Open access • Proceedings Article • DOI:10.1117/12.976236

Restoration for sampled imaging systems. — Source link

Lynnette Wood, Robert A. Schowengerdt, Dave Meyer

Institutions: University of Arizona, United States Geological Survey

Published on: 10 Dec 1986

Topics: Image restoration, Wiener filter, Point spread function, Image processing and Transfer function

RESTORATION FOR SAMPLED IMAGING SYSTEMS.

INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.

- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.

- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17"x 23" black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6"x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.

Wood, Lynnette

RESTORATION FOR SAMPLED IMAGING SYSTEMS

The University of Arizona PH.D. 1986

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs or pages ✓
2. Colored illustrations, paper or print _____
3. Photographs with dark background ✓
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages ✓
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received ✓
16. Other _____

_____

_____

RESTORATION FOR SAMPLED IMAGING SYSTEMS

by

Lynnette Wood

A Dissertation Submitted to the Faculty of the

COMMITTEE ON APPLIED MATHEMATICS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 8 6

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Lynnette Wood

entitled Restoration for Sampled Imaging Systems

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Arthur L. Wright	11/7/86 Date
Robert Schowengerdt	11/7/86 Date
Clark J. Benson	11/7/86 Date
	Date
	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Robert Schowengerdt
Dissertation Director

11/18/86
Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Signed: Lynnette Wood

This dissertation is dedicated to my parents, Wallace and Charlotte Wood, whose unceasing love and support through the years made it possible.

ACKNOWLEDGMENTS

This research was supported by the U. S. Geological Survey, EROS Data Center. I also wish to acknowledge my dissertation advisor Dr. Robert Schowengerdt, whose insight and direction were invaluable; Dr. Steve Park of the College of William and Mary, who contributed useful comments and suggestions on the mathematics and theory of the work; and Dave Meyer of the EROS Data Center who contributed to testing and evaluating software written as a result of this work.

TABLE OF CONTENTS

TABLE OF CONTENTS—Continued

TABLE OF CONTENTS—Continued

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS—Continued

LIST OF TABLES

ABSTRACT

Digital image restoration requires some knowledge of the degradation phenomena in order to attempt an inversion of that degradation. Typically, degradations which are included in the restoration process are those resulting from the optics and electronics of the imaging device. Occasionally, blurring caused by an intervening atmosphere, uniform motion or defocused optics is also included. Recently it has been shown that sampling, the conversion of the continuous output of an imaging system to a discrete array, further degrades or blurs the image. Thus, incorporating sampling effects into the restoration should improve the quality of the restored image.

The system transfer function (the Fourier transform of the point spread function), was derived for the Landsat Multi-Spectral Scanner and Thematic Mapper systems. Sampling effects were included, along with the relevant optical, instantaneous field of view and electronic filter data, in the system analysis.

Using the system transfer function, a least squares (Wiener) filter was then derived. A Wiener filter requires the ratio of the power spectra of the scene and noise, which is often, for simplicity, assumed to be a constant over frequency. The restoration method used here includes models for the power spectra which are based on the study of several different types of Landsat scenes. The Wiener filter is then inverse Fourier transformed to find a restoration filter which is spatially windowed to suppress ringing. Qualitative and quantitative evaluations are made of the restored imagery.

Comparisons are made to the approaches taken by other investigators, in particular, to one who has had success restoring the same type of imagery. it is found that the restoration method used here compares favorably with this previous work.

CHAPTER 1

INTRODUCTION

Image restoration is an example of what is known as an inversion problem, where some physical quantity is estimated from a set of measurements. In particular, image restoration is an attempt to recover losses suffered in an imaging process due to various degradations. These degradations may be the result of the intervening atmosphere, imperfect optics or electronic filters, motion blur, discretization of the continuous scene (sampling) and ground processing. Many of the imaging parameters can be determined or estimated from engineering design data and preliminary testing. Atmospheric effects may be modeled using a Monte Carlo approach. Motion blur is accounted for through a knowledge of the relative motion between object and imaging system, and ground processing is usually well documented. Sampling, the conversion of the continuous output of an imaging system to a discrete array, has not been an explicit component in previous restoration work. Sampling further degrades, or blurs, an image. Thus, incorporating sampling effects into the restoration should improve the quality of the restored image. One approach which includes sampling degradation in the restoration process is presented here.

Imaging can be modeled by a convolution of the function describing the imaging process with one describing the scene. The convolution of two functions is equivalent to the product of the Fourier transforms of the functions. This simplification leads to the concept of a restoration "filter", a third function which multiplies the existing product in an attempt to recover the function describing the original scene. The filter chosen for this work is the Wiener filter, which has been very successful in one-dimensional signal processing.

The Wiener filter, however, is derived for continuous functions. In particular, the Wiener filter requires certain parameters including an imaging system performance function called the "transfer function". One way of dealing with sampling is a stochastic approach in which sampling effects are accounted for by assuming a random distribution of point sources and averaging the overall system output over all possible point source locations. This results in an average system transfer function which can then be used in the Wiener filter. Since this average system transfer function is not continuous, the resulting Wiener filter is suboptimal in the least squares sense. However, it will be seen that using such a filter in image restoration leads to quite acceptable results, often even subjectively better results than those given by the optimal filter.

This suboptimal Wiener filter is then inverse Fourier transformed to create a spatial filter. it is required that this spatial filter be small enough to be efficiently applied to very large images. In particular, the spatial filter is applied to Landsat satellite images (typically consisting of 10^7 picture elements, or pixels) as part of a geometric correction procedure. The resulting images are sharper than images obtained using other methods.

The theory and development of this approach to image restoration is described in this dissertation. Chapter 2 discusses in detail a mathematical model for imaging and discusses various restoration procedures including the Wiener filter. Chapter 3 describes the Landsat imaging sensors, and Chapter 4 the geometric correction of these Landsat images and the use of the restoration filter in this context. Chapter 5 discusses sampling and its inclusion in the Wiener filter. Chapter 6 describes another approach which has also been applied successfully to Landsat images and how that approach relates to the Wiener filter approach. Chapter 7 covers the determination of the parameters necessary for

the Wiener filter and other implementation issues, and Chapter 8 shows some qualitative and quantitative results. Finally, Chapter 9 describes some topics for possible future study.

CHAPTER 2

IMAGE RESTORATION

An image is a representation of an object or scene which results from the recording of radiant energy emitted or reflected from the object (Andrews and Hunt, 1977). The object and image are often represented as continuous functions $f(x', y')$ and $g(x, y)$ which describe the distribution of light intensity at each point (x', y') and (x, y) in the object and image planes. A function h is also defined which describes the imaging process.

There are three assumptions on which the mathematics of noncoherent image formation is based: Image intensity is nonnegative, it is a neighborhood process and it obeys the superposition principle.

Since images are formed by the recording of radiant energy, it is reasonable to assume that the intensity distributions which are the object and its image are nonnegative, that is

$$f(x', y') \geq 0 \quad \text{and} \quad g(x, y) \geq 0 \tag{2.1}$$

so that nonnegativity is a property of noncoherent imaging.

A function h which describes the imaging system performance is the "point spread function". The point spread function describes the irradiance distribution at the image plane of an object which is an ideal point source. The point spread function of an ideal imaging system is a Dirac distribution (Arsac, 1966) or "delta function" (Greenberg, 1978) $\delta(x, y)$, which is defined in the theory of distributions in terms of the integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x - a, y - b) f(x, y) dx dy = f(a, b) \tag{2.2}$$

The ideal image of a point source would also be a point. However, realizable imaging systems introduce some blurring (or low pass filtering) so the intensity at a point in the image is a function of contributions from a neighborhood of points in the scene. For example, even the best optical system would be "diffraction limited", preventing the image of a point source from being an ideal point image (Gaskill, 1978).

Finally, image formation is usually assumed to obey the superposition principle under addition. The overall image intensity distribution can be thought of as a sum (or integral) of point sources. Summing all the point sources which make up the continuous image, the most general image formation equation is

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y,x',y',f(x',y'))dx'dy' \tag{2.3}$$

where x, y represents coordinates in the image plane and x', y' coordinates in the object plane. The point spread function h describes the transformation of energy from object to image plane.

A simplifying assumption is that the function h simply weights the object distribution. This is the case of a linear imaging system, and Eq. (2.3) reduces to

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y,x',y')f(x',y')dx'dy' \tag{2.4}$$

This is a linear integral equation, in particular, a Fredholm integral equation of the first kind. It is a Fredholm equation because the limits of integration are fixed and it is of the first kind because $f(x',y')$ appears only in the integrand (Lovitt, 1924; Greenberg, 1978).

Another simplifying assumption is that of shift invariance, that the image of a point source depends only on the intensity of the point and not on its location in image and object planes. In this case the image formation equation simplifies further to

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x-x',y-y')f(x',y')dx'dy' \tag{2.5}$$

which is the familiar convolution and will hereafter be denoted

$$g(x,y) = h(x,y) * f(x,y) \tag{2.6}$$

If the input signal consists of a single ideal point source then Eq. (2.5) becomes (with $\delta(x,y) = \delta(-x,-y)$)

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x-x',y-y')\delta(x',y')dx'dy' = h(x,y) \tag{2.7}$$

These equations are fundamental to linear system theory since they indicate that if the response to a impulse (delta function) i.e. the point spread function, is known, then the response to any input may be calculated by means of Eq. (2.6).

The imaging systems considered herein can be assumed to be linear and shift invariant with little error, so that much of the theory can be developed using the model of Eq. (2.6). Sampling, the conversion of the continuous output of an imaging system to a discrete array, violates the shift invariance assumption, and will be dealt with separately.

Image Restoration

When a scene is recorded by an imaging device certain degradations are inevitable. These degradations may be the result of the intervening atmosphere, imperfect optics or electronic filters, sensor noise, discretization of the continuous scene and ground processing. Image restoration is an attempt to recover losses suffered in the imaging process due to some or all of these various degradations. If we let the original scene be represented by $f(x',y')$, then the process of recording a scene, exclusive of sampling (Park, et. al., 1984) and noise, for linear shift invariant systems, can be described mathematically by Eq. (2.6). However, all physically realizable imaging systems introduce some noise, i.e. some unwanted signal. Some system noise may be periodic and easily dealt with. But there will usually be other noise which can only be characterized

statistically. If that noise $n(x,y)$ is additive and signal independent, the imaging process can be written

$$g(x,y) = h(x,y) * f(x,y) + n(x,y) \tag{2.8}$$

The scene $f(x',y')$ can no longer be perfectly restored since $n(x,y)$ is usually not deterministic.

Even without noise, image restoration is ill-conditioned (Andrews and Hunt, 1977; Phillips, 1962) and often even singular. In particular, if the object $f(x',y')$ in Eq. (2.4) consists of the sum of two functions $f_1(x',y')$ and $f_2(x',y')$ is orthogonal to the point spread function h over an interval $[a,b]$ and $f_1(x',y')$ is not, then

$$\int_a^b \int_a^b h(x,y,x',y')[f_1(x',y') + f_2(x',y')]dx'dy' = \int_a^b \int_a^b h(x,y,x',y')f_1(x',y')dx'dy'$$

The existence of the orthogonal component $f_2(x',y')$ cannot be determined from the image, and the correspondence between object and image is many-to-one.

Image restoration now becomes an attempt to find the "best" estimate $\hat{f}(x',y')$ of $f(x',y')$ in some sense. This estimate is referred to as the "restored image". The difference of one restoration method from another is often determined by the way in which "best estimate" is defined. Thus, there are innumerable processing approaches which can produce a "restored" image, and there are many ways one could categorize these methods such as constrained vs. unconstrained, a priori vs. a posteriori, linear vs. nonlinear and recursive vs. nonrecursive. Image restoration methods might also be categorized according to fields of mathematics: linear algebra approaches, probabilistic approaches, information theory approaches and transform approaches. A representative sample from each of these latter categories will be given here, and comparisons made.

This will by no means be an exhaustive list, but only meant to give a flavor of what has been done. Some of these approaches fair better than others in the presence of noise, and implementation issues are usually also of concern.

Linear Algebra Method

Since in much image restoration work the concern is with digital image restoration, we must consider sampled imagery. That is, the continuous scene $f(x',y')$ is imaged as $g(x,y)$ which is then sampled as $g_s(m\Delta x, n\Delta y)$ where

$$g_s(m\Delta x, n\Delta y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} g(x,y)\delta(x - m\Delta x, y - n\Delta y) \qquad (2.9)$$

for a uniform sampling grid (Andrews and Hunt, 1977). This suggests modeling the entire imaging process as a discrete process and implementing techniques of linear algebra for image restoration. The discrete counterpart of continuous convolution, Eq. (2.5), is

$$g(m,n) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(m-k, n-l) f(k,l) \qquad (2.10)$$

Or, for scenes and point spread functions with finite extent $[r,R]$,

$$g(m,n) = \sum_{k=max(r_k, m-R_p)}^{min(R_k, m-r_p)} \sum_{l=max(r_l, n-R_q)}^{min(R_l, n-r_q)} h(m-k, n-l) f(k,l) \qquad (2.11)$$

where $r_k \le k \le R_k$, $r_l \le l \le R_l$, $r_p \le p \le R_p$, $r_q \le q \le R_q$ and $m = R_k + R_p - 1$, $n = R_l + R_q - 1$. Eq. (2.11) can be written in matrix notation as

$$\mathbf{g} = \mathbf{H}\mathbf{f} \qquad (2.12)$$

where

$$\mathbf{g} = \begin{pmatrix} g^{(r_l+r_q)} \\ g^{(r_l+r_q+1)} \\ \vdots \\ g^{(R_l+R_q-1)} \\ g^{(R_l+R_Q)} \end{pmatrix} \quad \text{where} \quad g^{(\alpha)} = \begin{pmatrix} g(r_k+r_p, \alpha) \\ g(r_k+r_p+1, \alpha) \\ \vdots \\ g(R_k+R_p-1, \alpha) \\ g(R_k+r_p, \alpha) \end{pmatrix}$$

$$\mathbf{f} = \begin{pmatrix} f^{(r_l)} \\ f^{(r_l+1)} \\ \vdots \\ f^{(R_l-1)} \\ f^{(R_l)} \end{pmatrix} \quad \text{where} \quad f^{(\beta)} = \begin{pmatrix} f(r_k,\beta) \\ f(r_k+1,\beta) \\ \vdots \\ f(R_k-1,\beta) \\ f(R_k,\beta) \end{pmatrix}$$

and

$$\mathbf{H} = \begin{pmatrix} H^{(r_q)} & 0 & 0 & \cdots & 0 \\ H^{(r_q+1)} & H^{(r_q)} & 0 & \ddots & \vdots \\ \vdots & H^{(r_q+1)} & H^{(r_q)} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ H^{(R_q-1)} & \vdots & \ddots & \ddots & 0 \\ H^{(R_q)} & H^{(R_q-1)} & \ddots & \ddots & H^{(r_q)} \\ 0 & H^{(R_q)} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & H^{(R_q)} \end{pmatrix}$$

where

$$H^{(\gamma)} = \begin{pmatrix} h(r_p,\gamma) & 0 & 0 & \cdots & 0 \\ h(r_p+1,\gamma) & h(r_p,\gamma) & 0 & \ddots & \vdots \\ \vdots & h(r_p+1,\gamma) & h(r_p,\gamma) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ h(R_p-1,\gamma) & \vdots & \ddots & \ddots & h(r_p,\gamma) \\ h(R_p,\gamma) & h(R_p-1,\gamma) & \ddots & \ddots & \vdots \\ 0 & h(r_p,\gamma) & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & h(R_p,\gamma) \end{pmatrix}$$

for square matrices. Rectangular matrices can be padded with zero elements to make them square. This matrix formulation suggests regarding the discrete imaging equation, Eq. (2.11), as a system of linear equations where it is desired to solve for the (image) vector $\mathbf{f}$. Naively, Eq. (2.12) may be solved for $\mathbf{f}$ simply by finding the inverse of $\mathbf{H}$, and multiplying

$$\hat{\mathbf{f}} = \mathbf{f} = \mathbf{H}^{-1}\mathbf{g} \tag{2.13}$$

But there is no guarantee that $\mathbf{H}^{-1}$ exists. Even if $\mathbf{H}^{-1}$ does exist, the huge size of $\mathbf{H}$ (on the order 10^5x10^5 for a 512-by-512 image) seems to make solution of Eq. (2.13) prohibitive. However, point spread function matrices are sparse and often highly structured (Frieden, 1979; Pratt, 1975), so that efficient computational techniques for the solution of Eq. (2.13) are possible. For example, separable imaging systems, (systems for which the two-dimensional point spread function is a product of two one-dimensional line spread functions), which are also shift invariant, have symmetric point spread function matrices. Efficient methods exist (Stewart, 1973; Duff, 1981; Tewarson, 1973; Reid, 1971) for solving such systems, so that matrix inversion is not required. Those components of $\mathbf{f}$ which are in the null space of $\mathbf{H}$, however, can never be recovered.

Aside from this many-to-one nature of imaging, the problem with the discrete model of Eq. (2.12) is not, therefore, finding a feasible method of solution. The problem is that for physically realizable imaging systems the model itself is not realistic since it does not account for noise. A more realistic discrete imaging equation is

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n} \tag{2.14}$$

where $\mathbf{n}$ represents additive signal independent noise, an indeterministic quantity. The sparsity of $\mathbf{H}$ makes the inversion of Eq. (2.14) very unstable in the presence of noise, as $\mathbf{H}$ is usually ill-conditioned (Hunt, 1972; Phillips, 1962). Since for many imaging systems of interest (including those studied here) $\mathbf{H}$ is filled mostly with zeroes and small elements near the diagonal (Frieden, 1979), $\mathbf{H}^{-1}$ will have very large elements that will tend to enhance noise:

$$\hat{\mathbf{f}} = \mathbf{H}^{-1}\mathbf{g} - \mathbf{H}^{-1}\mathbf{n} \tag{2.15}$$

Many methods have been proposed to deal with this inherent instability, including smoothing methods (e. g. Phillips, 1962), iterative pseudo-inverse methods

(Maeda and Kazumi, 1984), iterative damped least squares methods (Maeda, 1985) and singular value decomposition methods (Huang and Narendra, 1975; Sasaki and Yamagami, 1985). Major disadvantages remain even with these methods. The pseudo-inverse method, for example, only works when the imaging system is separable. Iterative damped least squares methods work on nonseparable systems, but require an arbitrary parameter. Phillips' smoothing method has been successfully applied to image restoration (Hunt, 1973), but it also requires an arbitrary parameter. Other problems arise for those methods which do not constrain the solution $\mathbf{f}$: although it is known that $\mathbf{f}$ is nonnegative, $\hat{\mathbf{f}}$ may not be so. Methods which incorporate some constraints often give "better" results.

Information Theory Method

The maximum entropy method developed by Frieden (1972) includes such contraints. This method assumes that there are a fixed number of (positive energy) photons which must be distributed in the image (which is assumed to be subdivided into elemental "cells") in such a way as to be consistent with the image formation equation Eq. (1.5). The goal is to maximize the entropy

$$H = -\sum_{n=1}^{N} p_n \ln p_n \tag{2.16}$$

subject to the constraint Eq. (2.5), where the p_n represent the probabilities of infinitesimal units of intensity falling in the nth cell of the image. This method has been successfully applied to images of sources which can be described by Poisson statistics, such as astronomical images (Frieden and Wells, 1978), but have limited usefulness for other types of images.

Probabilistic Method

Other image restoration methods include viewing the whole imaging process as a stochastic problem, with the scene $\mathbf{f}$ and image $\mathbf{g}$ random vectors

which are samples from an ensemble of possible scenes and images. In this method, one wants to estimate the (discrete) restored image given the recorded image and some (statistical) knowledge about the noise; that is, to estimate $\mathbf{f}$ conditioned on knowledge of $\mathbf{g}$. Bayes' rule gives the a posteriori conditional density

$$P(\mathbf{f}|\mathbf{g}) = \frac{P(\mathbf{g}|\mathbf{f})P(\mathbf{f})}{P(\mathbf{g})} \tag{2.17}$$

Different criteria can be used in estimating $\mathbf{f}$: minimum mean square error estimates are the mean of the posterior density $P(\mathbf{f}|\mathbf{g})$, maximum a posteriori estimates are the mode of the posterior density, and maximum likelihood estimates are a special case of maximum a posteriori estimates where the posterior density is equal to the prior density $P(\mathbf{g}|\mathbf{f})$ (Hunt, 1977). Methods based on these criteria often give good results, but at the cost of computational efficiency as they often require (relatively) complex optimization algorithms (Hunt, 1977; Richardson, 1972).

Filtering Method

Finally, a method which has become very popular in image processing due to its success in one-dimensional signal processing (Cappellini, Constantinides and Emiliani, 1978) is the "filtering" method discussed below. This is the method used in this dissertation.

For linear shift invariant systems the imaging equation is just the convolution Eq. (2.8). Taking the Fourier transform of both sides of Eq. (2.8) results in considerable simplification since convolution in the spatial domain becomes simply multiplication in the frequency domain (Rosenfeld and Kak, 1982; Champeney, 1973). The frequency domain equivalent of the imaging equation is

$$G(u,v) = H(u,v)F(u,v) + N(u,v) \tag{2.18}$$

where G, H, F and N are the Fourier transforms of g, h, f and n, respectively, and u and v are the frequency domain coordinates corresponding to the spatial coordinates x and y. The Fourier transform and its inverse are defined as

$$F(u,v) = \mathcal{F}[f(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-2\pi i(xu+yv)}dxdy \tag{2.19}$$

and

$$f(x,y) = \mathcal{F}^{-1}[F(u,v)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{2\pi i(xu+yv)}dudv \tag{2.20}$$

$H(u,v)$ is called the "transfer function" and describes how well the imaging system "transfers" amplitude and phase information at different spatial frequencies. The equivalence of the convolution of two functions to the product of their fourier transforms, often called the "Convolution Theorem", is a very important relation which will be used repeatedly throughout this work.

Filtering methods consist of multiplication of the image spectrum $G(u,v)$ by a "filter function" $M(u,v)$. The product is then inverse Fourier transformed to yield the restored image. The filtering could be performed in the spatial domain as well. The filter function is inverse transformed and the resulting "deconvolution filter" is convolved with the image to create a restored image. Since the convolution of two functions in one domain is equivalent to multiplication of the Fourier transforms of the functions in the other domain, these two procedures are (theoretically) equivalent.

For example, if the transfer function $H(u,v)$ were known, for example from system modeling, and the system were noiseless, then one could, in theory, recover $f(x,y)$ exactly by dividing Eq. (2.18) (with $N(u,v) = 0$) by $H(u,v)$ and then taking the inverse Fourier transform, resulting in

$$f(x,y) = \mathcal{F}^{-1}[G(u,v)/H(u,v)] \tag{2.21}$$

In this example the term $1/H(u,v)$ is the filter function and is called an "inverse filter". This assumes, of course, that $H(u,v)$ is never zero. If it is the above

expression is undefined. If it is nonzero but very small, using this inverse filter would result in numerical instability. (This is just the Fourier domain equivalent of the ill-conditioned nature of imaging previously discussed.) In practice, this restoration method is not usually effective anyway because it assumes the imaging system is noiseless, which it rarely is. For additive signal independent noise, taking the Fourier transform of both sides of Eq. (2.8) and dividing by $H(u,v)$ results in the restoration

$$F(u,v) = \frac{G(u,v)}{H(u,v)} - \frac{N(u,v)}{H(u,v)} \tag{2.22}$$

Now if $|N(u,v)/H(u,v)|$ is very large relative to $|G(u,v)/H(u,v)|$, then, even in the case of a well behaved transfer function, the result will be considerable noise enhancement. Practical restoration methods must account for the inevitable noise present in the image.

One way of dealing with the noise amplification effects of the inverse filter is to reduce the value of the filter at high spatial frequencies, because it is at these frequencies that the signal-to-noise is usually least and, therefore, the noise amplification most severe. It would be preferable if the control of noise amplification were automatic. Such automatic control exists in the Wiener filter (Helstrom, 1967).

Wiener Filter

In deriving the Wiener filter, the concept of a random field is useful. A random field is a generalization of a stochastic process, which is, in turn, a generalization of a random variable. Whereas a random variable is a rule for assigning to every outcome w_i of an experiment whose set of possible outcomes is $W = \{w_1, w_2, ..., w_n\}$ a number $r(w_i)$, a random field is a rule for assigning to every outcome w_i a function of m variables $f(w_i, x_1, x_2, ..., x_m)$. If $m = 1$, Euclidean 1-space, then $f(w_i, x)$ is a stochastic process. For purposes of image

processing, it is useful to consider $m = 2$. Then, for example, if the underlying experiment is the selection of an image from a collection of images, the outcome w_i corresponds to the selection of the ith image. For a fixed w_i, $f(w_i, x, y)$ is a two-dimensional function representing the intensity distribution which defines the image.

The motivation for the Wiener filter is to try to minimize the mean square error between the original scene and the restored image (estimated scene):

$$\text{minimize} \quad e^2 \tag{2.23}$$

where

$$e^2 = E[(f(x,y) - \hat{f}(x,y))^2] \tag{2.24}$$

where both object and image coordinate systems are referenced to a common orthogonal coordinate system (x, y).

As previously mentioned, filtering by multiplying a filter function by the Fourier transform of the image is equivalent to convolving the inverse Fourier transform of the filter with the image itself. The Wiener filter approach is a linear least squares estimate: minimize the expectation in Eq. (2.24) subject to the constraint that the restoration $\hat{f}(x,y)$ is a linear function of the intensity levels in the image $g(x,y)$, i.e.

$$\hat{f}(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m(x,y,x',y') g(x',y') dx' dy' \tag{2.25}$$

where $m(x,y,x',y')$ is the deconvolution filter. For stationary random fields f and g Eq. (2.25) becomes a convolution (Papoulis, 1965)

$$\hat{f}(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m(x-x', y-y') g(x',y') dx' dy' \tag{2.26}$$

Substituting into Eqs. (2.23) and (2.24) yields

$$\text{minimize} \quad e^2 \tag{2.27}$$

where

$$e^2 = E[(f(x,y) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(x-x', y-y')g(x',y')dx'dy')^2] \qquad (2.28)$$

It can be shown (Theorem 1, Appendix A) that a function $m(x,y)$ minimizes Eq. (2.28) if and only if $m(x,y)$ satisfies

$$E[(f(x,y) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(x-x', y-y')g(x',y')dx'dy')g(x_0,y_0)] = 0 \qquad (2.29)$$

Eq. (2.29) can also be written

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(x-x', y-y')E[g(x',y')g(x_0,y_0)]dx'dy' = E[f(x,y)g(x_0,y_0)] \qquad (2.30)$$

Or, making use of the definitions of auto- and cross-correlation (Papoulis, 1984)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(x-x', y-y')R_{gg}(x',y',x_0,y_0)dx'dy' = R_{fg}(x',y',x_0,y_0) \qquad (2.31)$$

where R_{gg} and R_{fg} are the image autocorrelation and the image-scene cross-correlation, respectively. Since the random fields have been assumed to be stationary Eq. (2.31) reduces to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(x-x', y-y')R_{gg}(x'-x_0,y'-y_0)dx'dy' = R_{fg}(x'-x_0,y'-y_0) \qquad (2.32)$$

which is just a convolution. Taking the Fourier transform of both sides yields

$$M(u,v)S_{gg}(u,v) = S_{fg}(u,v) \qquad (2.33)$$

where S_{gg} and S_{fg} are the image power spectrum and the scene-image cross power spectrum, respectively. For stationary random fields and assuming the noise and scene are uncorrelated with zero mean noise, then

$$S_{fg}(u,v) = H^*(u,v)S_{ff}(u,v) \qquad (2.34)$$

and

$$S_{gg}(u,v) = S_{ff}(u,v)|H(u,v)|^2 + S_{nn}(u,v) \tag{2.35}$$

Substituting Eqs. (2.34) and (2.35) into Eq. (2.33) and solving for $M(u,v)$ yields the least squares or Wiener filter

$$M(u,v) = \frac{1}{H(u,v)} \frac{|H(u,v)|^2}{|H(u,v)|^2 + S_n(u,v)/S_f(u,v)} \tag{2.36}$$

where $S_n(u,v) = S_{nn}(u,v)$ and $S_f(u,v) = S_{ff}(u,v)$.

These assumptions about the stationarity of the scene and noise are strong and require comment. Stationarity of the scene is not usually the case, however, as will be discussed in Chapter 7. However, for the imaging systems studied here, the noise is so small that the question of stationarity is almost academic. As the noise $S_n(u,v)$ approaches zero, the Wiener filter approaches an inverse filter, which makes no assumptions about either the scene or the noise statistics.

The Wiener filter may be multiplied by the Fourier transform of the image and the product inverse Fourier transformed to yield the restored image, or the filter may be inverse transformed and the image convolved by the resulting deconvolution filter $m(x,y)$ to yield the restored image.

It is easy to show that the modulus and phase of the Wiener filter are

$$|M(u,v)| = \frac{1}{|H(u,v)|} \left[\frac{|H(u,v)|^2}{|H(u,v)|^2 + S_n(u,v)/S_f(u,v)} \right] \tag{2.37}$$

and

$$\arg M(u,v) = -\arg H(u,v) \tag{2.38}$$

The Wiener filter is an inverse phase filter and a modified inverse modulus filter.

The Wiener filter "controls" the noise enchancement of the inverse filter of Eq. (2.22), since even if $H(u,v)$ is very small or even zero for some values of u and v, the denominator will never be smaller than $S_n(u,v)/S_x(u,v)$.

Comments

It must be pointed out that there is no single "correct" or optimum method of image restoration. It has been suggested (Frieden, 1979), in fact, that the "best" restoration method depends on the type of object originally present (smooth or textured, edge-type or impulse-type), the purpose of the restoration (edge detection, pattern recognition, spectral classification), the amount and kind of a priori information available (scene or noise statistics) and strength of the noise present. In particular, due to the many-to-one nature of imaging, the more a priori information incorporated into the restoration, the fewer degrees of freedom in the "search" toward a restored image. Other considerations are purely pragmatic, involving, for example, availability of computing resources. It has also been asserted (Hunt, 1984) that most (75%) of image restoration can be accomplished using the Wiener filter, and that images which are not amenable to Wiener filtering are often unrestorable altogether as there is often too much noise present in such images.

It is also worthwhile to point out the general applicability of the mathematics of image restoration. Image restoration is one example of a general inversion problem which occurs in many fields of science. Some examples include indirect sensing in atmospheric physics (Twomey, 1977), x-ray (Smith, 1984) and ultrasound (Greenleaf, 1984) tomography in medicine, determination of earthquake epicenters (Jobert and Cisternas, 1978) and electromagnetic prospecting methods in geology (Barthes and Vasseur, 1978), and numerous problems related to differential equations which reduce to Fredholm equations (Mikhlin, 1957 and 1960).

CHAPTER 3

LANDSAT MULTI-SPECTRAL SCANNER AND THEMATIC MAPPER IMAGING SYSTEMS

The Landsat satellites are earth resources observation systems conceived following the photographic observations of early Mercury and Gemini orbital flights to record data about the earth's surface and transmit those data to ground observing stations. Their purpose was originally as a research tool, but they soon became an indispensable tool for natural resource management and environmental monitoring. The first Landsat satellite, called ERTS-A (Earth Resources Technology Satellite), was launched in late July of 1972. A second satellite was launched in late January of 1975, and the program was then renamed Landsat to distinguish it from a planned oceanographic satellite to be called Seasat. Since then three more Landsat satellites have been launched, a major change having been made in 1982 when the Thematic Mapper sensor was added to the payload along with the previously used Multi-Spectral Scanner. The Thematic Mapper provides higher spatial and radiometric resolution and a greater variety of spectral bands than the Multi-Spectral Scanner (Slater, 1979).

Landsat System Descriptions

The Landsat satellites are in "sun-synchronous" orbits so that data are always collected at the same local time of day over a particular geographical region. This provides repeatable sun illumination conditions which is desirable when mosaicing adjacent images and comparing yearly changes in landscape cover.

Each of the first three Landsat satellites crosses over the same point on the earth once every eighteen days, the last two once every sixteen days. This repetitive coverage is obviously important for agricultural purposes, but less obviously is also important for geological purposes. For example, in many areas cloud cover obscures the view of Landsat so that frequent passes are required to collect adequate data. Even in areas where cloud cover is not a factor, such as Tucson, Arizona, important information can be gained from frequent coverage. In Tucson, a comprehensive analysis of the hydrogeology requires imagery from two different seasons: Winter imagery is useful for landform and drainage analysis, while spring imagery is important because of the close association of some springtime vegetation with the shallow water table and areas of ground water discharge (Taranik, 1978).

Both the Multi-Spectral Scanner and the Thematic Mapper are remote sensing instruments known as "object-plane electromagnetic scanners". Radiant energy from the earth is first reflected off a scanning mirror, which rotates back and forth in a direction orthogonal to the orbital track to provide coverage in the "along scan" direction; the forward motion of the spacecraft provides the coverage in the "along track" direction. The energy is then focused by a telescope through spectral filters onto an array of visible and near infrared light detectors at the prime focal plane and, for the Thematic Mapper, to thermal infrared detectors in a cooled focal plane. These detectors convert the scene energy into analog electrical signals which are then filtered, sampled and quantized. The resulting digital signal is then encoded and transmitted to ground receiving stations and subsequently processed. The differences in the two systems are in the details of each system component.

Detailed System Descriptions

Both the Multi-Spectral Scanner and Thematic Mapper employ a two-mirror reflective telescope design called a Ritchey-Chretien telescope with effective focal lengths of 82 cm and 244 cm, respectively (Slater, 1979; Hughes Aircraft Santa Barbara Research Center, 1984).

The spectral filters determine the radiant energy frequency range (band) over which each detector will respond. The Multi-Spectral Scanner has four bands, two visible and two near infrared. (A thermal infrared band was also used briefly on Landsat-3.) The Thematic Mapper has seven bands, three visible, one near infrared, two short wavelength infrared and one thermal. A material on the earth's surface has specific reflectance and emission characteristics in each spectral band, called its "spectral signature". The sensor's spectral coverage defines what spectral signature will be provided by the sensor. Experience with the earlier Multi-Spectal Scanner indicated the need for more bands to satisfy an expanding user community. The seven bands were chosen to assure use of Landsat Thematic Mapper data in many fields including agriculture, forestry, geography, geology, hydrology, meteorology and oceanography.

The size of the detectors determine the "instantaneous field of view", the ground projected dimension of the detector at nadir. This quantity is related to the spatial resolution of detail on the ground. However, the tradeoff is that a small detector leads to reduced signal-to-noise which, in turn, leads to reduced radiometric resolution. The instantaneous field of view for the Multi-Spectral Scanner is 76 meters square (Park, et. al., 1984) and for the Thematic Mapper it is 30 meters square except for the thermal band for which it is 120 meters square (Hughes Aircraft Santa Barbara Research Center, 1984).

The signal from the detectors is amplified and filtered. The motivation for filtering is to reduce the effects of aliasing, an inevitable consequence of sampling. The disadvantage to filtering is that electronic analog filters usually "roll

off" gradually starting at low frequencies, thus losing also some low frequency modulation transfer. The filters used in the Landsat systems are designed to provide as sharp a cut-off as practicable. The Multi-Spectral Scanner uses what is called a "Butterworth filter" and the Thematic Mapper uses a modified Butterworth called a "Goldberg filter".

The filtered analog signal is then sampled and quantized. In the Multi-Spectral Scanner each sample is converted to a 6-bit binary word, giving a resolution of 64 gray levels. In the Thematic Mapper each sample is converted into an 8-bit binary word, a resolution of 256 gray levels.

System Component Transfer Functions

Each of the system components described above can be represented by a point spread function (or, equivalently, a transfer function) which can be derived either from engineering design and test data or from post-flight evaluations.

The optics of imaging systems are often characterized in terms of a "blur circle". This is commonly defined as the diameter of the circle which contains 90 percent of the total energy from a point source input. The ground projected value for this diameter is 30 meters for the early three Multi-Spectral Scanners (Slater, 1979), 23 meters for the later two (Markham, 1985) and 35 meters for both the Thematic Mappers (Hughes Aircraft Santa Barbara Research Center, 1984). The optical point spread function may then be modeled as a Gaussian with the blur circle diameter D as a characteristic width (Park, et. al., 1984)

$$h_o(x,y) = e^{-(x^2+y^2)/D^2} \tag{3.1}$$

The associated transfer function is also a Gaussian:

$$H_o(u,v) = \pi D^2 e^{-\pi^2 D^2 (u^2+v^2)} \tag{3.2}$$

The parameter D may be calculated to result in agreement between the model and the blur circle. It has been assumed that the spread and transfer functions

of the optics are separable and symmetric, a valid assumption for high quality imaging systems.

Light arriving at the focal plane of the telescope is transferred by fiber optics to the individual detectors in the case of the Multi-Spectral Scanner, and falls directly on the detectors in the case of the Thematic Mapper. If these fibers are assumed to be spatially uniform in their light transmission characteristics and the receiving detectors are assumed to be square, then the detector point spread function is a two-dimensional square with width and length equal to the detector dimension w,

$$h_d(x,y) = \begin{cases} 1, & |x| \leq \frac{w}{2},\ |y| \leq \frac{w}{2} \\ 0, & \text{elsewhere} \end{cases} \tag{3.3}$$

The corresponding transfer function is (Champeny, 1983)

$$H_d(u,v) = \frac{\sin \pi u w}{\pi u} \frac{\sin \pi v w}{\pi v} \tag{3.4}$$

The electronic filters used in both Landsat systems are critical to the overall system response. These filters, unlike the other components of the system, introduce a nonlinear phase shift into the output of the system. The transfer function of the Butterworth filter used in the Multi-Spectral Scanner is given by

$$H_B(u) = \frac{1}{1 + 2i(u/u_c) - 2(u/u_c)^2 - i(u/u_c)^3} \tag{3.5}$$

where u_c is a "cutoff" frequency (Markham, 1984; 1985). This filter is only effective in the along scan direction. It does not effect the data along track because the effective scanning velocity in that direction is slower by several orders of magnitude than the scan velocity of the mirror (Markham, 1985). This transfer function can be decomposed into magnitude and phase

$$|H_B(u)| = \left(\frac{1}{1 + (u/u_c)^6} \right)^{1/2} \tag{3.6}$$

$$\arg H_B(u) = -\arctan(u/u_c) - \arctan\left(\frac{u/u_c}{1 - u/u_c)^2} \right) \tag{3.7}$$

Eq. (3.7) indicates the nonlinear phase response of the Butterworth filter. This results in an asymmetric spread function along scan, as can be seen in Figure 1.

The transfer function of the Goldberg filter used in the Thematic Mapper, also only effective along scan, is given by (Markham, 1984; 1985)

$$H_G(u) = \frac{1}{(i(u/u_1)+1)} \frac{1}{1-(u/u_2)^2+2Li(u/u_2)} \tag{3.8}$$

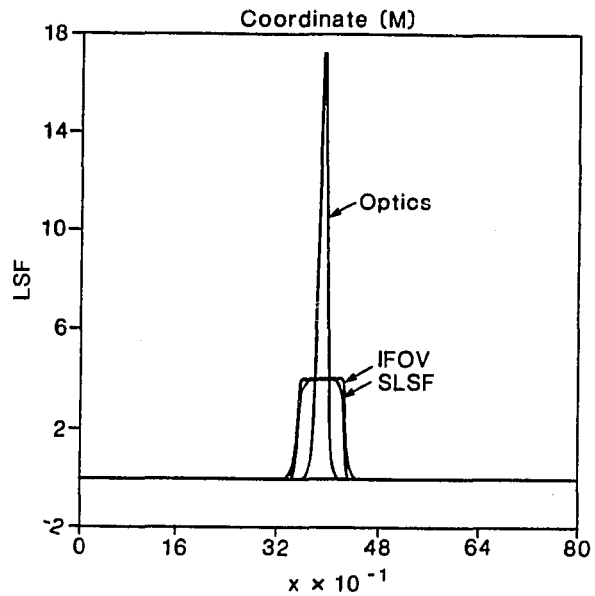
where u_1 and u_2 are cutoff frequencies and L is a damping ratio. The magnitude and phase of this filter with L = 0.5 (Markham, 1985) are

$$|H_G(u)| = \left(\frac{1}{((u/u_1)^2+1)} \frac{1}{1-(u/u_2)^2+(u/u_2)^4} \right)^{1/2} \tag{3.9}$$

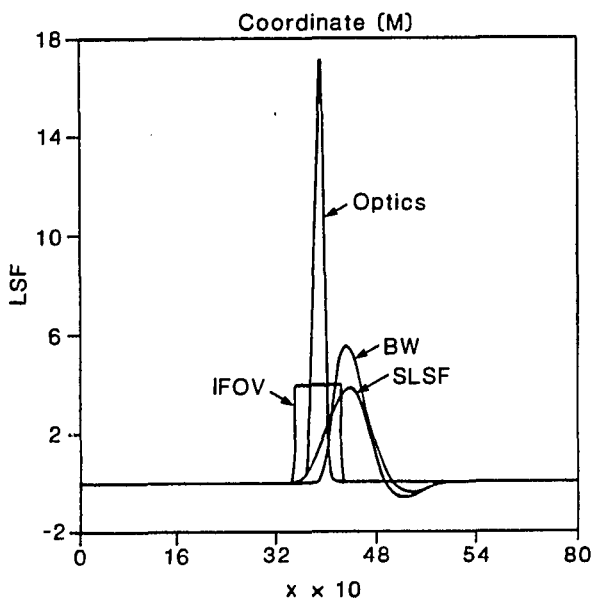
$$\arg H_G(u) = -\arctan(u/u_1) - \arctan\left(\frac{(u/u_2)}{1-(u/u_2)^2} \right) \tag{3.10}$$

This filter also has a nonlinear phase response resulting in the asymmetric spread function along scan shown in Figure 2. (The along track spread functions are similar to those for the Multi-Spectral Scanner.)

The filtered analog signal is sampled and quantized by an A/D converter. The time interval over which the signal is sampled, the "integration time", is 80 nanoseconds for the Multi-Spectral Scanner, which corresponds to a blur of less than 0.5 meters in the along scan direction (Slater, 1979). The Thematic Mapper does not integrate but uses a "track and hold" operation. The effective integration time for this operation is not available in the engineering reports; it is assumed to be small. The integration time for the Landsat scanners could be modeled as a linear along scan blur, but it is so small as to usually be considered negligible and ignored. The effect of sampling will be discussed in Chapter 5. However, it should be noted that the sample interval (or pixel width) is not necessarily the same as the instantaneous field of view. In the case of the Multi-Spectral Scanner, for example, the along scan samples correspond to overlapping instantaneous fields of view (thus the data is "oversampled"), while the



ALONG-TRACK LSFs FOR MULTI-SPECTRAL SCANNER



ALONG-SCAN LSFs FOR MULTI-SPECTRAL SCANNER

Figure 1. Spread Functions for the Multi-Spectral Scanner.

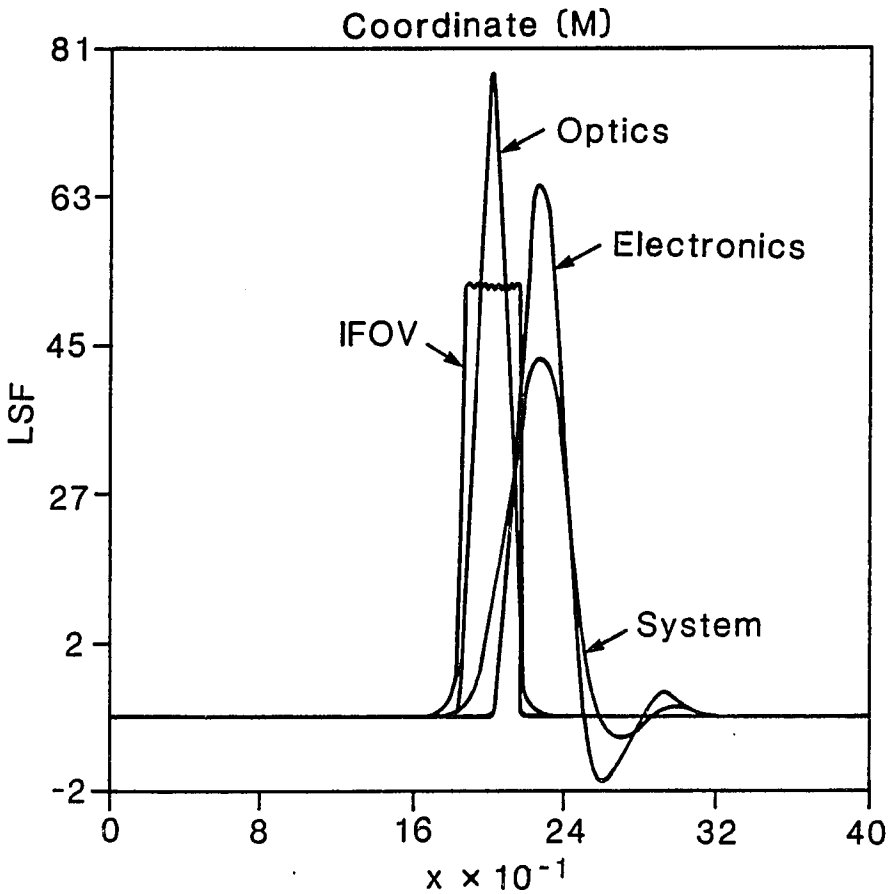


Figure 2. Spread Functions for the Thematic Mapper.

along track samples are not quite contiguous relative to the instantaneous field of view. Finally, error due to quantization can be modeled statistically, but this effect has not been included in this study.

These linear system components together with sampling combine to yield a net instantaneous field of view which is larger than that due to the detector alone. The effective instantaneous field of view is the spatial dimension equivalent to half a cycle of the spatial frequency where the modulus of the system transfer function falls to 0.5. The line spread function width at half the maximum is a comparable measure in the spatial domain (Markham, 1985). For the Thematic Mapper the effective instantaneous field of view has been estimated as 32 to 33 meters along track and 36 meters along scan, excluding sampling effects, (Markham, 1985) or 40.8 to 48.6 meters square, including sampling effects, (Schowengerdt, et. al., 1985a) and for the later Multi-Spectral Scanners as 70 to 75 meters along track and 79 to 82 meters along scan, excluding sampling, (Markham, 1985) or 65 meters along track and 77 meters along scan for the earlier systems, excluding sampling (Park, et. al., 1984).

The overall system transfer functions for each Landsat system (exclusive of sampling) can be found by cascading the transfer functions of the linear components.

$$H(u,v) = H_o(u,v)H_d(u,v)H_f(u,v) \tag{3.11}$$

or, assuming separability,

$$\begin{aligned} H(u) &= H_o(u)H_d(u)H_f(u) \quad \text{along scan} \\ H(v) &= H_o(v)H_d(v) \quad \text{along track} \end{aligned} \tag{3.12}$$

(Figures 3 and 4), where H_f denotes either of the electronic filters, and then incorporating sampling effects as discussed in Chapter 5. Important information about the actual performance of the system can only be gained, however, by in-flight experiments.

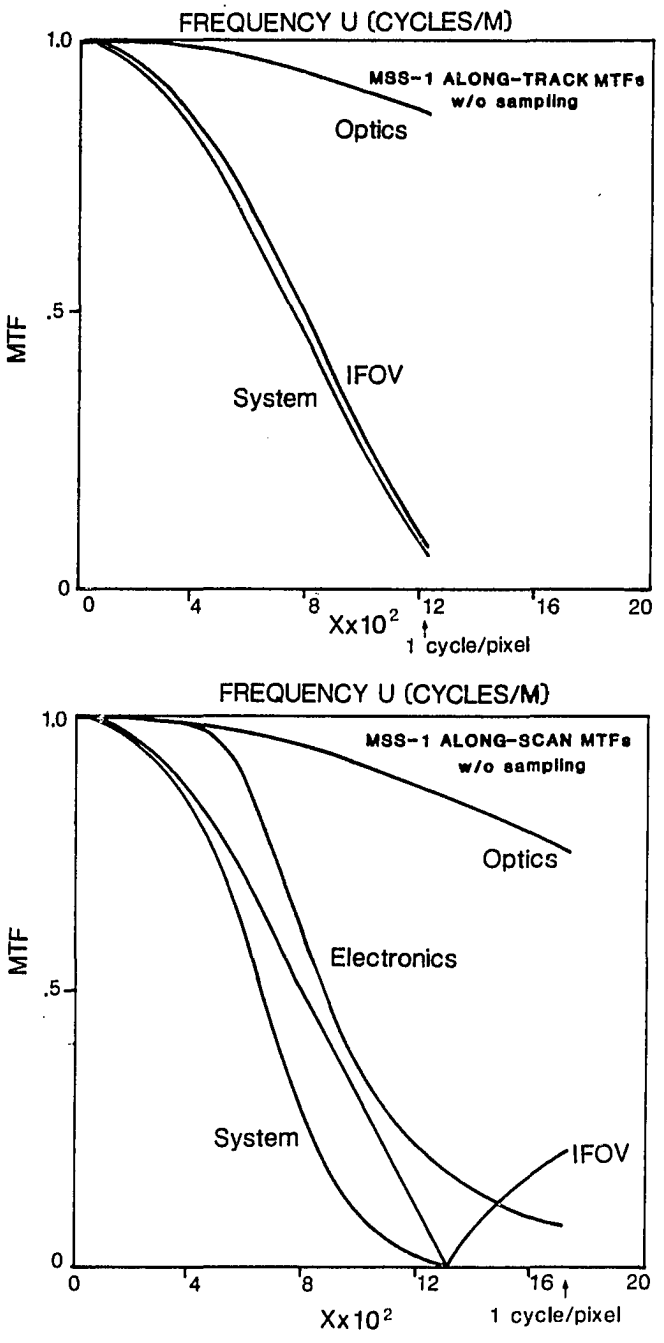


Figure 3. Transfer Functions for the Multi-Spectral Scanner.

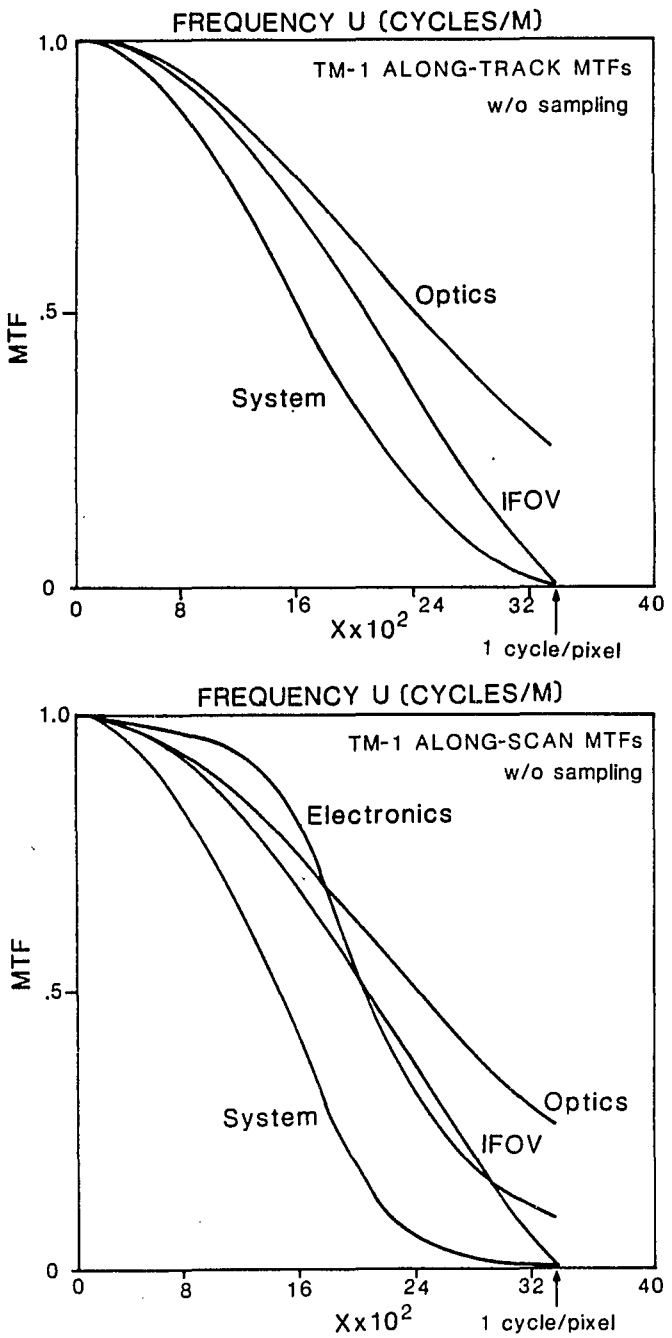


Figure 4. Transfer Functions of the Thematic Mapper.

In-Flight Evaluation

Recall that with the assumption of linearity, the response of a system is characterized completely by its point spread function $h(x,y)$ which describes the output of the system to a point source input. With the added assumption of shift invariance, the point spread function is independent of the location of the point in the object field, and the output to any arbitrary input is the two-dimensional convolution of the input with the point spread function. The direct measurement of the two-dimensional point spread function is generally not feasible, due to the inability to get sufficient energy concentrated at a point source. What is generally measured are one-dimensional sections of the two-dimensional point spread function, called "line spread functions". A line spread function $h_l(x)$ is the response of the system to an infinitely narrow line source and is defined as

$$h_l(x) = \int_{-\infty}^{\infty} h(x,y)dy \tag{3.13}$$

If the system is further assumed to be separable, then

$$h(x,y) = h_{l1}(x)h_{l2}(y) \tag{3.14}$$

the two-dimensional point spread function is the product of two one-dimensional line spread functions, one along track and one along scan.

Attempts to characterize the in-flight Landsat line spread functions require the availability of line sources on the surface of the earth. These line sources cannot be arbitrarily narrow because contrast must be maintained in the image, where "contrast" here is defined as the ratio of object gray level (intensity) to background gray level. At the same time, a source must be narrow compared to the width of the line spread function to qualify as a "line source". A suitable source for the Thematic Mapper is the San Mateo Bridge over the south end of San Francisco Bay (Schowengerdt, et. al., 1985b). The contrast

between bridge and water is high, the bridge is straight over most of its extent, and it is close enough to the water to minimize projected shadows. It is 18.3 meters wide, compared to 30 meters for the instantaneous field of view of the Thematic Mapper. The bridge is at an angle of 30° to the scan direction of the sensor, so that the bridge has a sampling phase associated with it (Chapter 5). Thus, measurements made using this data are, in effect, an average of the sampled system performance functions, sampling being an important effect not accounted for in the engineering reports.

The modulation transfer function for the Thematic Mapper was measured in the direction orthogonal to that of the bridge (Schowengerdt, et. al., 1985b). The modulation transfer function, the modulus of the Fourier transform of the one-dimensional point spread function, describes how the system modifies the amplitude of the various spatial frequency components of an input to the system. The line spread function which resulted from this study was wider than that acquired from the engineering design and test data. This is because, when measuring system response functions from images, the effects of atmospheric blurring, sampling and ground processing are being included in the data. It was found that this resulted in an average effective instantaneous field of view of 40 to 49 meters, larger than that measured for the imaging and electronics system alone. These results have been verified by others (e. g. Malaret, et. al., 1985) who also concluded that the overall resolution is less than that which the optics and electronics alone would dictate.

CHAPTER 4

GEOMETRIC RESAMPLING AND RESTORATION

The most important aspect of the ground processing of Landsat data is geometric manipulation. These geometric corrections usually require a two-dimensional "warping" of the image to compensate for geometric distortions in the image so that the image can be registered to a map or another image. The typical approach is to try to model the continuous scene that was sampled to form the image, and to interpolate pixel gray level (intensity) values at the intermediate locations required to form an image with the desired geometric characteristics. The particular interpolator used is an important issue, as the interpolator acts as an added convolution, or blurring, of the image. Using a restoration-derived interpolator would act to sharpen, rather than blur, the image.

Geometric Rectification

Numerous sources contribute to the distortion of unprocessed data which make geometric rectification necessary. These distortions can be grouped into two categories, those which are consistent over time and can be corrected in a systematic manner and those which are unpredictable and must be corrected using mathematical distortion models.

Predictable Distortions

The most obvious distortion which falls into the first category is that due to the curvature of the earth. This appears as an apparent spreading of the image off nadir.

A non-unity aspect ratio (ratio of pixel width to height) is introduced by the along scan oversampling of the Multi-Spectral Scanner relative to the along track direction. The aspect ratio is one for the Thematic Mapper.

Other distortions in this category result from the fact that data is sampled equally spaced in time, not space. The design of the Landsat Multi-Spectral Scanner also results in nonlinear motion of the oscillating mirror as it sweeps west to east due to its deceleration and acceleration at the ends of each sweep. (Data is not acquired during the east to west retrace.) This results visually in an apparent compression or stretching of the image at the ends of each scan line (Van Wie and Stein, 1977). Distortion also is introduced from one scan line to the next due to the relative velocity of the spacecraft to a rotating earth. The Landsat satellites travel south (on the daylight side of the earth) in nonpolar orbits and scan from west to east. The rotation of the earth causes each successive mirror sweep to begin further to the west. The resulting distortion is proportional to the cosine of the latitude, and is greatest at the equator. The Multi- Spectral Scanner senses six scan lines in each of four bands at each mirror sweep. The visual effect of this combined with the earth's rotation is for the image to be skewed in a step-wise fashion. This six-scan-per-mirror-sweep design also causes band-to-band misregistration with respect to observed ground positions.

Finally, small time delays between the readout of each of the six sensors give rise to small displacements between each of the six scan lines due to the movement of the scan mirror during this time (Van Wie and Stein, 1977).

All these sources of distortion are well understood and predictable and can be mathematically modeled. Furthermore, they will occur in every Landsat image so that all such distortions can be "systematically" corrected.

Unpredictable Distortions

Among the causes of unpredictable distortions is variation in the spacecraft attitude — yaw, roll and pitch. Yaw, rotation of the spacecraft about the vertical axis from it to the earth's center, causes the angle between the scan direction and the orbit to deviate from 90° and looks visually like an overall skew in the image. Roll, rotation of the spacecraft about its velocity vector (or longitudinal axis), appears in the image as an apparent stretching or shrinking, much like the distortion due to the nonlinear mirror velocity. Finally, pitch, the rotation of the spacecraft about the remaining orthogonal axis, changes the effective instantaneous field of view resulting in radiometric, rather than geometric, distortion (Gillespie, 1980).

Also included in this category are indeterministic variations in the spacecraft velocity, which result in distortion both within and between scan lines, changes in altitude, and ephemeris variation which results from variations in the location of the spacecraft with respect to the ground with each successive pass over a given region (Kashet and Sawchuk, 1983).

All of these sources of distortion, while also well understood and modeled, do not occur in every image or to the same extent between those images in which they do occur. Thus, such corrections need special treatment.

Geometric Rectification Implementation

Most of these distortions are essentially one-dimensional and can be corrected by rectifying the data in the along scan direction only (Friedmann, 1981). Some, such as skewing caused by the earth's rotation and spacecraft yaw, can be compensated for by simply adding an increment to the sample position of each scan line since such errors can be thought of as misalignment of the pixels in the vertical direction (Gillespie, 1980). Other corrections require interpolation or "resampling".

Geometric Registration

Registration to a specific map projection requires a two-dimensional "warping" of the entire image. For example, the data is often registered to a standard cartographic projection such as the Universal Transverse Mercator map projection, a cylindrical projection used by the United States Geological Survey and in the United States military mapping program, as well as by other nations (Van Wie and Stein, 1977; Robinson, 1960). (Other commonly used map projections include the Hotine Oblique Mercator and the Polar Stereographic projection (Slater, 1979).) For the merging of data from different sensors, it is important that all the data be registered to the same map to yield satellite independent images (Friedmann, 1981). In other cases it might be desired to register one image to another taken of the same area at another time. This "multitemporal" data is used to detect any changes that had occurred (Lillestrand, 1972). In each case the goal is to change the geometry of the image so that features are found in the same location in image and map or from one image to the other. Warping of the data to fit a map is also often called "rectification", although it is actually a distortion of the data to fit a map and not the removal of distortion.

Resampling

One possible method for geometric correction is to first perform rectification of the predicted distortions in the along scan direction, then perform rectification of the variable distortions along with registration in both dimensions. There are two ways in which this can be accomplished. The location of the pixels can be changed, or the original pixel grid can be retained and each pixel assigned a new gray level (intensity) value. The latter approach is the one usually taken, and is called "resampling". The motivation is to try to model the continuous scene that was sampled to form the image, and to resample this

model at the intermediate pixel locations required to form an image with the desired geometric characteristics. Both registration and the along scan rectification can be performed by resampling, but for the former, since the cause of the distortions is not known precisely, there is the need to use an empirical distortion model.

Geometric rectification and registration thus takes place in two steps: Determination of a projection function which will transform the geometry of the image to that of the map, thus reconstructing a (model) continuous image, and interpolation of the image gray levels to determine the correct gray level values to assign to the pixels.

For the "systematic corrections", the displacement of the sampling grid is known from prior knowledge of the distortions introduced in the imaging process. For the variable distortions and registration, a displacement function must first be found. First match points or "ground control points" whose locations are known precisely are located in both image and map. Such points often correspond to features like road or railway intersections. Location of these points may be done manually or digitally using cross-correlation techniques. The number, location and distribution of these control points is the most important step in the registration process (Shlien, 1979), since the accuracy of the registered image depends directly on the accuracy of these points.

Once the control points are identified, a bivariate polynomial surface may be used to model the geometry of the scene in the area between the control points. Often the image is broken up into quadrilaterals or triangular patches with the control points defining the vertices and a separate polynomial surface fit to each subimage. Alternatively, a single global polynomial may be used if the distortions are sufficiently smooth. A typical transformation is a polynomial of the form

$$x = a_o + a_1 x' + a_2 y' + a_3 x' y' + a_4 x'^2 + a_5 y'^2$$
$$y = b_o + b_1 x' + b_2 y' + b_3 x' y' + b_4 x'^2 + b_5 y'^2 \qquad (4.1)$$

where (x, y) are coordinates for the original image and (x', y') are for the map (Schowengerdt, 1983). Higher order polynomials may be used, however, it has been shown (Lillestrand, 1972) that little is gained by using polynomials of order greater that three, and much is lost by using high order polynomials (greater than five) since the surface then tends to follow noise.

This projection function defines a continuous image estimate (model). Once that model has been determined, it can be evaluated at the desired grid locations. It is necessary to determine the "best" gray levels to assign to the new pixel locations, given the set of samples from the original image. The problem then is to determine an appropriate interpolation or resampling function.

Nearest Neighbor

A straightforward approach might be simply to choose as the new gray level value that of the pixel nearest the one being interpolated. This "nearest neighbor" approach can be implemented quickly and efficiently and preserves the radiometric fidelity of the image. However, it can lead to position errors of up to $\pm 1/2$ pixel and result in a blocked appearance in the final image.

Bilinear Interpolation

A second approach is to assign a gray level value to the new pixel location which is a bilinear interpolation of the nearest four pixel gray levels. Figure 5 illustrates a bilinear interpolator.

Cubic Convolution

If the function describing the scene intensity distribution $f(x, y)$ is band limited, that is, its spectrum $F(u, v)$ is such that $F(u, v) = 0$ when $|u| \geq u_c$ or

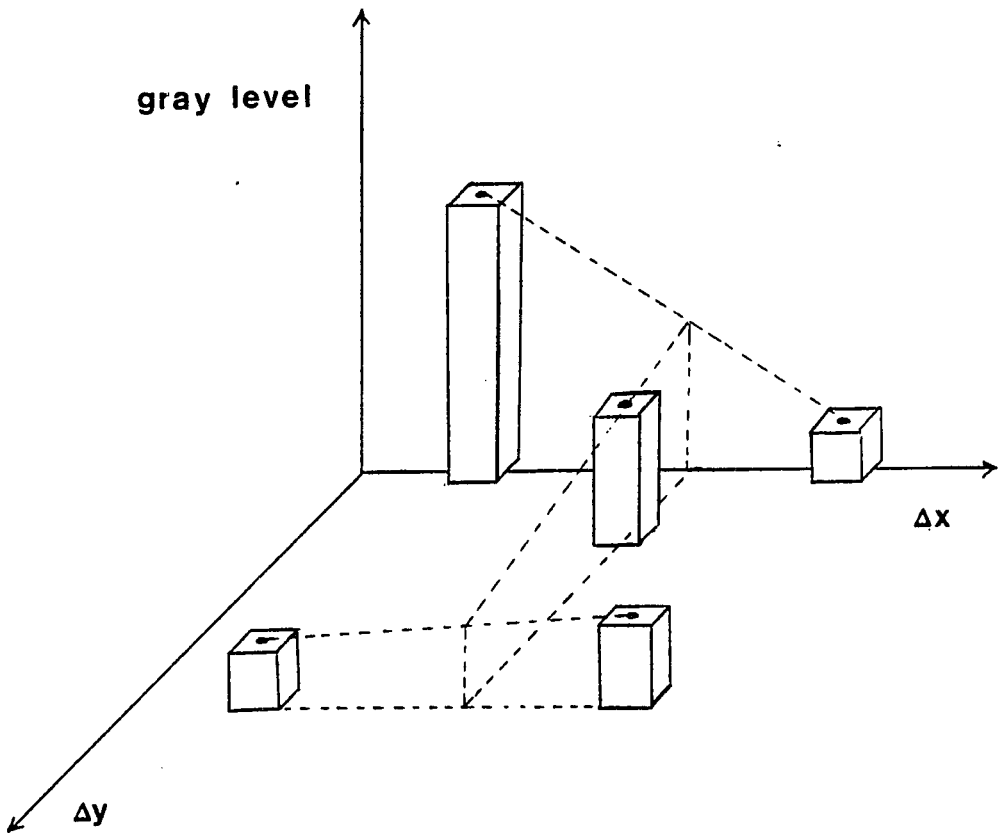


Figure 5. Bilinear Interpolator.

$|v| \geq v_c$, and sufficiently sampled (i. e. at a sampling rate at least twice per period for the highest frequency present in the signal), then $f(x,y)$ can be reconstructed exactly from convolution with a $(\sin(\pi x)/\pi x) \times (\sin(\pi y)/\pi y)$ function. (Proof in Appendix A.) It is impossible to implement this ideal interpolator numerically because of its infinite extent. Moreover, the function $f(x,y)$ is known only over a finite extent. If the interpolator were truncated, this would introduce ringing in the reconstructed $f(x,y)$, the familiar artifact referred to as Gibbs phenomenon. Thus a piecewise cubic polynomial of limited extent which approximates the $\sin(\pi x)/\pi x$ is usually implemented (typically using look-up tables, Appendix B). It is given, in one dimension, by (Park and Schowengerdt, 1983)

$$r(x) = \begin{cases} (\alpha+2)|x|^3 - (\alpha+3)|x|^2 + 1 & |x| < 1 \\ \alpha|x|^3 - 5\alpha|x|^2 + 8\alpha|x| - 4\alpha & 1 \leq |x| < 2 \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

This interpolation function is commonly known as "parametric cubic convolution" and has several important features. It is even, zero for all $|x| \geq 2$, $|y| \geq 2$, has continuous slope for all x, reconstructs the original image samples exactly and reproduces exactly a constant image.

Resampling as Convolution

All of these interpolators act as an added convolution, or blurring, of the image, which has important implications for the frequency content of the image. Since convolution in the spatial domain is equivalent to multiplication in the frequency domain, the form of the Fourier transform of these interpolators can modify significantly the spectrum of the image. Figure 6 shows these interpolators and their Fourier transforms. From these figures it is seen that nearest neighbor, which is equivalent to convolution with a rectangle function and bilinear interpolation, which is equivalent to convolution with a triangle function (Shlien, 1979) both suppress the higher frequencies leading to loss of

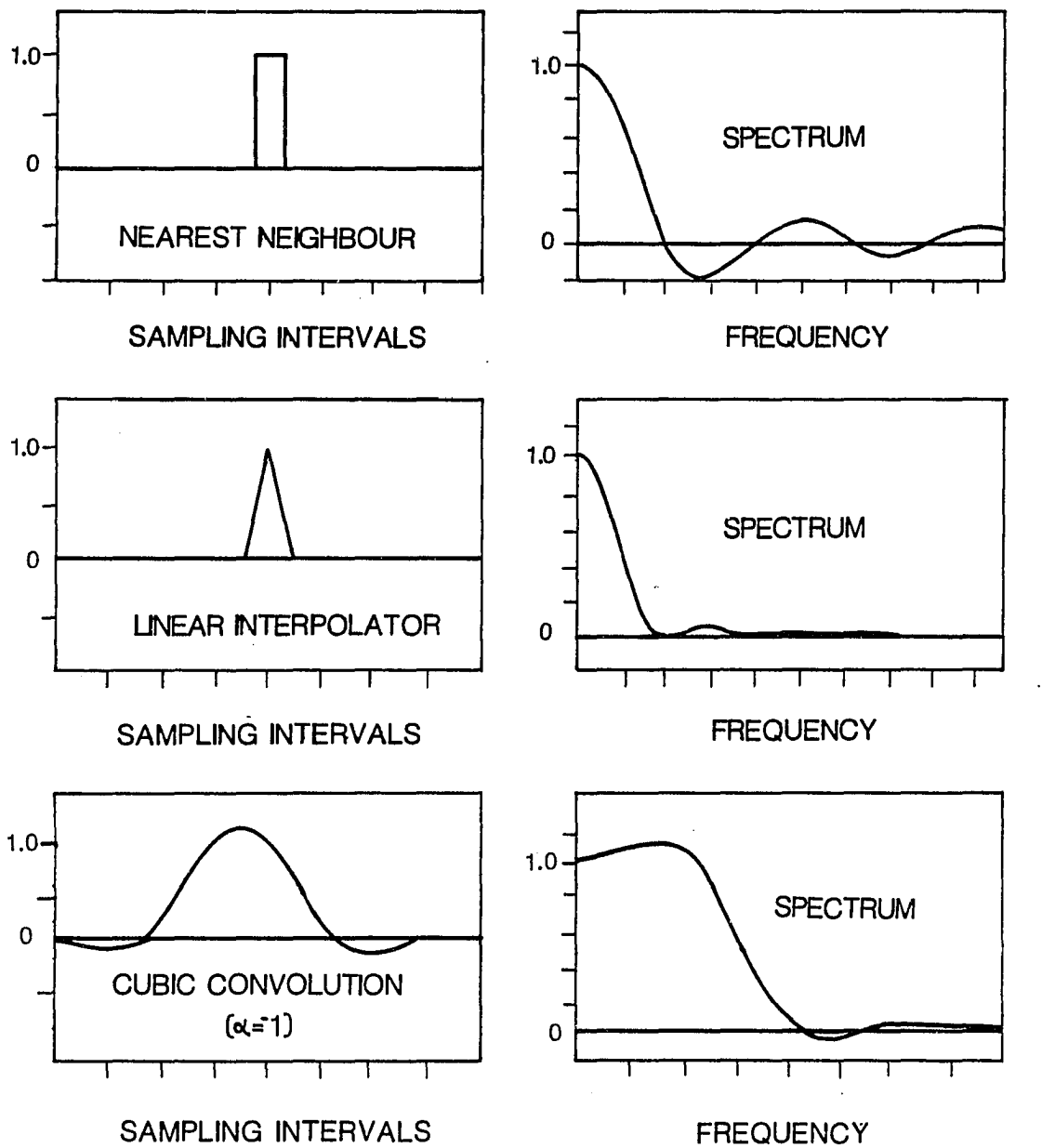


Figure 6. Interpolation Functions and Their Spectra.

resolution in the reconstructed image. The parametric cubic convolution interpolator preserves higher frequencies, but with $\alpha = -1$ exhibits ringing near edges.

Using a restoration filter as the interpolator would boost, rather than suppress, those frequencies which had already undergone some suppression in the imaging process. The idea is to use a restoration filter as the interpolator to account for blurring that had already occurred, rather than to introduce more blurring (high frequency suppression) in the image.

EROS Data Center Implementation

At the Earth Resources Observation System (EROS) Data Center in Sioux Falls, South Dakota, Landsat Multi-Spectral Scanner images are rectified and registered using a two-pass method. The images are first resampled in the along scan direction to account for the systematic corrections, then they are resampled a second time in both directions for registration to maps or other images. In the past, cubic convolution with $\alpha = -1$, or more recently with $\alpha = -1/2$ (Park and Schowengerdt, 1983), has been the resampling method of choice. Landsat Thematic Mapper images, on the other hand, are resampled for systematic corrections using cubic convolution before arrival at EROS, so that use of the restoration kernel can only occur in the registration process.

Resampling for restoration using the software developed as part of this dissertation has been successfully implemented at EROS for both Landsat scanners. Examples are shown in Chapter 8. Ideally, the restoration kernel should include a term which would account for the second resampling pass using cubic convolution, but it presently does not. This, and other modifications to be discussed in Chapter 9, would be logical extensions of this work.

CHAPTER 5

WIENER FILTER FOR SAMPLED IMAGING SYSTEMS

Much engineering system analysis is based on assumptions of linearity and shift invariance. Eq. (2.5), often the defining equation for system analysis, can also be the defining equation for combined linearity and shift invariance (Brown, 1962). Eq. (2.5) is a realistic representation of many types of systems (Gaskill, 1978), but it also has very attractive mathematical qualities. In particular, Eq. (2.5) represents spatial convolution, which reduces to simple multiplication in the Fourier domain. Furthermore, if a system is made up of several linear shift invariant components, the overall transfer function of the system is simply the product of the transfer functions of the individual components.

Most discussions of image restoration assume linearity and shift invariance of the imaging system, and, furthermore, that both the object and its image are continuous. However, many modern imaging devices, including the Landsat Multi-Spectral Scanner and Thematic Mapper, are "sampled" systems. That is, the continuous output of the electronic sensors is sampled and digitized before being coded and transmitted to the ground.

The importance of including the effect of sampling in the system analysis was pointed out by Park, et. al. (1984) in a study of sampling effects for the Landsat Multi-Spectral Scanner. It was shown that the effect of sampling is to increase the average amount of blurring in the image, especially in the along track direction. Sampling increases the effective instantaneous field of view from 77 m to 86 m along scan and from 65 m to 122 m along track. Thus, incorporation of sampling into the system transfer function, and therefore into

the restoration procedure, should improve the restored image since another blurring effect has been removed.

Sampled systems are not shift invariant. The output of a sampled system resulting from a point source input will depend not only on the intensity of the point source but also on its location with respect to the sampling grid. The output of a point source located precisely on a sample point will be different from one located, for instance, between two sample points.

This "sample scene phasing" has an important effect on the derivation of the system's transfer function. The mathematical function describing sampling cannot be simply cascaded (multiplied) with the other system element transfer functions to produce an overall system transfer function, because the process of cascading transfer functions depends on the assumptions of linearity and shift invariance of the component functions.

The major contribution of the present research is the explicit inclusion of sampling degradation in the design of the Wiener filter for restoration.

Stochastic Approach to Sampling

One way of dealing with this sampling phenomenon is a stochastic approach (Park, et. al., 1984) in which sampling effects are accounted for by assuming a random distribution of point sources and averaging the overall system output over all possible point source locations. This results in an average system transfer function, or in the spatial domain, an average system point spread function. This averaging is necessary because one does not know the exact distribution of the point sources in the scene. If one did then restoration would be pointless since the original scene would already be known!

The procedure used by Park, et. al., is to derive a shift variant sampled system point spread function $h_s(x, y; s, t)$, which is then averaged over all randomly located input point sources. The sampling of the continuous output of

the imaging (sub)system results in a sampled image which can be represented mathematically by

$$g_s(x,y) = [f(x,y) * h(x,y)] \sum_m \sum_n \delta(x-m, y-n) \tag{5.1}$$

where both the object and image are referenced to a common orthogonal coordinate system.

The system response to a single point source input is important because it defines the system response to any input (Gaskill, 1978).

$$h_s(x,y;s,t) = [h(x-s, y-t)] \sum_m \sum_n \delta(x-m, y-n) \tag{5.2}$$

where the notation indicates that the system is shift variant. The point spread function h (exclusive of sampling) represents an image formation system whose output from a point source input depends only on the intensity of the point source and not on its location in the object plane. The parameters s and t locate the point source with respect to the sampling grid. The effect of adding sampling, represented by the double sum in Eq. (5.2), is to make the overall system point spread function h_s lose that invariance. The output to a point source input now depends both on its intensity and its location.

If the parameters s and t are assumed to be random variables having a uniform distribution (that is, a point source is equally likely to be located anywhere with respect to the sampling grid), then an average system point spread function $h_a(x,y)$ can be found (Park, et. al., 1984) by adding (or integrating) the system point spread function over all possible point source locations. This results in

$$h_a(x,y) = \int_0^1 \int_0^1 h_s(x,y;s,t)\,ds\,dt \tag{5.3}$$

where the periodicity of the sampling grid has been used to restrict $0 \leq s \leq 1$, $0 \leq t \leq 1$. Using the assumption of the linearity of the imaging system (exclusive

of sampling) together with the linearity of averaging and the fact that s and t are assumed to be uniformly distributed yields

$$h_a(x,y) = (r(x-\frac{1}{2},y-\frac{1}{2}) * h(x,y)) \sum_m \sum_n \delta(x-m,y-n) \tag{5.4}$$

where

$$r(x-\frac{1}{2},y-\frac{1}{2}) = \begin{cases} 1 & 0 \le x < 1, 0 \le y < 1 \\ 0 & \text{elsewhere} \end{cases} \tag{5.5}$$

Taking the Fourier transform of Eq. (5.4) yields the average system transfer function, including sampling,

$$H_a(u,v) = T(u,v)e^{-\pi(u+v)i} \tag{5.6}$$

where

$$T(u,v) = \sum_m \sum_n (-1)^{m+n} \frac{\sin \pi(u-m)}{\pi(u-m)} \frac{\sin \pi(v-n)}{\pi(v-n)} H(u-m,v-n) \tag{5.7}$$

The infinite series represents the inevitable consequence of periodic replication caused by sampling (Brigham, 1974). Because of this replication, $H_a(u,v)$ is periodic while $H(u,v)$ is not, resulting in false information at frequencies beyond 0.5 cycles/pixel, the "Nyquist frequency".

The effect of sampling on the overall system transfer function is illustrated in Figure 7, which shows the modulus of the transfer functions of both Landsat systems with and without sampling. The figure shows an increased suppression of the higher frequencies (corresponding to fine detail in an image) in the sampled case. The effect of sampling, then, is to further degrade the image.

This average sampled system transfer function, Eq. (5.6), which accounts for the combined effects of imaging and sampling, may be used in the Wiener filter Eq. (2.36) resulting in

$$M(u,v) = \frac{1}{H_a(u,v)} \frac{|H_a(u,v)|^2}{|H_a(u,v)|^2 + S_n(u,v)/S_f(u,v)} \tag{5.8}$$

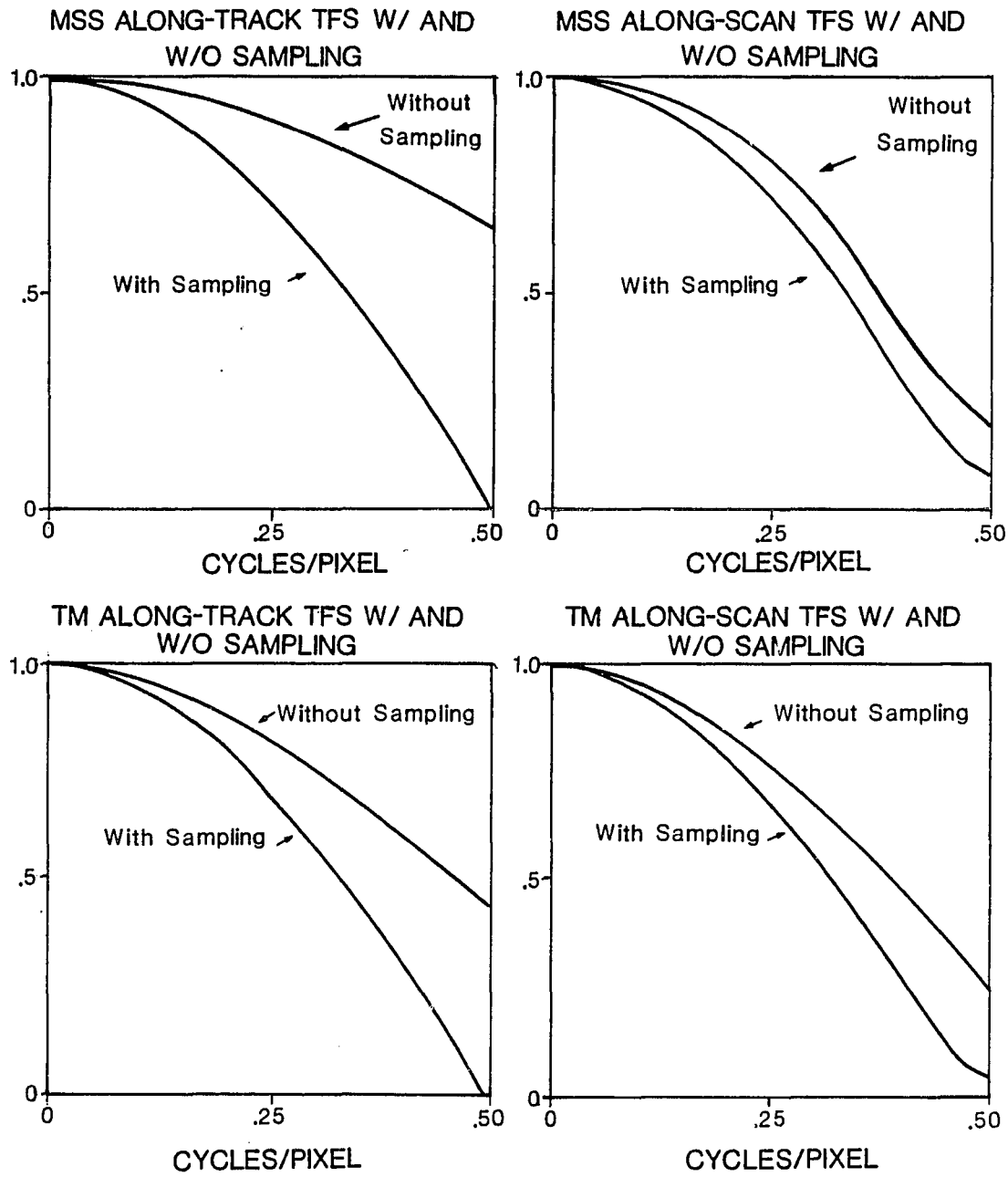


Figure 7. Landsat Transfer Functions With and Without Sampling.

As described in Chapter 2, the Wiener filter is derived for continuous functions and is that filter which minimizes the mean square error between the original scene and the restored image (estimated scene). Using the average system transfer function $H_a(u,v)$ in the Wiener filter results in a suboptimal filter, in the least squares sense, since $H_a(u,v)$ is not continuous. A Wiener filter which has been optimized for sampled imaging systems was derived by Huck, et al. (1985) and is given by

$$
\begin{aligned}
&M_s(u,v) = \\
&|S_f(u,v)|^2 H^*(u,v) \times \\
&1/\Big\{|S_f(u,v)H(u,v)|^2 + \sum \sum_{(m,n)\neq(0,0)} |S_f(u-m,v-n)H(u-m,v-n)|^2 \\
&+ (S_n(u,v))^2\Big\}. \qquad (5.9)
\end{aligned}
$$

This filter is shown in Figure 8 along with the filter of Eq. (5.8) for scene and noise statistics typical of Landsat imagery as discussed in Chapter 7 (in particular, for a variance of 400 and a correlation length of 5). For such statistics, the filter of Eq. (5.9) corresponds to the continuous Wiener filter of Eq. (2.36) (i. e., differs only in the 7th decimal place), since for these statistics the double sum in the denominator is negligable relative to the other terms there. (If the scene power spectrum contained a great portion of its power above 0.5 cycles/pixel, this would not be the case.) The effect of using the suboptimal Wiener filter is to shift the enhancement toward lower frequencies, that is, to those frequencies which have not been aliased. It could be argued that this corresponds to less enhancement of the fine detail in the image, but it could also be argued that this leads to less noise enhancement, noise also being a high frequency component. Indeed, although the mean squared error will, on the average, be greater for the suboptimal filter restorations, we will see in Chapter 8 that the subjective image quality will be improved.

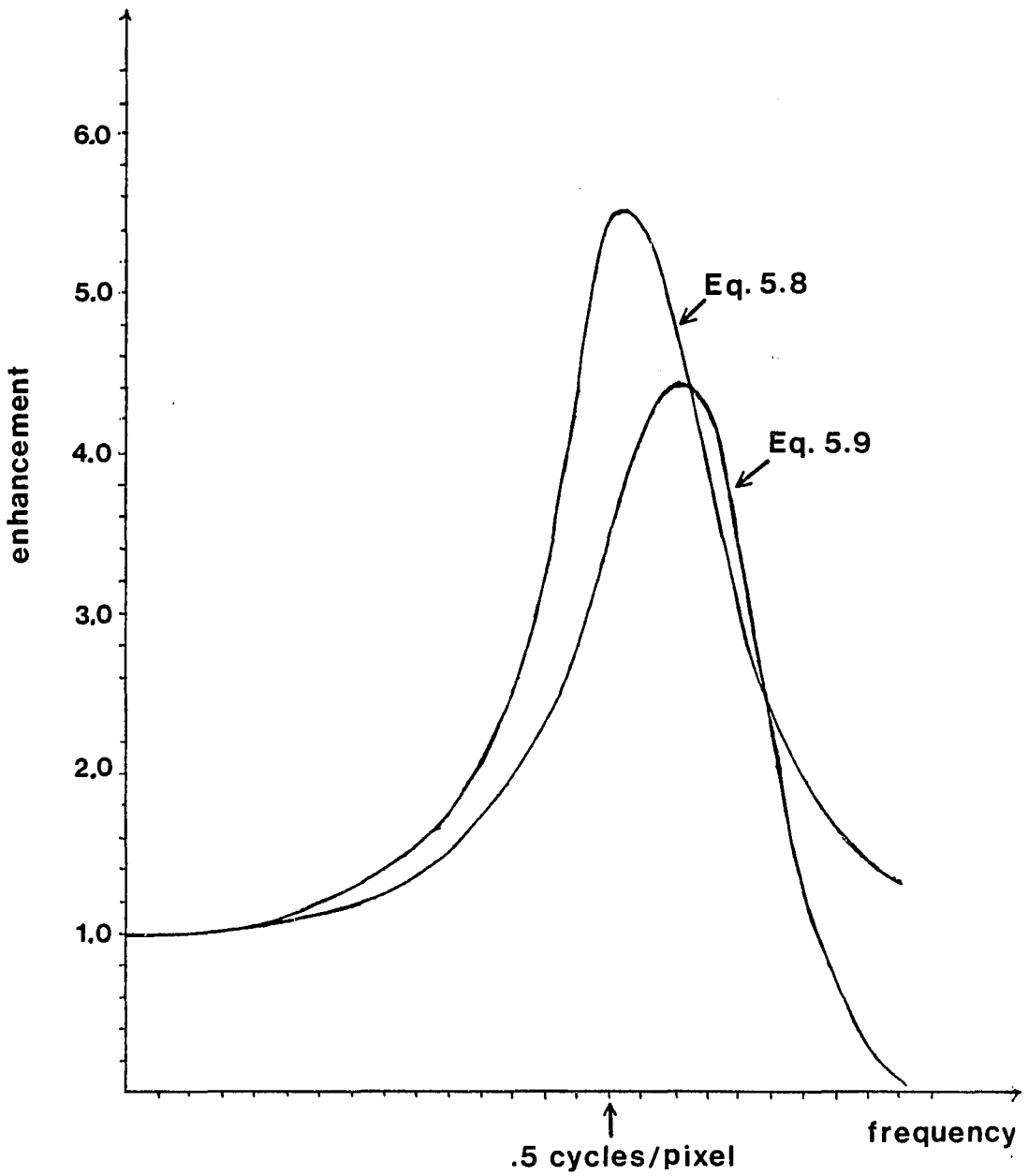


Figure 8. Comparison of Optimal and Suboptimal Wiener Filters with Sampling.

Separability of the Wiener Filter

In modeling object-plane scanners, such as the Multi-Spectral Scanner and the Thematic Mapper, the average system point spread function is separable since the acquisition of data for these systems in inherently bi-directional (Friedmann, 1981). Each of the two components of the system point spread function then corresponds to the system line spread function h_{li} in the corresponding direction i, and the average point spread function is given by

$$h_a(x,y;s,t) = \int_{-\infty}^{\infty} h_{l1}(x;s)ds \int_{-\infty}^{\infty} h_{l2}(y;t)dt \tag{5.10}$$

The separable version of the Fourier transform of Eq. (5.10) has been implemented in modeling the transfer functions for the Landsat imaging systems. The two-dimensional Wiener filter has also been treated as a separable filter given by

$$M_1(u) = \frac{1}{H_{a1}(u)} \frac{|H_{a1}(u)|^2}{|H_{a1}(u)|^2 + S_{n1}(u)/S_{f1}f(u)} \tag{5.11}$$

and

$$M_2(v) = \frac{1}{H_{a2}(v)} \frac{|H_{a2}(v)|^2}{|H_{a2}(v)|^2 + S_{n2}(v)/S_{f2}f(v)} \tag{5.12}$$

This is only an approximation, since although $H_a(u,v)$ is separable, that is, $H_a(u,v) = H_{a1}(u)H_{a2}(v)$, and assuming $S(u,v)$ is separable, it does not follow that $M(u,v)$ is separable.

It is easy to show that the Wiener filter of Eq. (5.8) is not separable, even when those functions which make it up are separable: Assume that $M(u,v)$

is separable, $M(u,v) = M_1(u)M_2(v)$. Then

$$
\begin{aligned}
&M_1(u)M_2(v) \\
&\quad = \frac{H^*_{a1}(u)}{H_{a1}(u)H^*_{a1}(u) + S_{n1}(u)/S_{f1}(u)} \frac{H^*_{a2}(v)}{H_{a2}(v)H^*_{a2}(v) + S_{n2}(v)/S_{f1}(v)} \\
&\quad = \frac{H^*_{a1}(u)H^*_{a2}(v)}{|H_{a1}(u)|^2|H_{a2}(v)|^2 + (S_{n1}(u)S_{n2}(v))/(S_{f1}(u)S_{f2}(v)) + P(u,v)} \\
&\quad = \frac{1}{H_a(u,v)} \frac{|H_a(u,v)|^2}{|H_a(u,v)|^2 + S_n(u,v)/S_f(u,v) + P(u,v)}
\end{aligned}
$$

where the "cross terms" $P(u,v)$ are given by

$$P(u,v) = |H_{a1}(u)|(S_{n2}(v)/S_{f2}(v)) + |H_{a2}(v)|(S_{n1}(u)/S_{f1}(u))$$

Thus, only if the value of these cross terms are negligible compared to $|H_a(u,v)|^2 + S_n(u,v)/S_f(u,v)$, can the two-dimensional Wiener filter be considered (approximately) separable. Fortunately, the signal-to-noise ratios typically encountered in Landsat data are high enough so that this is indeed the case. In particular,

$$
\begin{aligned}
|H_a(u,v)|^2 + \frac{S_n(u,v)}{S_f(u,v)} &= |H_{a1}(u)|^2|H_{a2}(v)|^2 + \frac{S_{n1}(u)}{S_{f1}(u)} \frac{S_{n2}(v)}{S_{f2}(v)} \\
&> |H_{a1}(u)|\frac{S_{n2}(v)}{S_{f2}(v)} + |H_{a2}(v)|\frac{S_{n1}(u)}{S_{f1}(u)} \\
\iff &\left(|H_{a1}(u)|^2 - \frac{S_{n1}(u)}{S_{f1}(u)}\right)\left(|H_{a2}(v)|^2 - \frac{S_{n2}(v)}{S_{f2}(v)}\right) > 0 \\
\iff &|H_{ai}(w)|^2 > \frac{S_{ni}(w)}{S_{fi}(w)} \quad \text{where } i = 1 \text{ or } 2,\ w = u \text{ or } v \qquad (5.13)
\end{aligned}
$$

For the signal-to-noise ratios typical of Landsat imagery and for values of u and v less than about 0.5 cycles, Eq. (5.13) does indeed hold, since $|H_{ai}(u)|^2$ does not fall below a value of about 0.1 and $S_{ni}(w)/S_{fi}(w)$ is usually about 0.01.

Comments

It should be pointed out that because one can never (even theoretically) restore the continuous scene because of aliasing, some of the information is forever lost. The goal, then, is to boost those frequencies which correspond to fine detail which have been suppressed by the low pass filtering effect of the system transfer function and thus sharpen fine details in the image. Since the effect of sampling is to further degrade or blur the image, adding sampling to the system transfer function in the Wiener filter should sharpen that detail even more. The inherent spatial variance of the system response is not removed by a spatially invariant Wiener filter, however, and remains in the restored image.

CHAPTER 6

COMPARISON TO PREVIOUS WORK

As described in Chapter 4, when it is required that an image be geometrically corrected, it is usually necessary to interpolate pixel intensity (or gray level) values for positions intermediate to those for which the data is given. At each line and sample coordinate of the corrected image, the location in the original (input) image from which the corrected (output) image pixel value will be extracted is given by a set of transformation equations. The coordinates of the pixel location in the input image found in this way will not, in general, correspond to actual pixel locations in the input image. Thus a set of interpolation equations must be developed for determining the pixel intensity value to be assigned to the "new" pixel in the output image.

As pointed out by several authors, among them Dye, (1975 and 1981), many interpolation rules are equivalent to an additional convolution which reduces the spatial resolution of the geometrically corrected image. To reduce such effects, Dye combines the system point spread function and the resampling point spread function into a synthesized point spread function which approximates an ideal (desired) point spread function in a least squares sense. Dye's synthesized point spread function has been successfully implemented at the Environmental Research Institute of Michigan (ERIM) in Ann Arbor. Dye's approach will be explained (since work for ERIM is proprietary a detailed description of Dye's work was never published) as will the relation between his approach and the Wiener filter.

Description of Dye's Approach

Consider the following image model,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{6.1}$$

where $\mathbf{x}$ is a vector representing the (original) scene radiance, $\mathbf{A}$ is a matrix describing the system point spread function at each pixel, $\mathbf{n}$ represents additive signal independent noise and $\mathbf{y}$ is the resulting data (image) vector. All of these vectors can be thought of as samples from an ensemble of random vectors having the following properties:

$$E[\mathbf{n}] = 0, \quad E[\mathbf{x}\mathbf{n}^{\mathbf{T}}] = E[\mathbf{n}\mathbf{x}^{\mathbf{T}}] = 0$$

so that the noise $\mathbf{n}$ has zero mean and the scene $\mathbf{x}$ and noise $\mathbf{n}$ are assumed to be uncorrelated. Scene and noise correlation matrices will be defined as follows:

$$E[\mathbf{x}\mathbf{x}^{\mathbf{T}}] = \mathbf{R}_{\mathbf{x}}, \quad E[\mathbf{n}\mathbf{n}^{\mathbf{T}}] = \mathbf{R}_{\mathbf{n}}$$

Now consider applying a resampling process with the desired result of estimating the data values that might have been attained from a synthesized point spread function positioned at the desired output pixel locations. The result of this resampling process is

$$\mathbf{z} = \mathbf{B}\mathbf{x} \tag{6.2}$$

where $\mathbf{B}$ is a matrix representing the synthesized point spread function positioned correctly on the desired output grid. The exact nature of $\mathbf{B}$ can be chosen in such a way as to alter the spatial resolution of the output data, i. e. the synthesized point spread function may be designed to be different from the original system point spread function. For restoration, $\mathbf{B} = \mathbf{I}$, thus Dye's approach is a more general approach than simple restoration. The derivation will be continued in this more general way, and $\mathbf{B}$ set equal to $\mathbf{I}$ at the end for comparison to the Wiener filter.

Now we estimate the desired result $\hat{\mathbf{z}}$ by an operator $\mathbf{C}$ on the data vector $\mathbf{y}$,

$$\hat{\mathbf{z}} = \mathbf{C}\mathbf{y}$$

$$\text{or} \quad \hat{\mathbf{z}} = \mathbf{C}\mathbf{A}\mathbf{x} + \mathbf{C}\mathbf{n} \tag{6.3}$$

We want to choose $\mathbf{C}$ so as to minimize the mean squared error,

$$\underset{\mathbf{C}}{\text{minimize}} \quad E[(\mathbf{z} - \hat{\mathbf{z}})^{\mathbf{T}}(\mathbf{z} - \hat{\mathbf{z}})] \tag{6.4}$$

If we let $\mathbf{e} = \mathbf{z} - \hat{\mathbf{z}}$ then

$$E[\mathbf{e}^{\mathbf{T}}\mathbf{e}] = E[e_1^2 + e_2^2 + \ldots + e_n^2]$$

But

$$E[\mathbf{e}\mathbf{e}^{\mathbf{T}}] = \begin{pmatrix} E[e_1^2] & E[e_1e_2] & \ldots & E[e_1e_n] \\ E[e_2e_1] & E[e_2^2] & \ldots & E[e_2e_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[e_ne_1] & E[e_ne_2] & \ldots & E[e_n^2] \end{pmatrix}$$

Thus

$$\begin{aligned} E[\mathbf{e}^{\mathbf{T}}\mathbf{e}] &= E[e_1^2 + e_2^2 + \ldots + e_n^2] \\ &= E[e_1^2] + E[e_2^2] + \ldots + E[e_1^2] \\ &= \mathrm{Tr}(E[\mathbf{e}\mathbf{e}^{\mathbf{T}}]) = E[\mathrm{Tr}(\mathbf{e}\mathbf{e}^{\mathbf{T}})] \end{aligned} \tag{6.5}$$

where "Tr" indicates the trace of a matrix. The significance of Eq. (6.5) is that now we may minimize equivalently the mean squared error given in Eq. (6.4) or the expectation of the trace given in Eq. (6.5), whichever is easier. That is,

$$\underset{\mathbf{C}}{\text{minimize}} \quad E[\mathbf{e}^{\mathbf{T}}\mathbf{e}] \quad = \quad \underset{\mathbf{C}}{\text{minimize}} \quad E[\mathrm{Tr}(\mathbf{e}\mathbf{e}^{\mathbf{T}})] \tag{6.6}$$

But

$$\begin{aligned}
E[\mathrm{Tr}(\mathbf{ee^T})] &= E[\mathrm{Tr}((\mathbf{z}-\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{\mathbf{T}})] \\
&= E[\mathrm{Tr}((\mathbf{Bx}-\mathbf{CAx}-\mathbf{Cn})(\mathbf{Bx}-\mathbf{CAx}-\mathbf{Cn})^{\mathbf{T}})] \\
&= E[\mathrm{Tr}(\mathbf{Bxx^TB}-\mathbf{Bxx^TA^TC^T}-\mathbf{Bxn^TC^T}-\mathbf{CAxx^TB^T} \\
&+ \mathbf{CAxx^TA^TC^T}+\mathbf{CAxn^TC^T}-\mathbf{Cnx^TB^T} \\
&+ \mathbf{Cnx^TA^TC^T}+\mathbf{Cnn^TC^T})]
\end{aligned}$$

Using $E[\mathrm{Tr}(\mathbf{M})] = \mathrm{Tr}(E[\mathbf{M}])$ where $\mathbf{M}$ is any matrix and $E[\mathbf{xn^T}] = 0 = E[\mathbf{nx^T}]$ yields

$$\begin{aligned}
E[\mathrm{Tr}(\mathbf{ee^T})] &= \mathrm{Tr}(E[\mathbf{Bxx^TB}]-E[\mathbf{Bxx^TA^TC^T}] \\
&- E[\mathbf{CAxx^TB^T}]+E[\mathbf{CAxx^TA^TC^T}]+E[\mathbf{Cnn^TC^T}]) \\
&= \mathrm{Tr}(E[\mathbf{BR_xB^T}-\mathbf{BR_xA^TC^T} \\
&- \mathbf{CAR_xB^T}+\mathbf{CAR_xA^TC^T}+\mathbf{CR_nC^T}]) \qquad (6.7)
\end{aligned}$$

since $E[\mathbf{Mxx^TN}] = \mathbf{ME}[\mathbf{xx^T}]\mathbf{N}$ for any matrices $\mathbf{M}$ and $\mathbf{N}$ and by definition $E[\mathbf{xx^T}] = \mathbf{R_x}$ and $E[\mathbf{nn^T}] = \mathbf{R_n}$.

Now we want to minimize Eq. (6.7) with respect to $\mathbf{C}$. Taking the derivative of this matrix equation (Graybill, 1983) with respect to $\mathbf{C}$ and setting the result equal to zero,

$$-2\mathbf{BR_xA^T}+\mathbf{C}(\mathbf{AR_xA^T}+\mathbf{R_n}) = 0$$

where the fact that $\mathbf{R_x}$ and $\mathbf{R_n}$ are symmetric, i. e. $\mathbf{R_x^T} = \mathbf{R_x}$ and $\mathbf{R_n^T} = \mathbf{R_n}$, has been used. Solving for $\mathbf{C}$ gives

$$\mathbf{C} = \mathbf{BR_xA^T}(\mathbf{AR_xA^T}+\mathbf{R_n})^{-1} \qquad (6.8)$$

the expression for $\mathbf{C}$ which gives the best estimate in the mean squared sense for approximating the original scene.

Relation to Wiener Filter

It has been shown that in order to minimize the mean squared error between a desired image $\mathbf{z} = \mathbf{Bx}$ and its estimate $\hat{\mathbf{z}}$, where we have only a noisy image of the scene $\mathbf{y} = \mathbf{Ax} + \mathbf{n}$, the required transformation matrix $\mathbf{C}$ which must be applied to the image $\mathbf{y}$ is given by

$$\mathbf{C} = \mathbf{B}\mathbf{R}_{\mathbf{x}}\mathbf{A}^{\mathbf{T}}(\mathbf{A}\mathbf{R}_{\mathbf{x}}\mathbf{A}^{\mathbf{T}} + \mathbf{R}_{\mathbf{n}})^{-1}$$

where $\mathbf{A}$ is the imaging system point spread function, $\mathbf{B}$ the desired point spread function, $\mathbf{R}_{\mathbf{x}}$ the scene autocorrelation matrix and $\mathbf{R}_{\mathbf{n}}$ the noise autocorrelation matrix. When $\mathbf{B} = \mathbf{I}$, the identity matrix, then the transformation is the usual restoration, which is to find the best estimate of the original scene $\mathbf{x}$.

Several assumptions are usually made at this point before implementation of the restoration is attempted. In particular, it is usually assumed that pixels in an image possess correlation only over some finite distance and that interpixel correlation depends only on the distance between two pixels and not on the actual locations of the pixels, i.e. stationarity (Andrews and Hunt, 1977). These assumptions result in a symmetric Toeplitz matrix (Grenander and Szegö, 1958), a symmetric band matrix for which all the elements on the principal diagonal are identical and all the elements on the ith lower subdiagonal are equal to the corresponding elements on ith upper subdiagonal, i. e. if the elements of the matrix are t_{ij} then $t_{mn} = t_{pq}$ whenever $m - n = p - q$. Assuming this structure for the correlation matrices $\mathbf{R}_{\mathbf{x}}$ and $\mathbf{R}_{\mathbf{n}}$ is equivalent to assuming a stationary variance for the image (to be discussed in Chapter 7). Frequently $\mathbf{R}_{\mathbf{x}}$ and $\mathbf{R}_{\mathbf{n}}$ are approximated by scalar matrices such as $\sigma_{\mathbf{x}}^2\mathbf{I}$ and $\sigma_{\mathbf{n}}^2\mathbf{I}$, where σ^2 represents variance (see for example Dye, 1975 and 1982). Such an approximation is not optimal (Trussell and Hunt, 1978), however, and a better restoration can be expected when more structure is introduced (again, see Chapter 7).

The major disadvantage of attempting to implement Eq. (6.8) directly is the necessity of a matrix inversion on a matrix which is frequently ill-conditioned, and perhaps even singular (Andrews and Hunt, 1977). That is, small perturbations of the elements in the original matrix may lead to large changes in its inverse. If the matrix to be inverted were diagonal, however, its inverse would be simply another diagonal matrix whose elements were the reciprocals of the diagonal elements of the original matrix. (Diagonalizing an ill-conditioned matrix doesn't make it well-conditioned, but it does accentuate exactly where the problems occur, thus making them more manageable.) Thus, we want to diagonalize $\mathbf{A}$, $\mathbf{B}$, $\mathbf{R_x}$ and $\mathbf{R_n}$ since sums, products and transposes of diagonal matrices are also diagonal.

If we make the assumptions necessary for $\mathbf{A}$, $\mathbf{B}$, $\mathbf{R_x}$ and $\mathbf{R_n}$ to be block circulant, (a circulant matrix is one for which each row is a one element shift of the previous row with "wraparound", and the first row is the shift of the last) then these matrices can easily be diagonalized by applying to them the discrete Fourier transform. In what follows it is shown that these matrices can be made block-circulant.

Circulant Point Spread Function Matrices

For the sake of illustration, consider a one-dimensional convolution of two discrete functions given by $h(0), h(1), h(2)$ and $f(0), f(1), f(2), f(3)$. For discrete convolution, one function is inverted and the second multiplied by it with the products summed as the second is moved over the first. Discrete convolution of the sampled functions given in this example yields

$$
\begin{aligned}
g(0) &= h(0)f(0) \\
g(1) &= h(0)f(1) + h(1)f(0) \\
g(2) &= h(0)f(2) + h(1)f(1) + h(2)f(0) \\
g(3) &= h(0)f(3) + h(1)f(2) + h(2)f(1) \\
g(4) &= h(1)f(3) + h(2)f(2) \\
g(5) &= h(2)f(3)
\end{aligned}
$$

or, in matrix notation,

$$
\begin{pmatrix} g(0) \\ g(1) \\ g(2) \\ g(3) \\ g(4) \\ g(5) \end{pmatrix} = \begin{pmatrix} h(0) & 0 & 0 & 0 \\ h(1) & h(0) & 0 & 0 \\ h(2) & h(1) & h(0) & 0 \\ 0 & h(2) & h(1) & h(0) \\ 0 & 0 & h(2) & h(1) \\ 0 & 0 & 0 & h(2) \end{pmatrix} \begin{pmatrix} f(0) \\ f(1) \\ f(2) \\ f(3) \end{pmatrix}
$$

This matrix can be extended to be made circulant as follows:

$$
\begin{pmatrix} g(0) \\ g(1) \\ g(2) \\ g(3) \\ g(4) \\ g(5) \end{pmatrix} = \begin{pmatrix} h(2) & h(1) & h(0) & 0 & 0 & 0 \\ 0 & h(2) & h(1) & h(0) & 0 & 0 \\ 0 & 0 & h(2) & h(1) & h(0) & 0 \\ 0 & 0 & 0 & h(2) & h(1) & h(0) \\ h(0) & 0 & 0 & 0 & h(2) & h(1) \\ h(1) & h(0) & 0 & 0 & 0 & h(2) \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ f(0) \\ f(1) \\ f(2) \\ f(3) \end{pmatrix}
$$

It is easy to see how this can be generalized to any sized discrete functions h and f for one-dimensional convolution.

For two-dimensional discrete convolution, $\mathbf{y} = \mathbf{H}\mathbf{x}$, the usual case with images, if the object and image are written as column vectors $\mathbf{x}$ and $\mathbf{y}$ as described in Chapter 2, then we may write

$$\begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{J-1} \\ \mathbf{y}_J \\ \vdots \\ \mathbf{y}_{J+M-1} \end{pmatrix} \begin{pmatrix} [\mathbf{H}_0] & 0 & \cdots & \cdots & 0 \\ [\mathbf{H}_1] & [\mathbf{H}_0] & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ [\mathbf{H}_{J-1}] & \ddots & \ddots & [\mathbf{H}_0] & 0 \\ 0 & [\mathbf{H}_{J-1}] & \ddots & \ddots & [\mathbf{H}_0] \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & [\mathbf{H}_{J-1}] \end{pmatrix} = \begin{pmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{J-1} \\ \mathbf{x}_J \\ \vdots \\ \mathbf{x}_{M-1} \end{pmatrix}$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ are the ith columns of the scene and image, respectively, stacked as columns to form the vectors $\mathbf{x}$ and $\mathbf{y}$, and $[\mathbf{H}_i]$ is a band matrix similar to the one-dimensional $\mathbf{H}$ matrix discussed above. Each $[\mathbf{H}_i]$ may be extended to be circulant, as in the one-dimensional example (Andrews and Hunt, 1977). Thus the point spread function for two-dimensional convolution may also be made to be block circulant.

Circulant Correlation Matrices

The correlation matrices $\mathbf{R}_\mathbf{x}$ and $\mathbf{R}_\mathbf{n}$ are symmetric for stationary random processes since $E[\mathbf{x}_i\mathbf{x}_j^\mathbf{T}] = E[\mathbf{x}_j\mathbf{x}_i^\mathbf{T}]$ where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the ith and jth column partitions of the ordered vector $\mathbf{x}$. In particular, $\mathbf{R}_\mathbf{x}$ is of the form

$$\mathbf{R}_\mathbf{x} = \begin{pmatrix} [\mathbf{R}_{0,0}] & [\mathbf{R}_{0,1}] & \cdots & [\mathbf{R}_{0,N-1}] \\ [\mathbf{R}_{1,0}] & [\mathbf{R}_{1,1}] & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ [\mathbf{R}_{N-1,0}] & \cdots & \cdots & [\mathbf{R}_{N-1,N-1}] \end{pmatrix}$$

where $[\mathbf{R}_{ij}] = [\mathbf{R}_{ji}]$ and $[\mathbf{R}_{ii}]$ is diagonal. Blocks off the diagonal describe the correlation of rows in the image matrix separated by a distance of $|i-j|$ pixels.

Covariance matrices $\mathbf{C}_\mathbf{x}$ and $\mathbf{C}_\mathbf{n}$ may be defined in an analogous fashion by removing the means $\bar{\mathbf{x}}$:

$$\mathbf{C}_\mathbf{x} = E[(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^\mathbf{T}]$$

Often, pixels in an image possess correlation only over some finite distance. Then

$$\mathbf{C_x} = \begin{pmatrix} [\mathbf{C}_{0,0}] & \cdots & \cdots & \cdots & [\mathbf{C}_{0,d-1}] & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ [\mathbf{C}_{d-1,0}] & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & [\mathbf{C}_{N-d,N-1}] \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & [\mathbf{C}_{N-1,N-1+d}] & \cdots & \cdots & \cdots & [\mathbf{C}_{N-1,N-1}] \end{pmatrix}$$

where each component matrix also has band structure. If, furthermore, the interpixel correlation depends only on the distance between two pixels and not on the actual location of the pixels (stationary second-order statistics) then $\mathbf{C_x}$ is block Toeplitz. Such matrices may easily be made circulant (Andrews and Hunt, 1977).

However, approximating $\mathbf{C_x}$ by a circulant matrix is not the same as approximating $\mathbf{R_x}$ by a circulant matrix since

$$\mathbf{R_x} = \mathbf{C_x} + [\mathbf{x}\mathbf{x}^{\mathbf{T}}] \tag{6.9}$$

Three cases need to be considered: Case I: A zero mean process. Since $\mathbf{x}$ is a vector of samples from an scene, positivity makes this case an impossibility and so it will not be considered further. Case II: A constant mean process. If it is assumed as part of the initial model that the underlying random process of the scene is stationary, then this case applies and

$$\mathbf{R_x} = \mathbf{C_x} + \text{additive constant} \tag{6.10}$$

Case III: The mean is not a constant, that is, the scene model is assumed to be nonstationary. This is the case where an adaptive filter would be used and this case cannot be analyzed using the methods described here.

For this analysis it will be assumed that the scene is a sample from a stationary (constant mean) process. Let

$$\mathbf{R_x} = \mathbf{C_x} \tag{6.11}$$

and perform the operations as though the data were zero mean to which a constant bias has been added. This bias must be accounted for later, but can be ignored for the purposes of the immediate derivation.

Note that this discussion is only of concern for $\mathbf{R_x}$. Since the noise is assumed to be zero mean, then $\mathbf{R_n} = \mathbf{C_n}$ with no difficulty.

Thus it has been shown that $\mathbf{R_n} = \mathbf{C_n}$ can be made block circulant and, with certain assumptions, so can $\mathbf{R_x}(= \mathbf{C_x})$.

Diagonalization of Circulant Matrices

It will now be shown how circulant matrices can be diagonalized using the discrete Fourier transform.

Consider a general circulant matrix $\mathbf{C}$ given by the following:

$$\mathbf{C} = \begin{pmatrix} c(0) & c(1) & c(2) & \dots & c(N-1) \\ c(N-1) & c(0) & c(1) & \dots & c(N-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c(1) & c(2) & c(3) & \dots & c(0) \end{pmatrix}$$

Let $W = e^{(2\pi i/N)}$, then

$$W^{kN} = 1 \qquad \text{for} \quad k = 0, 1, 2, \dots, N-1 \tag{6.12}$$

Consider the equation which results from multiplying the elements of the first row of the circulant matrix $\mathbf{C}$ by powers of W and adding them to get,

$$\lambda(k) = c(0) + c(1)W^k + c(2)W^{2k} + \dots + c(N-1)W^{(N-1)k}$$

By repeatedly multiplying this equation by W^k, using Eq. (6.12) and rearranging, we get the following sequence of equations

$$\begin{aligned}
\lambda(k)W^k &= c(N-1) + c(0)W^k + c(1)W^{2k} + \ldots + c(N-2)W^{(N-1)k} \\
\lambda(k)W^{2k} &= c(N-2) + c(N-1)W^k + c(0)W^{2k} + \ldots + c(N-3)W^{(N-1)k} \\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\
\lambda(k)W^{(N-1)k} &= c(1) + c(2)W^k + c(3)W^{2k} + \ldots + c(0)W^{(N-1)k}
\end{aligned}$$

If we now represent these W^k multipliers in a vector

$$\mathbf{w}(k) = (1 \quad W^k \quad W^{2k} \quad \ldots \quad W^{(N-1)k})^{\mathbf{T}}$$

then the above set of equations can be compactly written in matrix form as

$$\lambda(k)\mathbf{w}(k) = \mathbf{C}\mathbf{w}(k) \tag{6.13}$$

Thus the eigenvectors of $\mathbf{C}$ are $\mathbf{w}(0), \mathbf{w}(1), \ldots, \mathbf{w}(N-1)$ and the corresponding eigenvalues are $\lambda(0), \lambda(1), \ldots, \lambda(N-1)$. The eigenvectors may be written in a matrix $\mathbf{W}$ whose columns are the eigenvectors, i. e.

$$\mathbf{W} = (\mathbf{w}(0) \quad \mathbf{w}(1) \quad \ldots \quad \mathbf{w}(N-1)) \tag{6.14}$$

and the eigenvalues as a diagonal matrix whose elements are the eigenvalues

$$\Lambda = \begin{pmatrix} \lambda(0) & \ldots & \ldots & 0 \\ 0 & \lambda(1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & \lambda(N-1) \end{pmatrix} \tag{6.15}$$

Since the N-by-N matrix $\mathbf{C}$ has N linearly independent eigenvectors (Hunt, 1971), or alternatively N distinct eigenvalues, $\mathbf{C}$ can be diagonalized by the use of a similarity transformation whose columns are the eigenvectors of $\mathbf{C}$ (Strang,

1980). The resulting diagonal matrix will be the matrix whose diagonal elements are the eigenvalues of **C**. Thus

$$\mathbf{W}^{-1}\mathbf{C}\mathbf{W} = \mathbf{\Lambda} \tag{6.16}$$

The matrix $\mathbf{W}^{-1}$ represents the discrete Fourier transform.

The more familiar form of the discrete Fourier transform is its series form:

$$F(k) = \sum_{m=0}^{N-1} f(m)e^{-(2\pi imk/N)} \qquad k = 0, 1, \ldots, N-1 \tag{6.17}$$

Writing this out for each value of k yields

$$\begin{array}{lllllll}
F(0) & =f(0)+ & f(1) & + & \ldots & + & f(N-1) \\
F(1) & =f(0)+ & f(1)e^{-2\pi i/N} & + & \ldots & + & f(N-1)e^{-2(N-1)\pi i/N} \\
F(2) & =f(0)+ & f(1)e^{-4\pi i/N} & + & \ldots & + & f(N-1)e^{-4(N-1)\pi i/N} \\
\vdots & \vdots & \vdots & & & & \vdots \\
F(N-1) & =f(0)+ & f(1)e^{-2\pi i(N-1)/N} & + & \ldots & + & f(N-1)e^{-2\pi i(N-1)^2/N}
\end{array}$$

Writing this set of equations in matrix form yields

$$\begin{pmatrix} F(0) \\ F(1) \\ F(2) \\ \vdots \\ F(N-1) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & W_0 & W_0^2 & \ldots & W_0^{N-1} \\ 1 & W_0^2 & W_0^4 & \ldots & W_0^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_0^{N-1} & W_O^{2(N-1)} & \ldots & W_0^{(N-1)} \end{pmatrix} \begin{pmatrix} f(0) \\ f(1) \\ f(2) \\ \vdots \\ f(N-1) \end{pmatrix}$$

where $W_0 = e^{-2\pi i/N}$. Denote this matrix $\mathbf{W}^{-1}$. Due to the periodicity of $(W_0)^k$,

$$\mathbf{W}^{-1}_{(k,l)} = W_0^{k\times l(\mathrm{mod}N)} \tag{6.18}$$

with $k, l = 0, 1, 2, \ldots, N-1$ where $\mathbf{W}^{-1}_{kl}$ denotes the k, l^{th} element of $\mathbf{W}^{-1}$. For

example, with $N = 8$

$$\mathbf{W}^{-1} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & W^1 & W^2 & W^3 & W^4 & W^5 & W^6 & W^7 \\ 1 & W^2 & W^4 & W^6 & 1 & W^2 & W^4 & W^6 \\ 1 & W^3 & W^6 & W^1 & W^4 & W^7 & W^2 & W^5 \\ 1 & W^4 & 1 & W^4 & 1 & W^4 & 1 & W^4 \\ 1 & W^5 & W^2 & W^7 & W^4 & W^1 & W^6 & W^3 \\ 1 & W^6 & W^4 & W^2 & 1 & W^6 & W^4 & W^2 \\ 1 & W^7 & W^6 & W^5 & W^4 & W^3 & W^2 & W^1 \end{pmatrix}$$

The matrix $\mathbf{W}^{-1}$ is unitary, that is the inner product of any one column (row) with the conjugate of any other column (row) is zero unless the two columns (rows) are identical.

$$\sum_{m=0}^{N-1} \mathbf{W}_{ms}^{-1} (\mathbf{W}_{mt}^{-1})^* = \begin{cases} N & s = t \\ 0 & s \neq t \end{cases}$$

So its inverse is $(\mathbf{W}^{-1})^{-1} = ((\mathbf{W}^{-1})^*)^{\mathbf{T}}/N$.

Thus a matrix has been found which diagonalizes a circulant matrix and that matrix is the discrete Fourier transform. Furthermore, it has been shown that this matrix is unitary so that its inverse is easy to find.

Derivation of the Wiener Filter

We now have at our disposal all the mathematical machinery necessary to derive the Wiener filter from the least-squares restoration method given in Eq. (5.8). The procedure is as follows:

From $\mathbf{C} = \mathbf{B}\mathbf{R}_{\mathbf{x}}\mathbf{A}^{\mathbf{T}}(\mathbf{A}\mathbf{R}_{\mathbf{x}}\mathbf{A}^{\mathbf{T}} + \mathbf{R}_{\mathbf{n}})^{-1}$ do the following:

Step 1. Make all relevant matrices block circulant using the methods outlined above. Step 2. Diagonalize each block circulant matrix using the discrete Fourier transform. In particular, for $\mathbf{B}$, $\mathbf{R}_{\mathbf{x}}$, $\mathbf{A}$ and $\mathbf{R}_{\mathbf{n}}$, now block circulant, let

$$\mathbf{A} = \mathbf{W}\mathbf{D}_{\mathbf{A}}\mathbf{W}^{-1}, \quad \mathbf{B} = \mathbf{W}\mathbf{D}_{\mathbf{B}}\mathbf{W}^{-1},$$

$$\mathbf{R}_{\mathbf{x}} = \mathbf{W}\mathbf{D}_{\mathbf{x}}\mathbf{W}^{-1}, \quad \mathbf{R}_{\mathbf{n}} = \mathbf{W}\mathbf{D}_{\mathbf{n}}\mathbf{W}^{-1}$$

Substituting into Eq. (6.8) with $\mathbf{B} = \mathbf{I}$ (restoration) yields

$$\mathbf{F} = \mathbf{W}\mathbf{D}_\mathbf{x}(\mathbf{D}_\mathbf{A}^*)^\mathbf{T}(\mathbf{D}_\mathbf{A}\mathbf{D}_\mathbf{x}(\mathbf{D}_\mathbf{A}^*)^\mathbf{T} + \mathbf{D}_\mathbf{n})^{-1}\mathbf{W}^{-1}$$

where for the complex matrices the transpose becomes the conjugate transpose. The result has been denoted $\mathbf{F}$, for "filter". The diagonal elements of $\mathbf{D_A}$, $\mathbf{D_x}$ and $\mathbf{D_n}$ are the discrete Fourier coefficients of the corresponding circulant matrices. They can also be written in series notation as

$$H(u,v) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} h(m,n)e^{-(2\pi i(mu+nv)/NM)}$$

$$S_x(u,v) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} c_x(m,n)e^{-(2\pi i(mu+nv)/NM)}$$

$$S_n(u,v) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} c_n(m,n)e^{-(2\pi i(mu+nv)/NM)}$$

where $H(u,v)$ is the system transfer function (the Fourier transform of the point spread function), $S_x(u,v)$ is the power spectrum of the scene and $S_n(u,v)$ is the power spectrum of the noise.

Since the Fourier transform matrix $\mathbf{W}^{-1}$ is unitary and the point spread function matrix is real, $\mathbf{A}^T = (\mathbf{A}^*)^\mathbf{T}$ so that

$$\mathbf{A} = \mathbf{W}\mathbf{D}_\mathbf{A}\mathbf{W}^{-1} \Longrightarrow$$

$$(\mathbf{A}^*)^\mathbf{T} = ([\mathbf{W}\mathbf{D}_\mathbf{A}\mathbf{W}^{-1}]^*)^\mathbf{T} = ([\mathbf{W}^{-1}]^*)^\mathbf{T}(\mathbf{D}_\mathbf{A}{}^*)^\mathbf{T}(\mathbf{W}^*)^\mathbf{T} = \mathbf{W}(\mathbf{D}_\mathbf{A}{}^*)^\mathbf{T}\mathbf{W}^{-1}.$$

Thus we have

$$F(u,v) = \frac{S_x(u,v)H^*(u,v)}{H(u,v)S_x(u,v)H^*(u,v) + S_n(u,v)}$$

or

$$F(u,v) = \frac{1}{H(u,v)}\frac{|H(u,v)|^2}{|H(u,v)|^2 + S_n(u,v)/S_x(u,v)}$$

which is just the Wiener filter.

Thus, by diagonalizing the matrix **C** of Eq. (6.8) (Dye, 1977 and 1982) using the discrete Fourier transform, one derives the Wiener filter.

Discussion

The Wiener filter approach is equivalent mathematically to Dye's approach, Eq. (6.8), with Dye's synthesized point spread function **B** the identity matrix (restoration). The Wiener filter approach, however, avoids many of the pitfalls of Dye's approach.

Computationally, the matrix operations of Eq. (6.8) are not required; in particular, many of the issues involving matrix inversion are avoided. Using the fast Fourier transform one need only compute the discrete Fourier transforms of the image and the components of the restoration matrix **C**, invert the diagonalized matrix to find the "filter", multiply the Fourier transform of the image by the filter to obtain the Fourier transform of the estimate of the scene $\hat{\mathbf{z}}$, then compute the inverse Fourier transform of that product.

Using this approach, a solution can be found for which there is no associated ill-conditioned behavior. If $H(u,v)$ is very small or even zero for some values of u and v, the denominator will never be smaller than $S_n(u,v)/S_x(u,v)$. Thus, as long as $S_n(u,v)/S_x(u,v)$ is nonzero, even though the point spread function matrix **A** may be singular, a restored image can still be generated using this "frequency domain" approach.

If the system is noiseless, $S_n(u,v) = 0$, then the Wiener filter reduces to the inverse filter. If, on the other hand, the signal is absent, $S_x(u,v) = 0$, then the Wiener filter equals zero and the restored image is zero. This is reasonable since one does not expect to restore a nonexistant image.

CHAPTER 7

RESTORATION FILTER DETERMINATION

Implementation of the Wiener filter given by Eq. (2.36) or Eq. (5.8) requires knowledge of the scene and noise power spectra $S_f(u,v)$ and $S_n(u,v)$. Frequently the ratio $S_n(u,v)/S_f(u,v)$ is taken to be a constant, (Rosenfeld and Kak, 1982; Andrews and Hunt, 1977). However, better results can be expected if more information about signal and noise statistics is included. Toward this end models have been proposed for these quantities and their parameters measured for Multi-Spectral Scanner images.

The use of the power spectrum in the Wiener filter is valid for a stationary random process. However, studies of images indicate that this assumption is not valid for images of natural scenes (Trussel and Hunt, 1978; Hunt, 1980). The statistics of the scene may vary widely over the image. To account for this, one would need an adaptive filter, one which changed its parameters with the statistical variations encountered in the image. However, for reasons of mathematical tractability it is desirable to employ stationary image models. Such models work quite well in image restoration (Hunt, 1980), the major drawbacks being the possibility of under-restoring high frequency information (such as edges) and over-restoring low frequency regions (resulting in noise enhancement). The restoring filter parameters are chosen to yield an acceptable tradeoff between these effects.

Scene Autocorrelation Model

A simple mathematical model of an image might consist of a sequence of rectangular pulses of randomly determined height (corresponding to pixel gray

levels) and random duration (corresponding to the sizes of objects in the image). For such a model it can be shown (Papoulis, 1984) that the corresponding autocorrelation function is of the form of a first order Markov model which is given by

$$R(x,y) = \sigma^2 e^{-\sqrt{a^2x^2+b^2y^2}} \tag{7.1}$$

where σ is the standard deviation of the image gray levels. Studies of image autocorrelation data have been performed for several types of images in conjunction with the development of methods for image data compression (Kretzmer, 1952; O'Neal, 1966; Huang, 1965) . These studies indicate that this is indeed an excellent model for a wide variety of scanned pictorial data (Franks, 1966; Habibi and Wintz, 1971). The first order Markov function is rotationally symmetric, but not separable. For reasons of computational efficiency and mathematical tractibility the Markov process is often assumed to be separable having the form

$$R(x,y) = \sigma^2 e^{-a|x|-b|y|} \tag{7.2}$$

where $1/a$ and $1/b$ are the data correlation half-widths in two orthogonal directions (e. g. Hunt, 1980; Franks, 1966). This is the model we have used. With this autocorrelation model the corresponding model power spectrum is

$$S(u,v) = \frac{4ab\sigma^2}{(a^2+(2\pi u)^2)(b^2+(2\pi v)^2)} \tag{7.3}$$

which is also separable.

To measure a and b several 32-by-32 scene "windows" were taken from a (geometrically uncorrected) Landsat Multi-Spectral Scanner Washington D.C. image (ID# 84039218152). These windows were chosen to have different scene homogeneity characteristics. The autocorrelation of each window was then found by computing digitally the inverse Fourier transform of the power spectrum of the window. The parameters a and b were determined by fitting the model of Eq. (7.2) to the data.

Using the separability of the system's point spread function, the 32-by-32 pixel window autocorrelation data were studied for the along scan and along track directions separately, similarly for each band. This resulted in eight values for the correlation half-width parameter, one for each band along scan and along track. To reduce the number of image dependent parameters in the Wiener filter, it was hoped that a single, average, correlation half-width parameter could be fixed for the entire image. Only the variance would then need to be specified for a particular image. Figures 9 and 10 indicate that this is reasonable; the Wiener filter is relatively insensitive to the choice of correlation length compared to its sensitivity to the variance. The autocorrelation models used for the along scan and along track directions in each band are, then, given by:

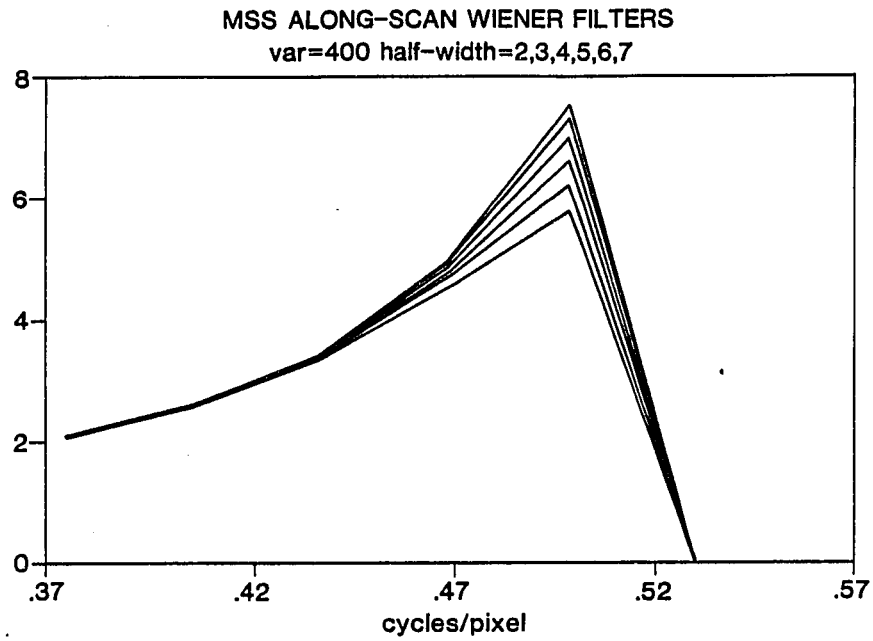
$$\begin{aligned} R_i(x) &= \sigma_i e^{-a_i|x|} + m_i \quad i = 1,2,3,4 \\ R_i(y) &= \sigma_i e^{-b_i|y|} + m_i \quad i = 1,2,3,4 \end{aligned} \tag{7.4}$$

where the variable x indicates the along scan direction, y the along track direction, the subscript i the band number, and m_i the local image grey level mean.

The correlation half-width parameters were found as follows. We consider the along scan direction only, the along track procedure being analogous. First, the mean was subtracted from the data, the result scaled, and the natural logarithm of the scaled result used in a linear least squares fit. The new data set is

$$w_k = \ln(s(d_k - m)) \tag{7.5}$$

where d is "raw" autocorrelation data, m the mean and $s = \sigma/(d_{max} - m)$. The scale factor s was used to fix the y-intercept of the linear fit, i. e. the image variance, since it is only the correlation half-width that is of interest. That is, we want to find only the one parameter a (or b). Taking the natural logarithm



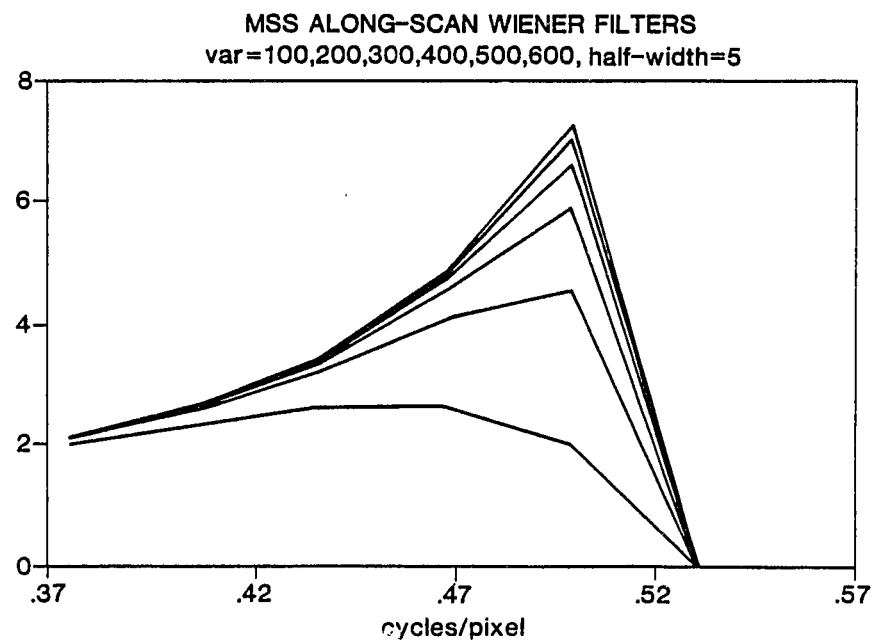


Figure 9. Sensitivity of the Wiener Filter to Parameter Changes for the Multi-Spectral Scanner.

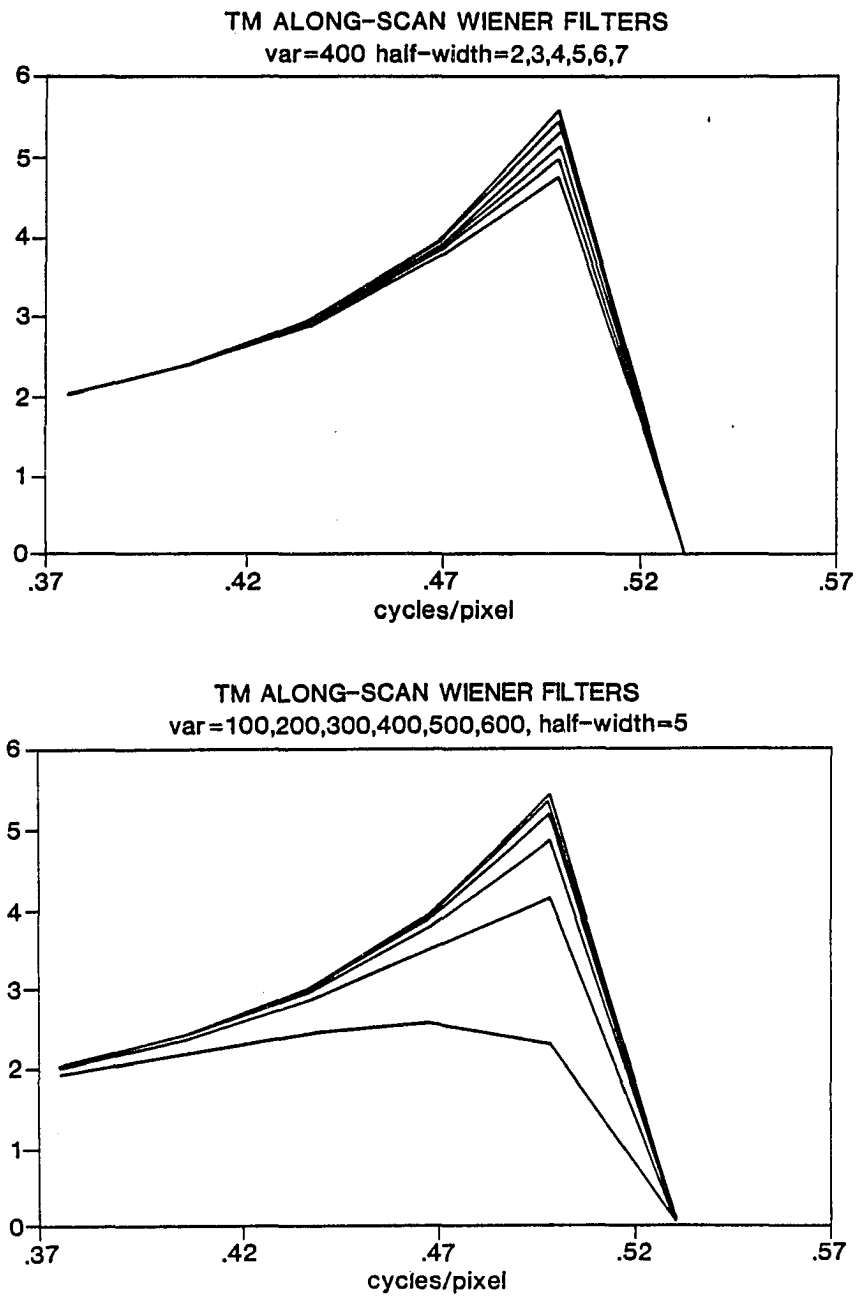


Figure 10. Sensitivity of the Wiener Filter to Parameter Changes for the Thematic Mapper.

of $(R(x) - m)$ and choosing $x \geq 0$ leaves

$$\ln(\sigma) - ax \tag{7.6}$$

We want to minimize the expected value of the sum of the squares between the new (scaled) data and this model, that is

$$\text{minimize} \quad E = \sum_{k=1}^{n} (w_k - \ln\sigma + ax_k)^2 \tag{7.7}$$

Taking the derivative of E with respect to a, setting it equal to zero and solving for a gives

$$a = \frac{\ln\sigma \sum_{k=1}^{n} x_k - \sum_{k=1}^{n} w_k x_k}{\sum_{k=1}^{n} x_k^2}$$

This was calculated for all the data sets giving the range of values shown in Table 1 of Figure 11. (As the variance (contrast) increased across the bands, the value inside the logarithm often was less than or equal to zero and that data had to be abandoned. This resulted in a loss of usable data for band 4 especially.)

The scaled data was plotted along with the linearized model and the residuals (differences between model and data) were also calculated and plotted. These residuals indicated that a better fit to the data might be made with a quadratic model. Thus a fit to a model of the form

$$R(x) = \sigma e^{-(a_1 x^2 + a_2 x)} \tag{7.9}$$

was attempted (similarly for the along scan direction.) This time the minimization problem is

$$\text{minimize} \quad E = \sum_{k+1}^{n} (w_k - \ln\sigma + a_1 x_k^2 + a_2 x_k)^2 \tag{7.10}$$

Table 1. Correlation Half-Widths for Test Windows

Band 1		Band 2		Band 3		Band 4	
scan	track	scan	track	scan	track	scan	track
.456	.238	.192		.220		.124	
.421	.743	.468		.264		.457	
.249	.515	.323		.309			
.183	.392	.178	.367	.307		.699	
.373	.136	.644					
.260	.452	.214	.335				
.477	.219	.444	.251	.507	.305		
.228		.255		.447			
	.271		.310				
.200		.200				.326	
	.705			.339	.329		.321

Table 2. Average Correlation Half-Widths in Pixels

	Band 1	Band 2	Band 3	Band 4
along scan	6.10	5.35	7.09	7.41
along track	4.29	4.33	5.40	4.37

Figure 11. Pixel Correlation Parameters

Again the constant term is dropped in order to match the y-intercept with the data. This time the derivatives of E with respect to both a_1 and a_2 are required, solving these for a_1 and a_2 results in the linear system

$$\begin{pmatrix} \sum x_k^4 & \sum x_k^3 \\ \sum x_k^3 & \sum x_k^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum (\ln \sigma - w_k) x_k^2 \\ \sum (\ln \sigma - w_k) x_k^2 \end{pmatrix}$$

This system was solved numerically for all the along scan band 1 test windows and the second order coefficients a_1 were found to be in the range 10^{-3} to 10^{-2}. It was felt that such a small contribution would not affect the restoration, so that refinement was abandoned.

Since it is desired to have a correlation half-width parameter which is the best fit over the whole image, that is for the average of the image statistics, the range of values found using the above procedure was then used as the range within which could be found the best parameter for the average of the data sets. The autocorrelation data was averaged together and the parameter which best fit this average was found. This time autocorrelation data, scaled as before to fix the y-intercept, was fit with the exponential model. No logarithms were taken and so there was no problem with nonpositive values appearing inside the logarithm. The choice of parameter was made based on the lowest root mean square error between the average data and the model. These results are shown in Table 2 of Figure 11. If one chooses to use one value of a for all the four bands along scan and one value of b for all bands along track, the average correlation half-widths would be

6.38 pixels or 370 meters along scan

4.56 pixels or 371 meters along track

These values apply to correlation half-widths of the Landsat Multi-Spectral Scanner images, which are degraded recordings of the scenes, and not to the scenes themselves. That is, the scene correlation half-width is not really

being measured, but rather the image correlation half-width. The image is the convolution of the scene with the system point spread function. So the autocorrelation function of the image is the convolution of the autocorrelation of the scene with the autocorrelation of the imaging system. That is, for complex-valued functions g, h and f, if $g(x,y) = h(x,y) * f(x,y)$, then $R_{gg}(x,y) = R_{ff}(x,y) * R_{hh}(x,y)$. This is easily proved: The power spectra corresponding to the correlations R_{gg}, R_{ff} and R_{hh} satisfy (Papoulis, 1984)

$$\begin{aligned} S_{gg}(u,v) &= S_{ff}(u,v)|H(u,v)|^2 \\ &= S_{ff}(u,v)H^*(u,v)H(u,v) \end{aligned}$$

Taking the Fourier transform of both sides and using the relation that $\mathcal{F}^{-1}[H^*(u,v)] = h^*(-x,-y)$ (Gaskill, 1978) yields

$$\begin{aligned} R_{gg}(x,y) &= R_{ff}(x,y) * [h^*(-x,-y) * h(x,y)] \\ &= R_{ff}(x,y) * R_{hh}(x,y) \end{aligned}$$

For convolution, the half-widths of the functions to be convolved add approximately to the half-width of the resulting function (Gaskill, 1978). The next step, then, is to correct for the imaging system.

The autocorrelation for the line spread functions were computed and their half-widths found to be

1.10 pixels or 63.8 meters along scan
0.42 pixels or 34.2 meters along track

The larger correction in the along scan direction is due to the electronic filtering of these data and the smaller pixel size, as described in Chapter 3, whereas the along track data is not filtered. Using these values to correct the image correlation half-widths yields the scene correlation half-widths:

5.28 pixels or 306 meters along scan

4.14 pixels or 337 meters along track

These correlation half-widths agree well with previous work (Craig, 1979; Lahart, 1979; Shlien, 1979) indicating a total correlation length of 10 pixels, twice the average half-width. Because these values are scene parameters, the imaging system effects having been removed, they may be used in the Wiener filter for restoration of imagery from both Landsat systems.

Noise Autocorrelation

A major source of noise in photoelectronic imaging systems is random thermal noise in the circuits which acquire and process the incoming signal. Thermal noise, due to the random motion of thermally excited free electrons in electronic circuits, is usually described by a zero mean Gaussian probability density function with a uniform ("white") power spectrum (Davenport, 1958), so called because the contribution to the variance is the same at any frequency, or "white". In practice, it is sufficient for the noise to be white over the region of interest, in particular, from zero frequency up to the Nyquist frequency.

Furthermore, it can be shown (Papoulis, 1984) that for a zero-mean random process such as noise, the area under the power spectrum of a signal, i. e. its total power, is equal to the variance of the corresponding signal.

$$\int_{-\infty}^{\infty} S_{nn}(u,v)dudv = R_{nn}(0,0) \tag{7.11}$$

For the noise power spectrum, then, we have assumed white Gaussian noise with a variance of one grey level (Schowengerdt, 1983) and used the relation between power spectrum and variance to set $S_n = 1.0$ for the region of interest. This value is verified by principal components analysis (Appendix C) performed on several representative Landsat images.

Windowing

From the restoration filter calculated using the Wiener filter one can derive the restoration (or deconvolution) line spread function by taking the filter's inverse Fourier transform. The resulting restoration line spread function will have significant side lobes due to the multiplication of it by a rectangle function which is implicit when using a finite set of data, (Figure 12). If applied directly to an image, this restoration filter would result visually in ringing at object edges and in considerable noise enhancement. For this reason it is advisable to apply a "window" function to the line spread function which acts to reduce the side lobes while keeping the main lobe relatively unchanged. A trade-off must be made, as any decrease in the side lobes, while reducing ringing and noise enhancement, will also reduce the amount of restoration obtainable.

Many criteria have been proposed for the proper window to use which will give the "best" results (Rabiner, 1971; Frieden, 1974; Rabiner and Gold, 1975; Hamming, 1983). The user first decides on a criterion and expresses it as a positive number such that poorer window designs have larger numbers associated with them. Then the window parameters are adjusted to make this number as small as possible. The criterion is user defined and always heuristic.

Windowing in the spatial domain is equivalent to convolution in the frequency domain, so that any modification of the restoration line spread function has an effect on the frequency response of the restoration filter. If the spatial window is too abrupt, then in the frequency domain the result will be the familiar Gibbs phenomenon in which there is significant overshoot at high frequencies which may result in objectionable visual artifacts in the restored image. If the window is too smooth, then in the frequency domain there will be a loss of power at all frequencies resulting in loss of resolution in the restored image. A choice of window which results in a good tradeoff of these factors is the Hanning or

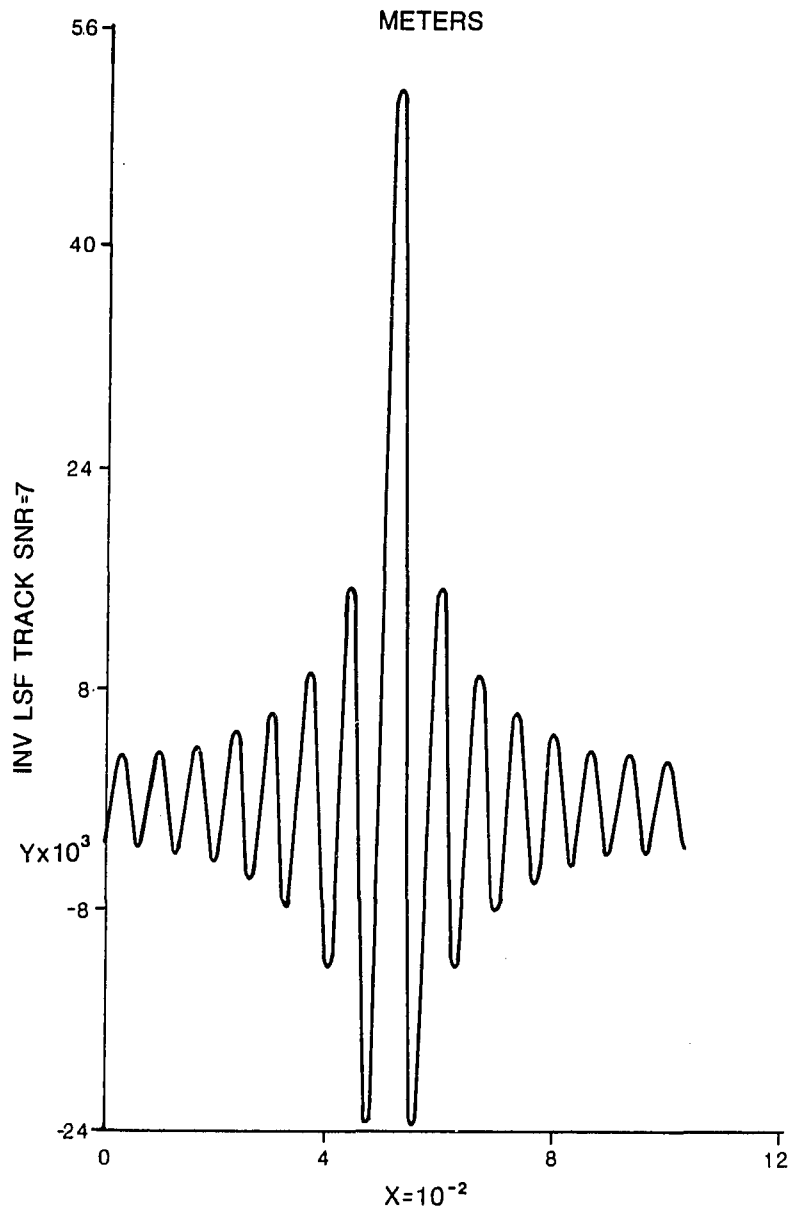


Figure 12. Restoration Spread Function Before Windowing.

Von Han window (Hamming, 1983). The Hanning window is of the form

$$w(x) = \begin{cases} (1/2)(1+\cos(\pi x/\tau)) & |x| < \tau \\ 0 & \text{otherwise} \end{cases} \tag{7.12}$$

where τ is the half-width of the window. This window was applied to the restoration line spread function resulting from the Wiener filter discussed previously. Various values for τ were used, and the resulting windows multiplied by the restoration line spread function. The Fourier transform of the windowed spread fucntions were taken, and the area under the curves above the folding frequency (0.5 cycles/pixel) calculated for each value of τ. A typical area vs. frequency curve can be found in Figure 13. The best value for τ is assumed to be the one that minimizes this side lobe power. It is seen from the figure that there is often no absolute minimum. However, as the window size increases other effects come into play which can be seen from a study of restored edge profiles, to be discussed in Chapter 8. In such cases, a tradeoff must be made.

Figure 14 shows the increase in high frequency modulation obtainable from the restoration. This figure includes the system transfer function, restoration filter without windowing and their product. It also shows the increased modulation obtainable from windowing. (However, attempting to increase this modulation too much by decreasing the window size will result in increased ringing and noise enhancement.)

This attempt to minimize the power under the side lobes of the Fourier transform of the restoration line spread function was verified by another criterion, maximizing the amount of restoration (slope of edge) to be described in Chapter 8, to optimize the spatial window width. It was determined that kernel sizes of 6-by-6 coefficients are appropriate for restoration of Landsat images.

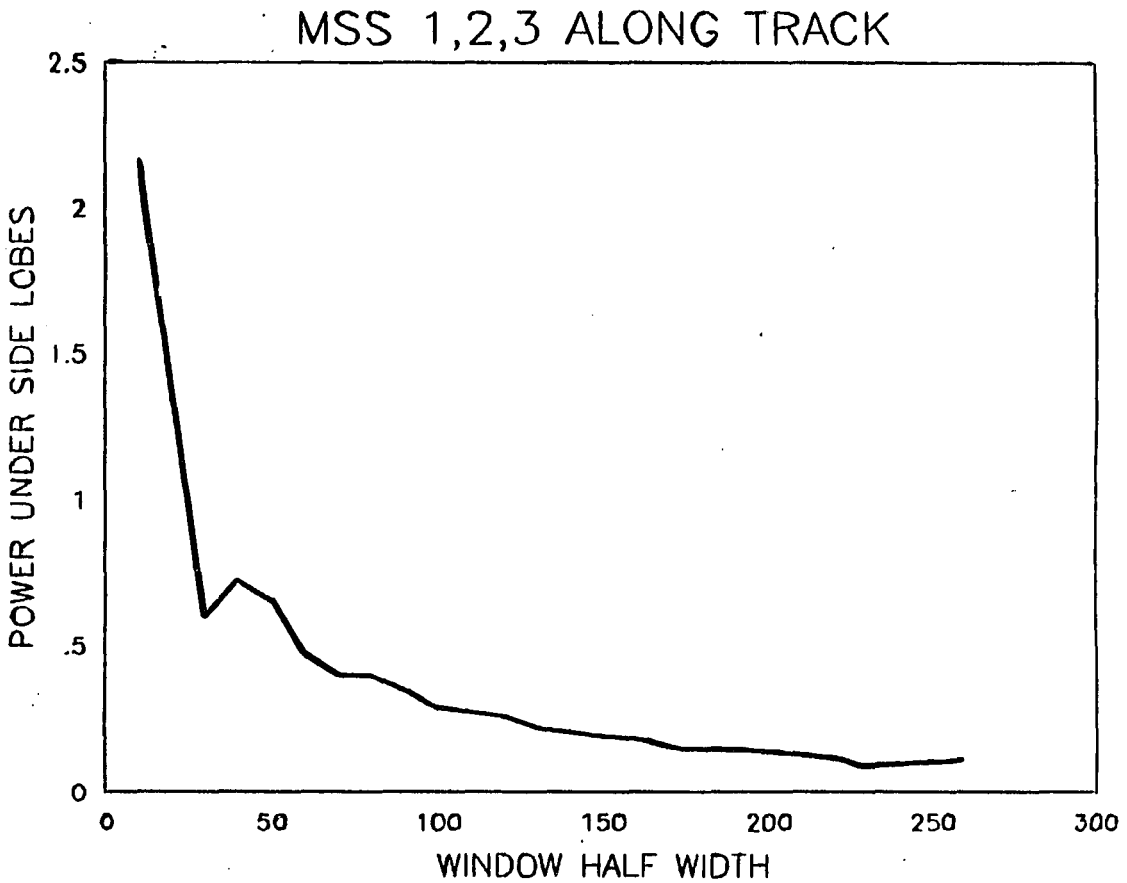


Figure 13. Area (power) vs. Frequency for Restoration Filters.

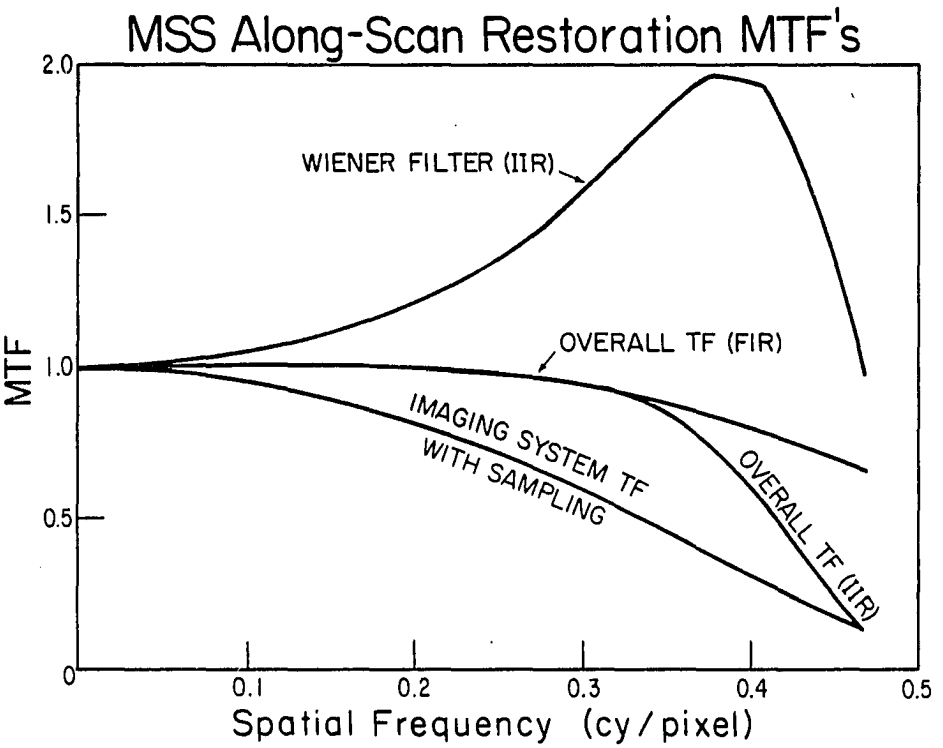


Figure 14. Typical Net Transfer Function: System × Restoration.

CHAPTER 8

QUANTITATIVE AND QUALITATIVE RESULTS

Several tests were made in order to verify the parameters used in the Wiener restoration filter and to illustrate the improvement obtained with a restoration technique over the usual resampling techniques described in Chapter 4.

Edge Profile Restorations

The image of a simple edge was studied in order to verify the best value to assign to τ of the Hanning window of Eq. (7.12), which is especially important in cases where the side lobe power as a function of τ did not have an absolute minimum. This edge image was simply a two gray level 256-by-256 image half of which was assigned the gray level 60 and the other the gray level 190. The image was blurred using the line spread functions which model each of the Landsat imaging systems, then subsampled to the pixel interval of the appropriate system in order to simulate an edge image for that system. The resulting (blurred, subsampled) edges were then resampled using a restoration line spread function derived from Eq. (5.8), with different Hanning window half-width values.

The amount of restoration which occurred could be measured by looking at the slope of the restored edge, a perfect edge having infinite slope. Thus the "restoration" in these terms is maximized when this slope is maximized. A typical edge slope vs. window size relation is shown in Figure 15. As the window size increases, the slope and so the amount of restoration increases. However, it does so very slowly after an intermediate half-width as shown in Figure 15.

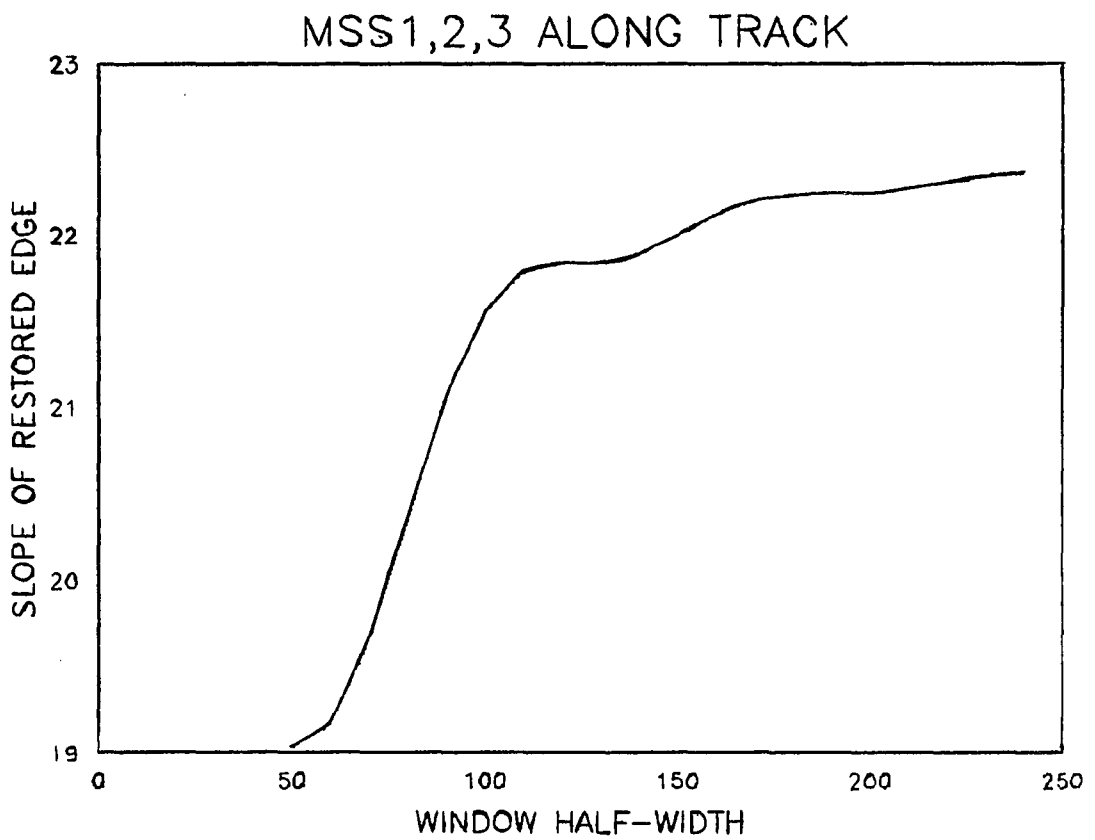


Figure 15. Edge Slope vs. Window Size for Edge Restoration Simulation.

Furthermore, the amount of ringing at the edge becomes objectionable for larger values of τ.

In the case of first three Multi-Spectral Scanners, two restored edge profiles are shown in Figure 16, which compares the profiles for large and small window sizes. Figure 17 shows a comparison of a restored edge to one resampled using cubic interpolation with $\alpha = -1$. The slope of the restored edge is significantly steeper than that of the resampled edge, indicating a sharpening of this feature in the case of restoration.

It should be noted that the slope of the restored edge was not used as a windowing criterion, as it is image dependent. That is, the restoration of this edge depends on the location of the edge with respect to the sampling grid, as discussed in Chapter 5. The edge slope was only used to support the optimization of the "side lobe power" criterion discussed in Chapter 7.

This phase dependence can be accounted for by shifting the edge by incremental amounts corresponding to, say, 1/8th of a pixel, and a comparison can be made between the restored edge and the cubic convolved edge. This was done, the errors between the original edge (scene) and these resampled edges are shown in Figure 18. It is clear that the mean square error for the restored edge is less than that for the cubic convolved edge, and, for this particular image, less yet when sampling is included.

Delta Function Approximation

As discussed in Chapter 2, a perfect imaging system would have a Dirac delta function for its point spread function. Thus, one measure of the restoring ability of the restoration point spread function is how close the net point spread function, the imaging point spread function convolved with the restoration point spread function, approximates a delta function. This convolution was studied for various window sizes and it was found that the value of τ which gave the

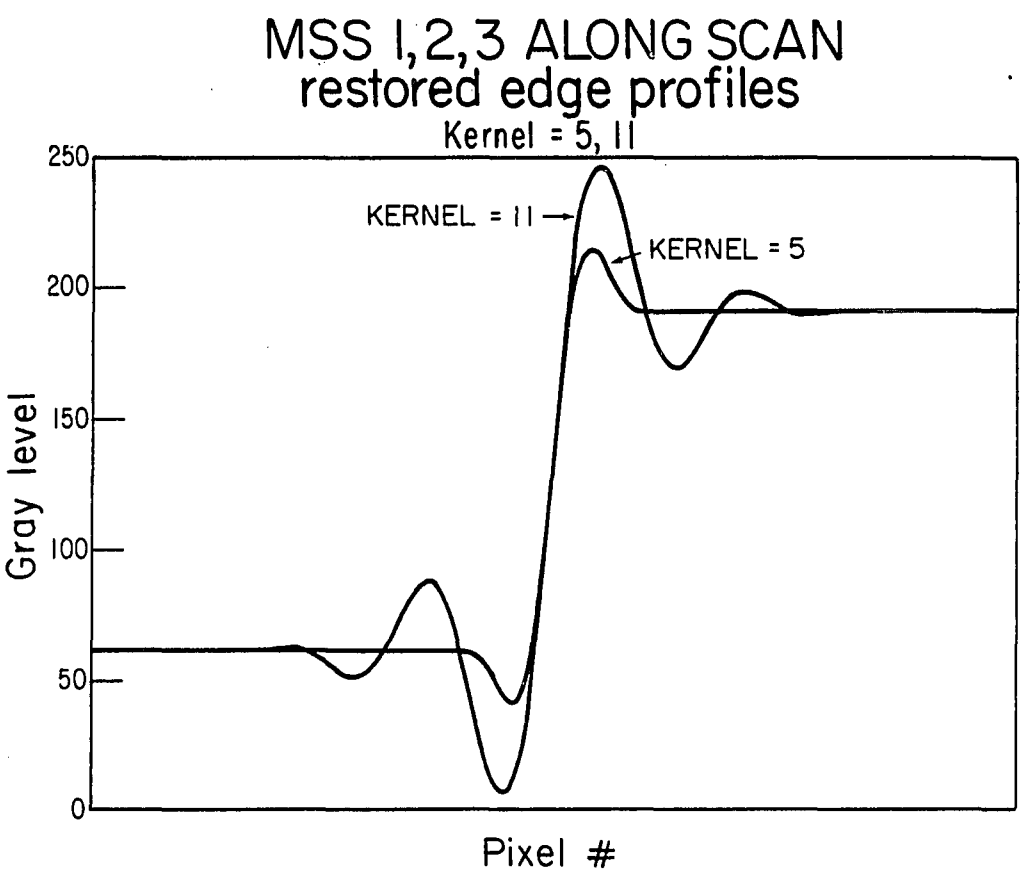


Figure 16. Edge Profiles for Large and Small Window Sizes.

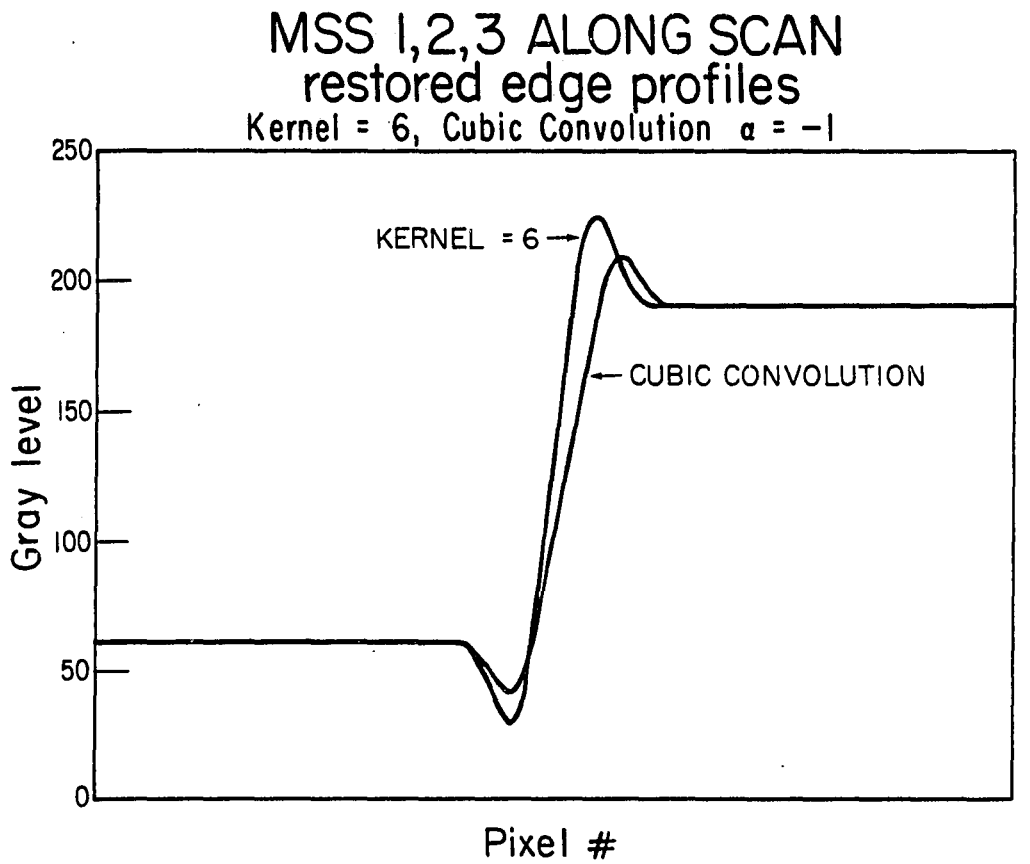


Figure 17. Edge Profiles for Restoration vs. Cubic Convolution Resampling.

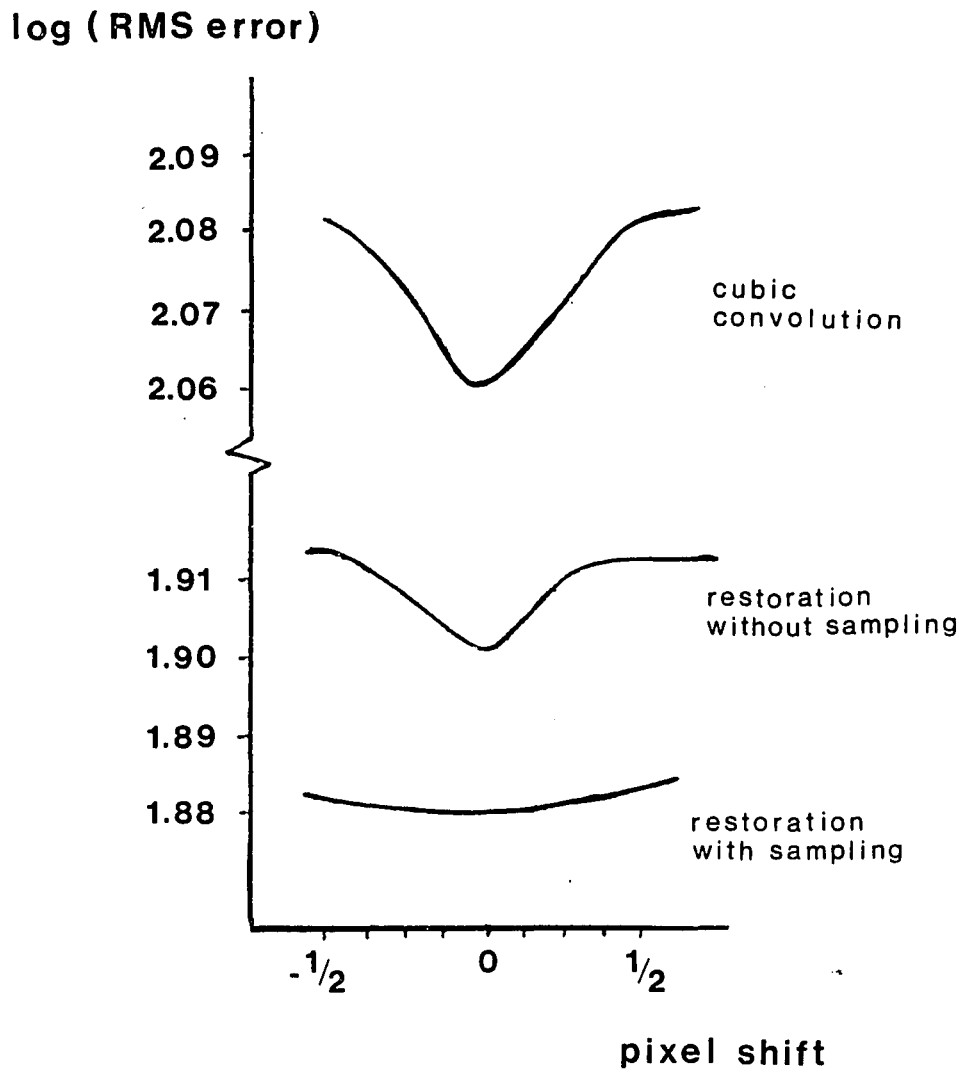


Figure 18. Error Comparison for Edge Simulation.

closest approximation to a delta function (in terms of the width of the curve and the minimization of side lobes) was a value of τ which corresponds to a kernel size of about 5. The resulting convolution, for the Multi-Spectral Scanner, is shown in Figure 19.

Simulation Tests

A third test was made using a high resolution aerial image which was digitally magnified in three ways: using bilinear interpolation, cubic convolution and the restoration filter. The results are shown in Figure 20. The restored image (lower right) is clearly sharper than the images magnified with bilinear interpolation (upper right) and cubic convolution (lower left), although there is some noise enhancement which is apparent in the relatively homogeneous region of the water.

As a more quantitative test, the high resolution aerial image was digitally degraded using the known system point spread function of the Thematic Mapper, then subsampled to the instantaneous field of view of that system and resampled using cubic convolution and the restoration filter. The mean square error between the original aerial image and the restored and resampled images were calculated. The error between the original and the image resampled using cubic convolution was 8.92 gray levels. The error between the original and a restored image (without sampling) using a constant signal-to-noise ratio in the Wiener filter was 8.46 and, when the Lorentzian model was used in the Wiener filter, the error reduced further to 8.45. However, when the effect of sampling was included in the Wiener filter, (i.e. Eq. (5.8) was used as opposed to Eq. (2.36)) the mean square error actually increased again, to 8.76 for a constant signal-to-noise and to 9.27 when the Lorentzian model was used. This may correspond to an increase in contrast which results from the added boosting of high frequencies with the inclusion of sampling as shown in Figure 20.

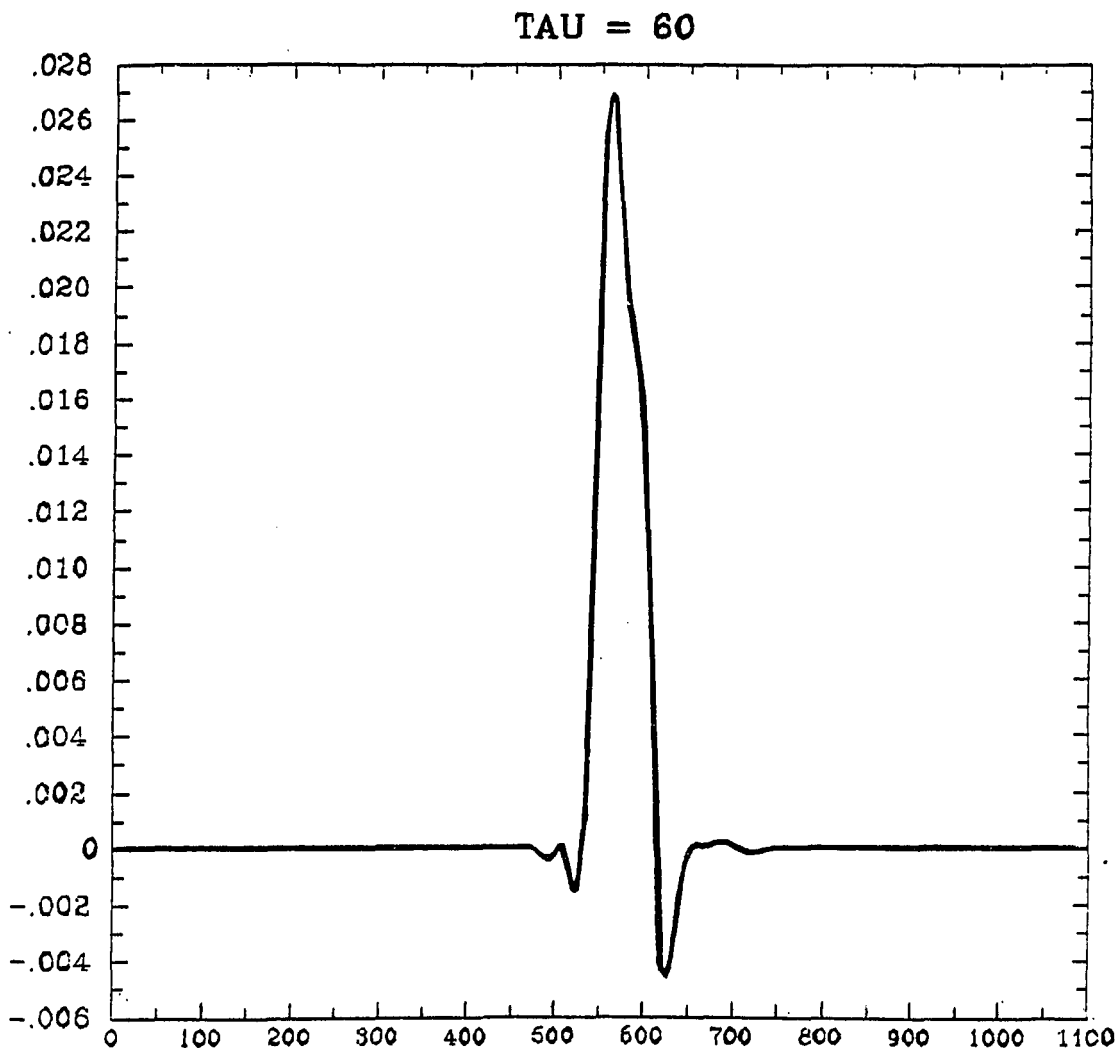


Figure 19. Net Spread Function Delta Function Approximation: System × Restoration.

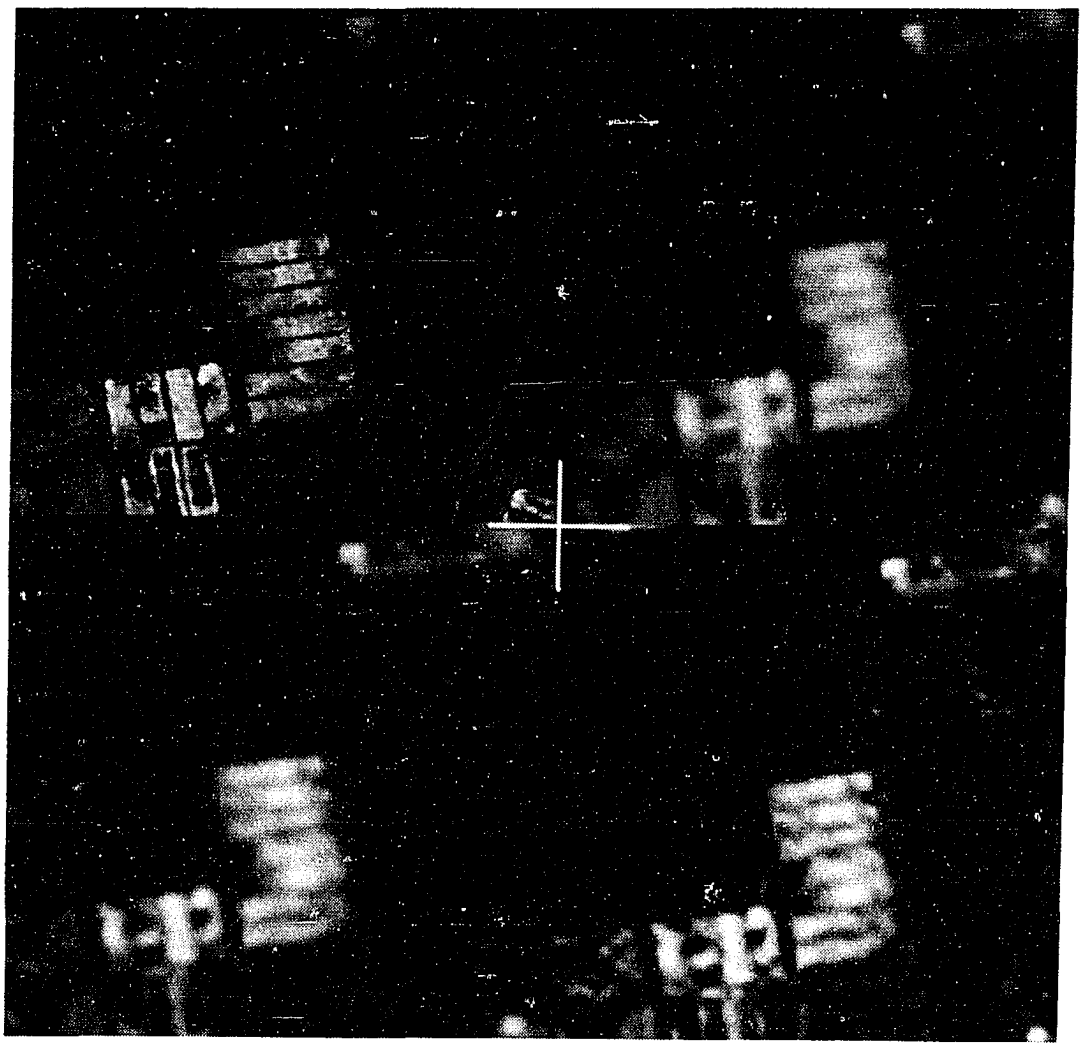


Figure 20. Qualitative Comparison of Interpolation Methods.

Clockwise from upper left: Aerial image, simulated Landsat image, restored image, cubic convolved image.

White Sands Missile Range Target

Finally, the restoration was tested on a man-made "target" located at the White Sands Missile Range in White Sands, New Mexico. The target consists of an array of black tar 15-by-15 meter squares laid out on the highly reflective white sands of the Missile Range. They are spaced so that each target falls at a different inter-sample position with respect to the Thematic Mapper sampling grid (Schowengerdt, et. al., 1985b; Rauchmiller and Schowengerdt, 1986). Figure 21 compares the effects of interpolation and restoration for the resampling of this Thematic Mapper image. The restoration improves the contrast of the dark squares against the light background relative to the bilinear (upper left) and cubic interpolated (upper right) images. Further contrast improvement in seen when the effect of sampling is included in the restoration (lower right) as compared to when it is not included (lower left).

Conclusions

All the tests reinforce the notion that restoration does, indeed, improve (i. e. sharpen) the processed images over other resampling techniques typically used. The subjective improvement, however, seems to be greater than objective measures such as mean square error might indicate. This result, however, has been observed by many other investigators (e. g. Huhns, 1975): lower mean square error does not always correspond to visual improvement.

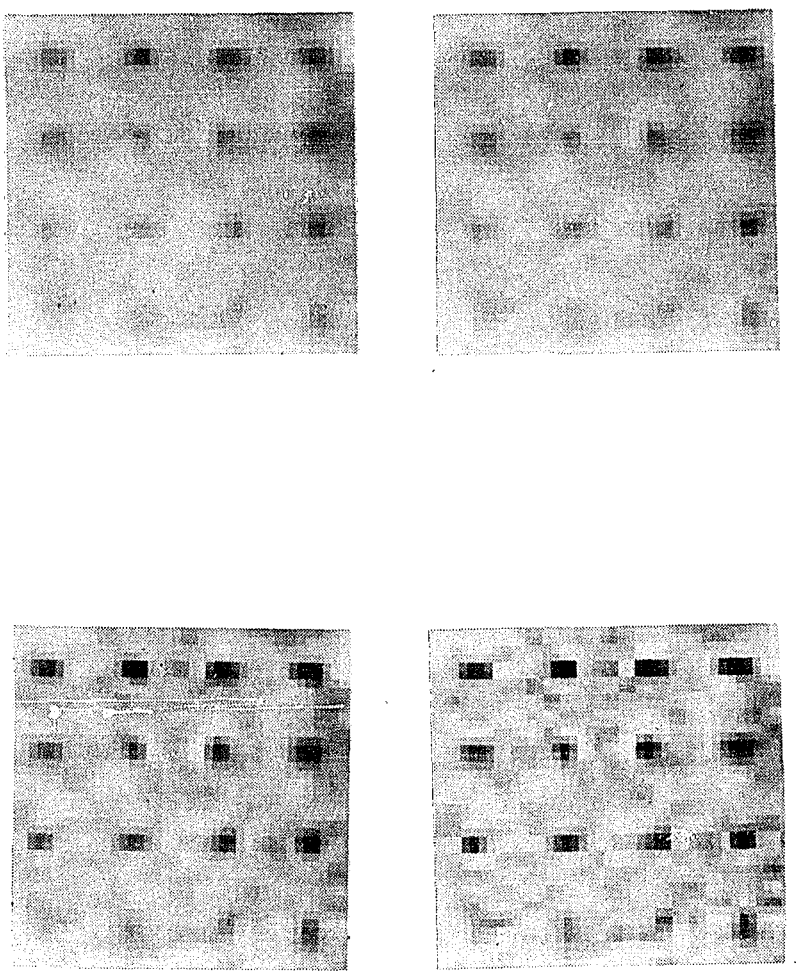


Figure 21. Comparison of Interpolation Methods on Real Landsat Images.

Clockwise from upper left: Bilinear interpolation, cubic convolution, restored including sampling, restored without sampling.

CHAPTER 9

TOPICS FOR POSSIBLE FUTURE STUDY

Several issues have been encountered during the undertaking of this research which have not been pursued, but which should be mentioned as they are potential areas for future study. These topics will be discussed now.

Double Resampling

The EROS Data Center employs a two-pass resampling for the geometric rectification and registration of images. For Multi-Spectral Scanner data this is done at their option. But for Thematic Mapper data this is forced upon them as that data has already been resampled in the along scan direction before being acquired by the Data Center. Since resampling can be viewed as an added convolution or blurring of the data, as discussed in Chapter 4, this two-pass procedure degrades the image that much more.

There are two ways in which this can be avoided. One is to resample the data just once, that is, perform all the geometric corrections, both rectification and registration, in one pass. The second is to add a term to the restoration filter which accounts for the effect of prior resampling. In the latter case, the resampling transfer function may be considered as another component in the overall system, so that the "system" now includes some ground processing. The system transfer function is then

$$H_s(u,v) = H_a(u,v)H_r(u,v) \tag{9.1}$$

where $H_a(u,v)$ is the average system transfer function described in Chapter 5, $H_r(u,v)$ is the transfer function associated with the interpolator chosen for the resampling as described in Chapter 4, and $H_s(u,v)$ is the overall system

transfer function which now accounts for the imaging system, sampling and ground processing.

At the time this research was being carried out, EROS had not yet decided which method to pursue, so that an implementation study was not included in this research.

Pre-Restoration

The idea of adding the resampling interpolator transfer function to the overall system transfer function suggests another intriguing possibility, to pre-restore for anticipated degradations. In particular, terms can be included in the overall system transfer function which will account for all the ground processing including degradations introduced by the display media (CRT, photographic printing or screening, lithography). It is even conceivable to restore for the human visual system, which acts as a band-pass filter with a circularly symmetric transfer function given by (Hunt, 1975)

$$H_e(u) = \sqrt{(1 + 0.005u^2)e^{-(u/50.0)^2}} \quad \text{for} \quad 0 \leq u \leq 100 \quad \text{cy/dist} \qquad (9.2)$$

where the units cy/dist (cycles per unit distance) depend on the viewing distance. This transfer function is determined by experimentation, and would vary from one individual to another.

Gray Level Quantization

The variation in light intensity from a scene is continuous, and therefore the potential number of gray levels in an image is infinite. Thus, the number of distinct output values from Landsat analog sensors is so large as to require a quantization strategy before encoding.

A quantizer is a mapping from an essentially continuous domain to a domain consisting of a limited number of integers. These integers become the values that are then encoded and transmitted (or stored).

A one-dimensional quantizer may be defined as follows: Let $u \in \Re^n$ represent the brightness, contrast, or other quantity associated with a picture element. Divide $\Re^n$ into a finite set of K subintervals, bins, or decision regions Δd_0, Δd_1, ..., Δd_k, where $\Delta d_0 = (-\infty, d_1]$, $\Delta d_{k-1} = (d_{k-1}, +\infty)$ and $\Delta d_i = (d_i,\ d_{i+1}]$ for $i \neq 0, K-1$. If u lies in the interval Δd_i, then u is mapped to a quantization level q_i, $i \in [0, K-1]$. The quantity q_i is the quantized value associated with u. If all the subintervals Δd_i, $i \neq 0, K-1$ are of the same width, then the quantizer is a uniform quantizer. Nonuniform quantizers allow Δd_i to vary.

Quantization is an irreversible procedure; given an output value, one cannot, in general, determine its corresponding input value. (For this reason it is the major source of error in image data compression.)

The mean squared error between the input value u and its associated quantized value q_k for the case when u is uniformly distributed over the bins or decision regions Δd_k is given by

$$e = \int_{d_k}^{d_k+1} (q_k - u)^2 du \tag{9.3}$$

If u is not uniformly distributed within the bins then the squared error $(q_k - u)^2$ must be weighted by the probability density function $p(u)$ and

$$e = \int_{d_k}^{d_k+1} (q_k - u)^2 p(u) du \tag{9.4}$$

and the total quantizer error becomes a weighted average of these terms. In particular, to find a quantizer which minimizes the mean squared error for a given number K of quantization levels, one must minimize

$$\sum_{k=0}^{K+1} \int_{d_k}^{d_{k+1}} (q_k - u)^2 p(u) du \tag{9.5}$$

Ideally, a quantizer would be used which minimizes this error Eq. (9.5) in some sense. Much research has been done on finding such a quantizer (e. g. Panter

and Dite, 1951; Max, 1960; Roe, 1964; Wood, 1969) for various assumptions about the uniformity of the decision regions Δd_k, the number of quantization levels K (or quantization regions in N-space) and the form of the density function $p(u)$. Less effort has been expended on restoring the effects of the quantization.

One major effort (Huhns, 1975), however, models the scene as an N-dimensional random vector $\mathbf{x} = (x_1 \quad x_2 \quad \ldots \quad x_n)^{\mathbf{T}}$ and the restored value of the quantized scene as the vector $\mathbf{y}_k = (y_1 \quad y_2 \quad \ldots \quad y_n)^{\mathbf{T}}$. Then minimizing the error between scene and restoration given by

$$\sum_{k=0}^{K-1} \int_{\mathcal{R}_k} w(\mathbf{x} - \mathbf{y}_k) p(\mathbf{x}) d\mathbf{x} \tag{9.6}$$

where $\mathcal{R}_k$ represent the N-dimensional decision regions corresponding to the one-dimensional decision regions Δd_k above, and w is any weighting function, (one possibility being the mean-square error criterion of Eq. (8.5)), yields the restored vector

$$\mathbf{y}_k = \frac{\int_{\mathcal{R}_k} \mathbf{x} p(\mathbf{x}) d\mathbf{x}}{\int_{\mathcal{R}_k} p(\mathbf{x}) d\mathbf{x}} \tag{9.7}$$

This quantity was evaluated by Huhns for several types of scene statistics. It was shown that a decrease in mean square error was obtained in all simulations, but a corresponding subjective improvement in image quality was not always observed.

Another approach is to treat quantization as a source of noise and include it in the noise-to-signal term in the Wiener filter. For $K = 2^b$ quantization intervals from 0 to $g_{\max}$, where $g_{\max} = \bar{g} + 6\sigma_g$ is typically chosen (Oppenheim and Schafer, 1975), and where $\bar{g}$ is the mean gray level and σ_g the standard deviation, an α can be found such that

$$g_{\max} = \alpha \bar{g} \quad \text{where} \quad \alpha > 1 \tag{9.8}$$

Then for b large enough, greater than about 5 or 6, it can be shown (Oppenheim and Schafer, 1975) that the quantization noise within an interval is well

approximated by a uniform probability distribution. Thus the root mean square quantization noise n_q is given by

$$n_q = \frac{\Delta d_i}{\sqrt{12}} \tag{9.9}$$

where Δd_i is the width of the interval. For a uniform quantizer, where $\Delta d_i = \alpha \bar{g}/2^b$,

$$n_q = \frac{\alpha \bar{g} 2^{-b}}{\sqrt{12}} \tag{9.10}$$

Quantization effects might be accounted for in this way in the Wiener filter restoration approach, but based on Huhns' study, it might be wise to first determine if the payoff would be worth the effort.

APPENDIX A

THEOREMS

Theorem 1: A function $m(x,y)$ minimizes Eq. (2.28) if and only if $m(x,y)$ satisfies

$$E[(f(x,y) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(x-x',y-y')g(x',y')dx'dy')g(x_0,y_0)] = 0 \quad \text{(A.1)}$$

for all points (x,y) and (x_0,y_0) in the xy-plane, and where $g(x,y)$ is defined as in Eq. (2.8).

Proof: Suppose $m(x,y)$ satisfies Eq. (A.1). Then it will be shown that any other choice $m'(x,y)$ for the deconvolution filter will result in a larger mean square error than that given by Eq. (2.28).

Let e'^2 be the mean square error Eq. (1.28) with $m'(x,y)$ for $m(x,y)$:

$$e'^2 = E[(f(x,y) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m'(x-x',y-y')g(x',y')dx'dy')^2] \quad \text{(A.2)}$$

It will now be shown that e'^2 is minimized when $m'(x',y') = m(x',y')$. Eq. (A.2) can be written

$$\begin{aligned} &E[(f(x,y) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(x-x',y-y')g(x',y')dx'dy' \\ &+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(m(x-x',y-y') - m'(x-x',y-y'))g(x',y')dx'dy')^2] \end{aligned} \quad \text{(A.3)}$$

or

$$\begin{aligned} &e^2 + E[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(m(x-x',y-y') - m'(x-x',y-y'))g(x',y')dx'dy')^2] \\ &+ 2E[(f(x,y) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(x-x',y-y')g(x',y')dx'dy') \\ &\times (\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(m(x-x',y-y') - m'(x-x',y-y'))g(x',y')dx'dy')] \end{aligned} \quad \text{(A.4)}$$

The second term of Eq. (A.4) is always nonnegative. The third term is zero by assumption (Eq. A.1) since by changing the variable of integration in the second integral from (x', y') to (x_0, y_0) and interchanging the integration with expectation yields

$$2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[(f(x,y) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(x-x', y-y')g(x',y')dx'dy') \times g(x_0,y_0)](m(x-x_0, y-y_0) - m'(x-x_0, y-y_0))dx_0dy_0 \quad \text{(A.5)}$$

which is zero. Thus $e'^2 \geq e^2$.

Conversely, if $m(x,y)$ minimizes Eq. (2.28) then $m(x,y)$ satisfies Eq. (A.1). This result follows immediately from the Orthogonality Principle (Gardner, 1986). $\diamondsuit$

Theorem 2: If the function describing the scene intensity distribution $f(x,y)$ is band limited, that is, its spectrum $F(u,v)$ is such that $F(u,v) = 0$ when $|u| \geq u_c$ or $|v| \geq v_c$, and sufficiently sampled (i. e. at a sampling rate at least twice per period for the highest frequency present in the signal), then $f(x,y)$ can be reconstructed exactly from convolution with a $(\sin(\pi x)/\pi x) \times (\sin(\pi y)/\pi y)$ function.

Proof: Suppose $F(u,v) = \mathcal{F}[f(x,y)]$, where $F(u,v)$ is band-limited. If $f(u,v)$ is sampled, so that

$$f_s(x,y) = \sum_{m=-\infty}^{m=+\infty}\sum_{n=-\infty}^{n=+\infty} \delta(x-m, y-n)f(x,y) \quad \text{(A.6)}$$

$$= \sum_{m=-\infty}^{m=+\infty}\sum_{n=-\infty}^{n=+\infty} f(m,n)\delta(x-m, y-n) \quad \text{(A.7)}$$

Then

$$F_s(u,v) = \sum_{m=-\infty}^{m=+\infty}\sum_{n=-\infty}^{n=+\infty} F(u-m, v-n) \quad \text{(A.8)}$$

So that

$$F(u,v) = U(u,v)F_s(u,v) \tag{A.9}$$

where

$$U(u,v) = \begin{cases} 1 & |u| < \frac{1}{2},\ |v| < \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases} \tag{A.10}$$

So that

$$\begin{aligned} f(x,y) &= \mathcal{F}^{-1}[U(u,v)F_s(u,v)] \\ &= \mathcal{F}^{-1}[U(u,v)] * \mathcal{F}^{-1}[F_s(u,v)] \\ &= \frac{\sin(2\pi x)}{2\pi x}\frac{\sin(2\pi y)}{2\pi y} * f_s(x,y) \end{aligned} \tag{A.11}$$

Thus, the original scene $f(x,y)$ is reconstructed exactly. $\diamond$

APPENDIX B

TABLE LOOK-UP RESTORATION KERNEL

In actual implementation of any resampling procedure, computational efficiency usually demands that the interpolation weights are kept in a look-up table rather than calculated at each use. The question then arises: How big should the look-up table be to guarantee a certain accuracy?

Currently, much resampling is done using a cubic convolution interpolator. This method is well understood and documented, and the interpolator can be written in closed form, as in Eq. (4.2). Thus, the equation describing cubic interpolation can be coded into the resampling routine, and an interpolation weight calculated at each step. Alternatively, a table of weights can be generated in which the weight used for an intermediate value is the nearest, say, 1/32nd interval value. The difference between the data resampled using the look-up table and that using the "continuous" equation can be computed and an error between the two found.

A line profile through one band of a typical Landsat Multi-Spectral Scanner image was taken and interpolated values found by calculating the appropriate weight at each step using the cubic convolution equation directly. Then interpolated values were found by using look-up tables of various sizes and the results compared.

Figure 22 shows the root mean squared error between the resampled image profile using the interpolation equation directly and using look-up tables of various sizes. In particular, it shows that as the size of the look-up table increases, the error decreases. This is as expected, since an infinite look-up table would be equivalent to calculating the weights directly. (Actually, finite

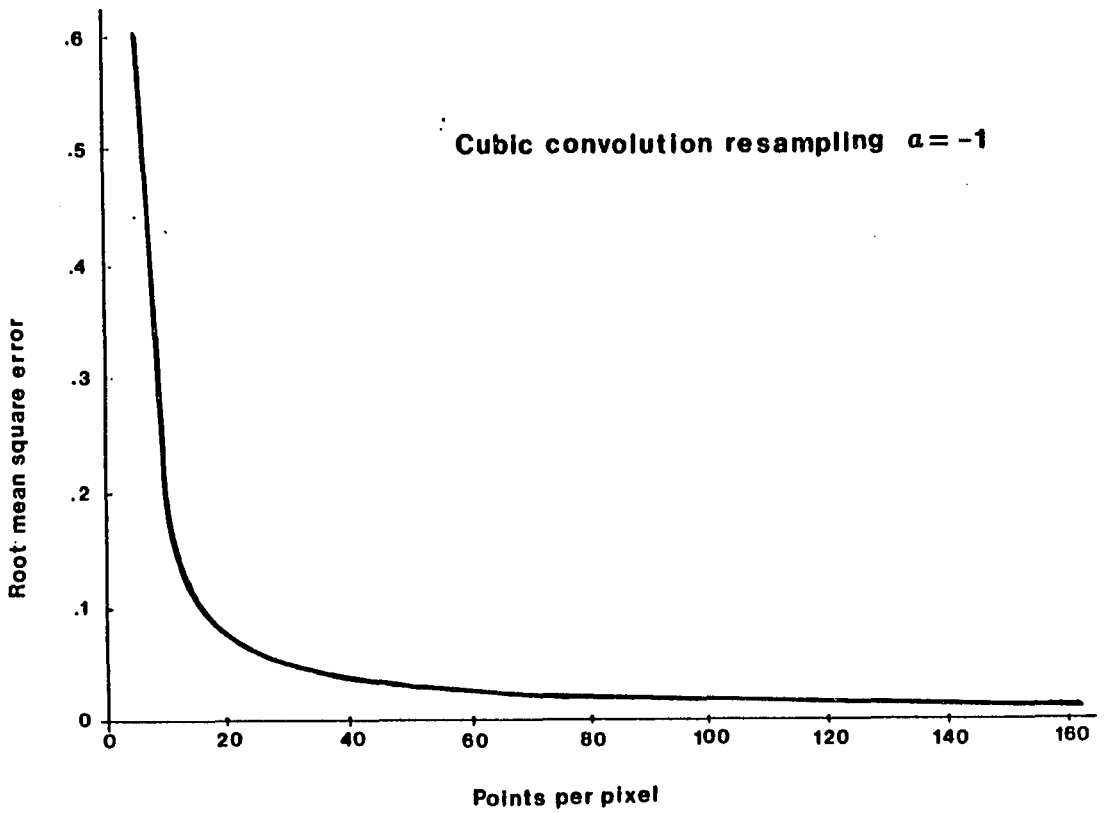


Figure 22. Effect of Look-up Table Size on Resampling Error.

machine precision would make the two comparable far short of infinity.) It is seen that after about 32 points per pixel in the look-up table, the added precision would probably not be worth the added effort.

It is impossible to do the same comparison using the restoration coefficient values, because the restoration line spread function is not available in a closed form which could be used for comparison. However, based on these studies of the cubic convolution interpolator it is felt that 32 points per pixel is also appropriate for restoration resampling because restoration functions resemble closely a cubic-type function. This is the size of the look-up table used at the EROS Data Center.

APPENDIX C

PRINCIPAL COMPONENT ANALYSIS

Many imaging processes result in a high degree of correlation among picture elements (pixels). For example, the bands of a Landsat multispectral image are highly correlated (Schowengerdt, 1983). A transformation may be performed on the data which produces relatively uncorrelated data. There are several such transformation techniques which may be employed depending on the particular situation.

In general, the best transformation would be one that results in statistically independent variables. However, this requires knowledge of the joint density function of all the pixels. Using linear transformations, the best one can do is to produce uncorrelated coefficients (Wintz, 1972).

A particular linear transformation which has found wide application in image processing is the principal component transformation (also called the Karhunen-Loève, Hotelling or eigenvector transformation). This transformation takes advantage of the statistical redundancy in an image.

An n-dimensional random vector $\mathbf{x}$ is selected. This random vector $\mathbf{x}$ could represent the "stacked" image or the spectral intensities associated with each image point. The random vector $\mathbf{x}$ may, without loss of generality, be represented as a linear combination of linearly independent column vectors $\mathbf{a}_i$ which span the space containing $\mathbf{x}$:

$$\mathbf{x} = \mathbf{A}\mathbf{y} = \sum_{i=1}^{n} y_i \mathbf{a}_i \tag{C.4}$$

where $\mathbf{A} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_n)$, $\det(\mathbf{A}) \neq 0$, and $\mathbf{y}$ is also an n-dimensional

random vector. If it is also assumed that $\mathbf{A}$ is orthogonal, i.e. $\mathbf{A}^{-1} = \mathbf{A}^{\mathbf{T}}$, then

$$\mathbf{y} = \mathbf{A}^{\mathbf{T}}\mathbf{x} \quad \Longrightarrow \quad y_i = \mathbf{a}_i^{\mathbf{T}}\mathbf{x} \tag{C.5}$$

The n-dimensional random vector $\mathbf{y}$ can be thought of as a vector of weights, each of which multiplies one of the basis vectors, which are then added to equal $\mathbf{x}$. An estimate $\hat{\mathbf{x}}(k)$ of $\mathbf{x}$ occurs when only $k < n$ components of $\mathbf{y}$ are used. In particular, one could estimate $\mathbf{x}$ by replacing the omitted components of $\mathbf{y}$ by preselected constants:

$$\hat{\mathbf{x}}(k) = \sum_{i=1}^{k} y_i \mathbf{a}_i + \sum_{i=k+1}^{n} b_i \mathbf{a}_i \tag{C.6}$$

The difference between the estimated and true vectors can be thought of as an error,

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}(k) = \sum_{i=k+1}^{n} (y_i - b_i)\mathbf{a}_i \tag{C.7}$$

The mean squared value of $\mathbf{e}$ may be chosen to measure the effectiveness of the estimate, and the b_i's chosen to minimize this mean squared error. Linear mean square estimation theory applied to this problem results in the optimum choice for the constants b_i as $b_i = E[y_i]$, the mean values of the "weights" (Papoulis, 1984). These optimum constants depend on the transformation matrix as follows,

$$b_i = \bar{y}_i = \mathbf{a}_i^{\mathbf{T}} E[\mathbf{x}] = \mathbf{a}_i^{\mathbf{T}} \bar{\mathbf{x}} \tag{C.8}$$

The mean squared error $\overline{\mathbf{e}^2(k)}$ is then

$$\begin{aligned} \overline{\mathbf{e}^2(k)} = E[\mathbf{e}^{\mathbf{T}}\mathbf{e}] &= \sum_{i=k+1}^{n} (y_i - \bar{y}_i)(y_i - \bar{y}_i)^{\mathbf{T}} \\ &= \sum_{i=k+1}^{n} \mathbf{a}_i^{\mathbf{T}} E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^{\mathbf{T}}]\mathbf{a}_i \\ &= \sum_{i=k+1}^{n} \mathbf{a}_i^{\mathbf{T}} \mathbf{C}_{\mathbf{x}} \mathbf{a}_i \end{aligned} \tag{C.9}$$

where the fact that $\mathbf{a}_i^{\mathbf{T}}\mathbf{a}_j = \delta_{ij}$ has been used. But $E[(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^{\mathbf{T}}] = \mathbf{C}_{\mathbf{x}}$ is the covariance matrix associated with $\mathbf{x}$, its diagonal elements are the variances of the components of $\mathbf{x}$, which are always nonnegative. Since the off-diagonal elements are the covariances of the components x_i and x_j, $\mathbf{C}_{\mathbf{x}}$ is symmetric for stationary random processes.

If the basis vectors $\mathbf{a}_i$ are chosen as the eigenvectors of $\mathbf{C}_{\mathbf{x}}$, that is if $\mathbf{C}_{\mathbf{x}}\mathbf{a}_i = \lambda_i\mathbf{a}_i$, then since $\mathbf{a}_i^{\mathbf{T}}\mathbf{a}_j = \delta_{ij}$, $\lambda_i = \mathbf{a}_i^{\mathbf{T}}\mathbf{C}_{\mathbf{x}}\mathbf{a}_i$. It follows that

$$\overline{\mathbf{e}^2(k)} = \sum_{i=k+1}^{n} \lambda_i \tag{C.10}$$

The mean squared error can be minimized by ordering the eigenvalues so that

$$\lambda_1 > \lambda_2 > \ldots > \lambda_n > 0$$

Thus, if a component y_i is deleted, the mean squared error increases by λ_i. By deleting components corresponding to the smallest eigenvalues, the mean squared error is minimized.

When this transformation is applied to multispectral images, the resulting principal component images are uncorrelated and ordered by decreasing gray level variance, since the eigenvalues λ_i are the variances of the principal components. If variance is considered a measure of information content, then the transformation optimizes the redistribution of variance in that the entropy defined over the data variance distribution is minimized. In particular, it can be shown (Watanabe, 1965) that

$$-\sum_k \lambda_k \log \lambda_k < -\sum \rho_i \log \rho_i \tag{C.11}$$

where $\rho_i = \sigma_i^2 / \sum \sigma_j^2$ and where the σ_i^2's are the variances of the original spectral bands. Thus, the result of the principal component transformation is twofold: (1) The removal of the correlation that was present between the spectral bands

of the original image and (2) the creation of a new set of images for which fewer of the component images contain more of the total variance.

A principal component transformation is typically performed on a noisy image $\mathbf{z} = \mathbf{x} + \mathbf{n}$. If the noise is represented by a vector $\mathbf{n}$ whose elements are assumed to be uncorrelated, zero mean and identically distributed random variables, then the associated covariance matrix $\mathbf{C_n}$ is given by

$$\mathbf{C_n} = \sigma_n^2 \mathbf{I}$$

where σ_n^2 is the noise variance. Then for additive signal independent noise

$$\mathbf{C_z} = \mathbf{C_x} + \mathbf{C_n} = \mathbf{C_x} + \sigma_n^2 \mathbf{I}$$

and

$$\begin{aligned} \mathbf{C_z a}_i &= (\mathbf{C_x} + \mathbf{C_n})\mathbf{a}_i \\ &= (\mathbf{C_x} + \sigma_n^2 \mathbf{I})\mathbf{a}_i \\ &= (\lambda_i + \sigma_n^2)\mathbf{a}_i \end{aligned}$$

Thus each eigenvalue of the noisy image is the sum of the eigenvalue of the noiseless image with the variance of the noise. Since $\lambda_n \approx 0$ for correlated data (Ready and Wintz, 1973), this indicates a method of estimating the noise in correlated data:

$$\lambda_n \approx \sigma_n^2$$

The principal component transformation was performed on several Landsat scenes and the value of the nth eigenvalue taken as the variance of the noise for each image. The results are shown in Table 3 of Figure 23. These values, which range between 1 and 2 gray levels, verify the value of 1.0 used for the noise variance as discussed in Chapter 7.

Table 3. Eigenvalues from Principal Components

Test Image 1	Test Image 2	Test Image 3	Test Image 4
5.77	287	46.9	199
2.87	5.88	4.66	21.2
1.92	2.14	1.87	2.41
1.32	1.95	1.43	2.01

Figure 23. Principal Component Eigenvalues

REFERENCES

Andrews H. C., "Digital image restoration: A survey",*IEEE Computer*, **7** pp. 6–15 (1974).

Andrews, H. C. and B. R. Hunt, *Digital Image Restoration* Prentice-Hall, Englewood Cliffs (1977).

Arsac, J., *Fourier Transforms and the Theory of Distributions* Prentice-Hall, Englewood Cliffs (1966).

Barthes, V. and G. Vasseur, "An inverse problem for electro- magnetic prospection", *Lecture Notes in Physics, Vol. 85: Applied Inverse Problems* P. C. Sabatier, Ed. Springer-Verlag, Berlin (1978).

Brigham, E. O.,*The Fast Fourier Transform* Prentice- Hall, Englewood Cliffs (1974).

Brown, R. G., *Introduction to Linear Systems Analysis* John Wiley and Sons, New York (1962).

Cappellini, V. , A. G. Constantinides and P. Emiliani, *Digital Filters and Their Applications*, Academic Press, London (1978).

Champeney, D. C., *Fourier Transforms and Their Physical* Applications, Academic Press, London (1973).

Craig, R. G., "Autocorrelation in Landsat data" *Proceedings of the 13th International Symposium on Remote Sensing of the Environment*, April 1979, pp. 1517-1524.

Davenport, W. B. and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York (1958).

Duff, I. S. , Ed., *Sparse Matrices and Their Uses*, Academic Press, London (1981).

Dye, R. H., "Restoration of Landsat images by discrete two- dimensional deconvolution", *Proceedings of the 10th International Symposium on Remote Sensing of the Environment*, October 1975, pp. 725– 729.

Dye, R. H., "A quantitative assessment of resampling errors", *Proceedings of the NASA Workshop on Registration and Rectification*, JPL Publ. 82-23, pp. 371–376 (1982).

Fletcher, R., *Practical Methods of Optimization*, Vol. 1., John Wiley and Sons, Chichester (1980).

Franks, L. E., "A model for the random video process',' *The Bell System Technical Journal*, **45**, pp. 609–630 (1966).

Frieden, B. R., "Restoring with maximum likelihood and maximum entropy", *Journal of the Optical Society of America*, **62**, pp. 511–518 (1972).

Frieden, B. R., "Image restoration by discrete convolution of minimal length", *Journal of the Optical Society of America*, **64**, pp. 682–686 (1974).

Frieden, B. R., "Restoring with maximum entropy. III. Poisson sources and backgrounds", *Journal of the Optical Society of America*, **68**, pp. 93–103 (1978).

Frieden, B. R., "Image enhancement and restoration", *Topics in Applied Physics, Vol. 6: Picture Processing and Digital Filtering*, T. S. Huang, Ed. Springer-Verlag, New York (1979).

Friedman, D. E., "Operational resampling for correcting images to a geocoded format", *Proceedings of the 15th International Symposium on Remote Sensing of the Environment*, May 1981, pp. 195–212.

Gardner, W. A., *Introduction to Random Processes*, MacMillan Publishing Company, New York (1986).

Gaskill, J. D., *Linear Systems, Fourier Transforms and Optics*, John Wiley and Sons, New York (1978).

Gillespie, A. R., "Digital Techniques of Image Enhancement", *Remote Sensing in Geology*, B. S. Siegal and A. R. Gillespie, Eds. , John Wiley and Sons, New York (1980).

Goldman, S., *Information Theory*, Prentice-Hall, New York (1953).

Graybill, F. A., *Matrices with Applications in Statistics*, Wadsworth Publishing Company, Belmont (1983).

Greenberg, M. D., *Foundations of Applied Mathematics*, Prentice-Hall, Englewood Cliffs (1978).

Greenleaf, J. F., "Computed tomography from ultrasound scattered by biological tissues", *SIAM-AMS Proceedings*, **14** pp. 53– 63 (1984).

Grenander, U. and G. Szegö, *Toeplitz Forms and Their Applications*, Cambridge University Press, London (1958).

Habibi, A. and P. A. Wintz, "Image coding by linear transformation and block quantization", *IEEE Transactions on Communication Technology*, **19**, pp. 50–62 (1971).

Hamming, R. W., *Digital Filters*, Prentice-Hall, Englewood Cliffs (1983).

Harris, J. L., "Image evaluation and restoration", *Journal of the Optical Society of America*, **56**, pp. 569– 574 (1966).

Helstrom, C. W., "Image restoration by the method of least squares", *Journal of the Optical Society of America*, **57**, pp. 297-303 (1967).

Huang, T. S., "The subjective effect of two-dimensional pictorial noise", *IEEE Transactions on Information Theory*, **11**, pp. 43–53 (1965).

Huang, T. S. and P. M. Narendra, "Image restoration by singular value decomposition", *Applied Optics*, **14**, pp. 2213–2216 (1975).

Huck, F. O., C. L. Fales, N. Halyo, R. W. Samms and K. Stacy, "Image gathering and processing: Information and fidelity", *Journal of the Optical Society of America*, **2**, pp. 1644- 1666 (1985).

Hughes Aircraft Santa Barbara Research Center, *Thematic Mapper: Design Through Flight Evaluation, Final Report* (October 1984).

Huhns, M. N., "Optimum restoration of quantized correlated signals", *University of Southern California Image Processing Institute Report No. 600* (August 1985).

Hunt, B. R., "A matrix theory proof of the discrete convolution theorem", *IEEE Transactions on Audio and Electroacoustics*, **19** pp. 285–288 (1971).

Hunt, B. R., "A theorem on the difficulty of numerical deconvolution", *IEEE Transactions on Audio and Electroacoustics*, **20**, pp. 94–95 (1972).

Hunt, B. R., "The application of constrained least squares estimation to image restoration by digital computer", *IEEE Transactions on Computers*, **22**, pp. 805–812 (1973).

Hunt, B. R., "Digital image processing",*Proceedings of the IEEE*, **63**, pp. 693–708 (1975).

Hunt, B. R., "Bayesian methods in nonlinear digital image restoration", *IEEE Transaction on Computers*, **26**, pp. 219–229 (1977).

Hunt, B. R., "Nonstationary statistical image models (and their application to image data compression)", *Computer Graphics and Image Processing*, **12** pp. 173–186 (1980).

Jobert, G. and A. Cisternas, "On the inverse problem of local seismic foci", *Lecture Notes in Physics, Vol. 85: Applied Inverse Problems.* P. C. Sabatier, Ed., Springer-Verlag, Berlin (1978).

Kashet, B. G. and A. A. Sawchuk, "A survey of new techniques for image registration and mapping", *Proceedings of the SPIE*, **432**, pp. 222–239 (1983).

Kretzmer, E. R., "Statistics of Television Signals", *The Bell System Technical Journal*, **31**, pp. 751–763 (1952).

Lahart, M. J., "Local image restoration by a least-squares method", *Journal of the Optical Society of America*, **10**, pp. 1333–1339 (1979).

Lillestrand, R. L., "Techniques for change detection", *IEEE Transactions on Computers*, **21**, pp. 654–659 (1972).

Lovitt, W. V., *Linear Integral Equations*, McGraw-Hill, New York (1924).

Maeda, J., "Image restoration by an iterative damped least- squares method with non-negativity constraint", *Applied Optics*, **24**, pp. 751–757 (1985).

Maeda, J. and K. Murata, "Image restoration by an iterative regularized pseudoinverse method", *Applied Optics*, **23**, pp. 857–861 (1984).

Malaret, E., L. A. Bartolucci, B. F. Lozano, P. E. Anuta and C. D. McGillem, "Landsat-4 and Landsat-5 Thematic Mapper data quality analysis", *Photogrammetric Engineering and Remote Sensing*, **LI**, pp. 1407–1416 (1985).

Markham, B. L., "Characterization of the Landsat Sensors' Spatial Responses", *NASA Technical Memorandum 86130*, (July, 1984).

Markham, B. L., "The Landsat Sensors' Spatial Responses", *IEEE Transactions on Geoscience and Remote Sensing*, **23** pp. 864–875 (1985).

Max, J., "Quantizing for minimum distortion", *IRE Transactions on Information Theory*, **6**, pp. 7–12 (1960).

Mikhlin, S. G., *Integral Equations*, Pergamon Press, London (1957).

Mikhlin, S. G., *Linear Integral Equations*, Hindustan Publishing Corporation, Delhi (1960).

O'Neal, J. B., "Predictive quantizing systems (differential pulse code modulation) for the transmission of television signals", *The Bell System Technical Journal*, **5**, pp. 689–721 (1966).

Oppenheim, A. and R. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs (1975).

Panter, P. F. and W. Dite, "Quantization distortion in pulse- count modulation with nonuniform spacing of levels", *Proceedings of the IRE*, **39**, pp. 44–48 (1951).

Papoulis, A., *Probability, Random Variables and Stochastic Processes*, McGraw-Hill Book Co., New York (1965).

Papoulis, A. *Probability, Random Variables and Stochastic Processes*, McGraw-Hill Book Co., New York (1984).

Park, S. K. and R. A. Schowengerdt, "Image Reconstruction by Parametric Cubic Convolution", *Computer Vision, Graphics and Image Processing*, **23**, pp. 258–272 (1983).

Park, S. K., R. Schowengerdt and M. Kaczynski, "Modulation- transfer-function analysis for sampled image systems", *Applied Optics*, **23**, pp. 2572–2582 (1984).

Phillips, D. L., "A technique for the numerical solution of certain integral equations of the first kind", *Journal for the Association of Computing Machinery*, 9, pp. 84-97 (1962).

Pratt, W. K., "Vector space formulation for two-dimensional signal processing operations", *Computer Graphics and Image Processing*, **4**, pp. 1–24 (1975).

Rabiner, L. R., "Techniques for designing finite-duration impulse-response digital filters", *IEEE Transactions in Communications Technology*, **19**, pp. 188–195 (1971).

Rabiner, L. R. and B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs (1975).

Rauchmiller, R. F. and R. A. Schowengerdt, "Measurement of the Landsat Thematic Mapper MTF Using a 2-Dimensional Target Array", *Proceedings of the SPIE*, **697**, August, 1986.

Ready, P. J. and P. A. Wintz, "Information extraction, SNR improvement and data compression in multispectral imagery", *IEEE Transactions on Communications*, **21**, pp. 1123–1130 (1973).

Reid, J. K., *Large Sparse Sets of Linear Equations*, Academic Press, London (1971).

Richardson, W. H., "Bayesian-based iterative method of image restoration", *Journal of the Optical Society of America*, **62**, pp. 55–59 (1972).

Robinson, A. H., *Elements of Cartography*. John Wiley and Sons, New York (1960).

Roe, G. M., "Quantizing for minimum distortion", *IEEE Transactions on Information Theory*, **10**, pp. 384–385 (1964).

Rosenfeld, A. and A. C. Kak, *Digital Picture Processing*, Vol. 1, Academic Press, New York (1982).

Sasaki,O. and T. Yamagami, "Image restoration by iterative estimation of the expansion coefficients of an object in singular vector space", *Optics Letters*, **10**, pp. 433–435 (1985).

Schowengerdt, R. A., *Techniques for Image Processing and Classification in Remote Sensing*, Academic Press, New York (1983).

Schowengerdt, R. , C. Archwamety and R. C. Wrigley, "Operational MTF for Landsat Thematic Mapper", *Proceedings of the SPIE*, **549**, pp. 110–118 (1985a).

Schowengerdt, R. A., C. Archwamety and R. C. Wrigley, "Landsat Thematic Mapper Image-Derived MTF", *Photogrammetric Engineering and Remote Sensing*, **51**, pp. 1395–1406 (1985b).

Shlien, S., "Geometric correction, registration, and resampling of Landsat imagery", *Canadian Journal of Remote Sensing*, **5**, pp. 74–89 (1979).

Slater, P. N., *Remote Sensing: Optics and Optical Systems*, Addison-Wesley, Reading, MA (1980).

Smith, K. T., "Inversion of the x-ray transform", *SIAM-AMS Proceedings*, **14**, pp. 41–52 (1984).

Stewart, G. W., *Introduction to Matrix Computations*, Academic Press, New York (1973).

Strang, G., *Linear Algebra and its Applications*, Academic Press, New York (1980).

Taranik, J. V., "Characteristics of the Landsat Multi-Spectral Data System", *USGS Open-File Report*, #78-187 (1978).

Tewarson, R. P., *Sparse Matrices*, Academic Press, New York (1973).

Trussel, H. J. and B. R. Hunt, "Sectioned methods for image restoration", *IEEE Transactions on Acoustics, Speech and Signal Processing*, **26**, pp. 157–164 (1978).

Twomey, S., *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*, Elsevier Scientific Publishing Co., Amsterdam (1977).

Watanabe, S., "Karhunen-Loève expansion and factor analysis. Theoretical remarks and applications", *Transactions of the 4th Prague Conference on Information Theory*, pp. 635–660 (1965).

Van Wie, P. and M. Stein, "A Landsat digital image rectification system", *IEEE Transactions on Geoscience Electronics*, **15**, pp. 130–137 (1977).

Wintz, P. A., "Transform picture coding", *Proceedings of the IEEE*, **60**, pp. 809–820 (1972).

Wood, R. C., "On Optimum Quantization", *IEEE Transactions on Information Theory*, **15**, pp. 248–252 (1969).