of generating classes without any knowledge of prototype classification. The essential characteristic is the sorting of the data into subsets such that each subset contains data points that are as much alike as possible.

A drawback of supervised classification techniques for multispectral data is associated with the high variability of the spectral signatures. Application of supervised techniques requires, therefore, obtaining the reference signatures from training areas which form part of, or are near, each particular survey area. Even with this practice, iterations and considerable human judgement are required to select proper training areas such that the classification is of acceptable accuracy [49].

Unsupervised classification groups multispectral data into a number of classes based on some intrinsic similarity within each class. It avoids reference signatures and the physical identification of each class is done after processing by checking a small area belonging to each class. Because of this reversed order with respect to supervised techniques, the investigator will know where to select the ground truth data based on the resulting classification map.

The sequential clustering program described in [49] is available at LMES.

REFERENCES

[1] McGlamery, B. L.: Digital Image Processing. Summary Report, Visibility Laboratory, Univ. of California, San Diego 1972.

[2] Goodman, J. W.: Introduction to Fourier Optics. New York: McGraw-Hill 1968.

[3] Falconer, D. G.: Image Enhancement and Film-Grain Noise. Optica Acta 17 (1970), 693-705.

[4] Stockham, T. G.: Image Processing in the Context of a Visual Model. Proc. IEEE 60 (1972), 828-842.

[5] Sawchuk, A. A.: Space-Variant Motion Degradation and Restoration. Proc. IEEE 60 (1972), 854-861.

[6] Robbins, G. M., Huang, T. S.: Inverse Filtering for Linear Shift-Variant Imaging Systems. Proc. IEEE 60 (1972), 862-872.

[7] Rosenfeld, A.: Picture Processing by Computer. Academic Press, New York-London 1969.

[8] Cooley, J. W., Tuckey, J. W.: An Algorithm for the Machine Calculation of Complex Fourier Series. Math. of Comp. 19 (1965), 297-301.

[9] Cooley, J. W., Lewis, P. W., Welch, P. D.: The Fast Fourier Transform and its Applications. IEEE Trans. Educ. 12 (1969), 27-34.

[10] Bracewell, R.: The Fourier Transform and its Applications. McGraw-Hill, New York 1965.

[11] Barnes, G. H., R. M. Brown, M. Kato, D. J. Kuck, D. L. Slotnik: The ILLIAC IV Computer. IEEE Trans. Computers C-17 (1968), 746-757.

[12] Kuck, D. J.: ILLIAC IV software and application programming. IEEE Trans. Computers C-17 (1968), 758-770.

[13] Pease, M. C.: An Adaptation of the Fast Fourier Transform for parallel processing. J. ACM 15 (1968), 252-264.

[14] Bergland, G. D.: Fast Fourier Transform Hardware Implementations — a Survey. IEEE Trans. Audio and Electroacoustics 17 (1969), 109-119.

[15] Johnston, Emily, G.: The PAX II Picture Processing System. In Lipkin, B. S., Rosenfeld, A., eds., Picture Processing and Psychopictorics. Academic Press, New York-London 1970.

[16] Courant, R., Hilbert, D.: Methods of Mathematical Physics. Interscience New York 1953.

[17] Andrews, H. C.: Computer Techniques in Image Processing. Academic Press, New York-London 1970.

[18] Davenport, W. B., Root, W. J.: Random Signals and Noise. McGraw-Hill, New York 1958.

[19] Harris, J. L.: Image Evaluation and Restoration. J. Opt. Soc. Amer. 56 (1966), 569-574.

[20] Roetling, P. G., Trabka, E. A., Kinzly, R. E.: Theoretical Prediction of Image Quality. J. Opt. Soc. Amer. 58 (1968), 342-346.

[21] Horner, J. L.: Optical Spatial Filtering with the Least-Mean-Square-Error Filter. J. Opt. Soc. Amer. 59 (1969), 553-558.

[22] Slepian, D.: Linear Least-Squares Filtering of Distorted Images. J. Opt. Soc. Amer. 57 (1967), 918-922.

[23] McAdam, D. P.: Digital Image Restoration by Constrained Deconvolution. J. Opt. Soc. Amer. 60 (1970), 1617-1627.

[24] Moik, J. G.: SMIPS-A Small Interactive Image Processing System-Users Manual. GSFC X-Document X-650-73-283.

[25] Nathan, R.: Picture Enhancement for the Moon, Mars and Man. In Cheng, G. C., et al, eds., Pictorial Pattern Recognition, Washington, Thompson, 1968, 239-266.

[26] Twomey, S.: On the numerical solution of Fredholm integral equations of the first kind by the inversion of the Linear system produced by quadrature. J. Assoc. Comp. Mach. 10 (1963), 97-101.

[27] Smith, H. A.: Improvement of the resolution of a Linear scanning device. SIAM J. Appl. Math. 14 (1966), 23-40.

[28] Prewitt, Judith, M. S.: Object Enhancement and Extraction. In Lipkin, E. S. and Rosenfeld, A., eds., Picture Processing and Psychopictorics. Academic Press, New York-London 1970, 75-149.

[29] Freeman, H.: On the Encoding of Arbitrary Geomatric Configurations. IEEE Trans. EC-10 (1961), 260-268.

[30] Montanari, U.: A Note on Minimal Length Polygonal Approximation to a Digitized Contour. Comm. ACM 13 (1970), 41-47.

[31] Narasimhan, R.: Labeling Schemata and Syntactic Descriptions of Pictures. Information and Control 7 (1964), 151-179.

[32] Rosenfeld, A., Pfaltz, J. L.: Sequential Operations in Digital Picture Processing. J. ACM 13 (1966), 471-494.

[33] Andrews, H. C., Tescher, A. G., Kruger, R. P.: Image Processing by Digital Computer. IEEE Spectrum 9 (1972), 20-32.

[34] Cooley, J. W., Lewis, A. W., Welch, P. D.: Application of the Fast Fourier Transform to Computation of Fourier Integrals, Fourier Series and Convolution Integrals. IEEE Trans. Audio and Electroacoust. 15 (1967).

[35] Hunt, B. R.: A Matrix Theory Proof of the Discrete Convolution Theorem. IEEE Trans. Audio and Electroacoust. 19 (1971), 285-288.

[36] Uhr, L., Vossler, C.: A Pattern Recognition Program that Generates, Evaluates and Adjusts its own Operators. Proc. WJCC (1961), 555-569.

[37] Block, H. D., Nilson, N. J., Duda, R. O.: Determination and Detection of Features in Patterns. In Tou, J. T., Wilcox, R. H., eds., Computer and Information Sciences, New York, Spartan, 1964, 75-110.

[38] Kazmierczak, H., Steinbuch, K.: Adaptive Systems in Pattern recognition. IEEE Trans. EC 12 (1963), 822-835.

[39] Montgomery, W. D., Broome, P. W.: Spatial Filtering. J. Opt. Soc. Amer. 52 (1962), 1259-1275.

[40] Deutsch, E. S., Wann, Z. F.: On the Use of the Hadamard Transform in Digital Image Matching. Univ. of Maryland Techn. Rep. TR-216, 1973.

[41] Arcese, A., Menger, P. H., Trombini, E. W.: Image Detection through Bipolar Correlation. IEEE Trans. IT-16 (19   ).

[42] Andrews, H. C.: Automatic Interpretation and Classification of Images by Use of the Fourier Domain. In Grasselli, A., ed., Automatic Interpretation and Classification of Images. New York, Academic Press 1968, 187-198.

[43] Duda, R. O., Hart, P. E.: Pattern Classification and Scene Analysis. Wiley-Interscience, New York, London, 1973.

[44] Andrews, H. C.: Introduction to Mathematical Techniques in Pattern Recognition. Wiley-Interscience, New York, London, 1972.

[45] Allison, L. J., Cherrix, G. T.: The Application of Color Display Techniques for the Analysis of Nimbus Infrared Radiation Data. NASA Technical Note D-6786, 1972.

[46] Vander Lugt, A.: A review of optical data-processing techniques. Optica Acta 15 (1968), 1-33.

[47] Hawkins, J. K., Munsey, C. J.: Image Processing by Electron-Optical Techniques. J. Opt. Soc. Amer. 57 (1967), 914-918.

[48] Sheppard, J. J., Stratton, R. H., Gazley, C. G.: Pseudo-color as a Means of Image Enhancement. Am. J. Optom., 46 (1969), 735-754.

[49] Su, M. Y.: The Composite Sequential Clustering Technique for Analysis of Multispectral Scanner Data. NASA Contractor Report CR-128999, 1972.

[50] Rindfleisch, T. C., et al.: Digital Processing of the Mariner 6 and 7 Pictures. J. of Geophysical Research 76 (1971), 394-417.