A UNIFIED SUPERRESOLUTION APPROACH FOR OPTICAL AND SYNTHETIC APERTURE RADAR IMAGES

By

FRANK M. CANDOCIA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1998
To Yudit
ACKNOWLEDGEMENTS

It is easy to feel overwhelmed by the amount of time and effort required of the Ph.D. degree. These feelings were, often times, too familiar. With patience, persistence and the help of many individuals, my doctoral studies provided for a very memorable and fruitful four years. I would like to thank these people for their help and friendship throughout my studies.

First, I wish to acknowledge and thank Dr. Jose C. Principe for his role as advisor and mentor throughout my doctoral studies. Our discussions and exchanging of ideas were fundamental in the shaping of this work. I would also like to thank him for providing a stimulating work environment in CNEL through which my engineering horizons were expanded. Thank you to my supervisory committee members: Dr. John M. M. Anderson, Dr. A. Antonio Arroyo, Dr. John G. Harris and Dr. Janice C. Honeyman – your interest and suggestions were appreciated. Dr. John G. Harris deserves additional thanks, particularly for our fruitful discussions and his suggestion to me to work on optical images. I would like to thank Dr. Leslie Novak of MIT Lincoln Laboratory for providing me the opportunity to work with him during the summer of 1997. It was a truly rewarding experience.

My colleagues in and around CNEL also deserve recognition for their help and friendship. These include Victor Brennan, Craig Fancourt, John Fisher, Chen-Hsien Wu,
Dongxin Xu and Li-Kang Yen. Thanks for your insight into all the issues and questions I came to you with.

I wish to give a special thanks to my parents and the Rodriguez family. All of the pastelitos you sent from Miami were a definite morale boost. Your love, encouragement, and support are indelibly etched in my memory.

Last, but certainly not least, I wish to thank Yudit. Words cannot fully describe how special you have been. Your love, support, patience and sacrifice have been immeasurable.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTERS</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2 AN APPROACH TO SUPERRESOLUTION</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Introducing the Problem</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Comments and Observations</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Motivation and Description of the Superresolution Architecture</td>
<td>15</td>
</tr>
<tr>
<td>3 EXISTING APPROACHES TO RECONSTRUCTION OF OPTICAL AND SYNTHETIC APERTURE RADAR IMAGES</td>
<td>22</td>
</tr>
<tr>
<td>3.1 Optical Image Interpolation with a Single Kernel</td>
<td>23</td>
</tr>
<tr>
<td>3.1.1 Common Kernels</td>
<td>24</td>
</tr>
<tr>
<td>3.1.2 Increasing the Sample Density</td>
<td>28</td>
</tr>
<tr>
<td>3.1.3 Interpolation as Projections</td>
<td>29</td>
</tr>
<tr>
<td>3.2 Review of Reconstruction for Optical Images</td>
<td>30</td>
</tr>
<tr>
<td>3.3 SAR Imaging via Nonparametric Spectral Estimation</td>
<td>42</td>
</tr>
<tr>
<td>3.3.1 Phase History Data</td>
<td>43</td>
</tr>
<tr>
<td>3.3.2 Standard Approaches</td>
<td>45</td>
</tr>
<tr>
<td>3.3.3 Minimum Variance Approaches</td>
<td>46</td>
</tr>
<tr>
<td>3.3.4 Estimating a Full Rank Autocorrelation Matrix</td>
<td>49</td>
</tr>
<tr>
<td>3.4 Review of Reconstruction for Synthetic Aperture Radar Images</td>
<td>51</td>
</tr>
<tr>
<td>4 SUPERRESOLUTION OF OPTICAL IMAGES</td>
<td>58</td>
</tr>
<tr>
<td>4.1 Optical Image Acquisition Model</td>
<td>59</td>
</tr>
<tr>
<td>4.2 Using Local Information</td>
<td>61</td>
</tr>
<tr>
<td>4.2.1 Relation Between Correlation and Euclidean Distance</td>
<td>61</td>
</tr>
</tbody>
</table>
4.2.2 Existence of Locally Correlated Information in Images ... 66
  4.2.2.1 Interblock Correlation ........................................ 67
  4.2.2.2 Scale Interdependencies ........................................ 71
4.3 Architecture Specifics ................................................ 79
4.4 Training Phase ......................................................... 83
  4.4.1 The Preprocessing .................................................. 83
    4.4.1.1 Decimation .................................................... 84
    4.4.1.2 Neighborhood Extraction ..................................... 84
  4.4.2 Hard Partitioning the Input Space ................................ 86
    4.4.2.1 Kohonen’s Self-Organizing Feature Map ...................... 87
    4.4.2.2 Cluster Formation ............................................. 88
  4.4.3 Associative Memories ............................................. 89
    4.4.3.1 Least Mean Squares .......................................... 89
    4.4.3.2 Multilayer Perceptrons ....................................... 93
    4.4.3.3 Association of Neighborhoods ................................. 94
4.5 Reconstruction Phase .................................................. 97
4.6 Relation to the Mixture of Experts .................................. 99

5 EXPERIMENTAL RESULTS FOR OPTICAL IMAGES .......................... 103
  5.1 Analysis and Comparison ............................................. 103
    5.1.1 Linear Associative Memories ................................... 106
      5.1.1.1 Standard Images ........................................... 106
      5.1.1.2 Texture Images ............................................. 119
      5.1.1.3 Additional Results and Comments ............................ 124
    5.1.2 Nonlinear Associative Memories .................................. 129
    5.1.3 Mixture of Experts .............................................. 134
  5.2 Feature Extraction .................................................... 139
    5.2.1 Image Features .................................................. 139
    5.2.2 Correlated Image Structure ..................................... 142

6 SUPERRESOLUTION OF SYNTHETIC APERTURE RADAR IMAGES ...................... 144
  6.1 Imaging with Multiple Models ....................................... 145
  6.2 Architecture Specifics .............................................. 146
  6.3 Models Used .......................................................... 148
  6.4 The Methodology ...................................................... 151
    6.4.1 The Imaging Criteria ........................................... 151
    6.4.2 Selecting Between Models ....................................... 152
    6.4.3 The Superresolution Process .................................... 153
  6.5 Experimental Results ................................................ 155
    6.5.1 MSTAR Slicy Data ................................................. 155
    6.5.2 MSTAR Target Data ............................................... 157
    6.5.3 Detection Performance ........................................... 158
Abstract of Dissertation Presented to the Graduate School
Of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

A UNIFIED SUPERRESOLUTION APPROACH FOR OPTICAL AND
SYNTHETIC APERTURE RADAR IMAGES

By
Frank M. Candocia

May 1998

Chairman: Dr. Jose C. Principe
Major Department: Electrical and Computer Engineering

This work addresses the issue of superresolving optical and synthetic aperture
radar (SAR) images. Superresolution is the process of obtaining an image at a resolution
higher than that afforded by the sensor used in the imaging. The concept of signal
resolution is shown to be intimately related to the notion of perfect reconstruction. As
such, the major issue addressed in this work is in establishing an appropriate set of bases
for the reconstruction of images in these domains in terms of a finite set of collected
samples.

The existing theoretical foundations for perfect reconstruction in the
aforementioned domains is expressed in terms of linear projections and infinite extent data.
However, practical restrictions leading to finite collected data limits the resolution
afforded by the theoretically established bases – the sinc bases in the spatial domain and the Fourier bases in the frequency domain. Superresolution deals with this issue by incorporating a priori information into the process of establishing an appropriate set of projections for reconstruction.

In the optical domain, a priori information is extracted locally in an unsupervised manner from low and high resolution versions of digital images of the same scene. This is in agreement with the local weighting of the theoretically established sinc bases for perfect reconstruction. In the SAR domain, a priori assumed models are inherent to the reconstruction. The information is provided globally in terms of a best match to these models. This is also in agreement with the global weighting of the theoretically established Fourier bases for perfect frequency reconstruction.

The superresolution of images is accomplished by means of a modular architecture using a two-step approach. First, each image neighborhood is assigned to a priori determined clusters for the image class. Second, each cluster is linked to its optimal reconstructor which was trained from the collected set of samples. The results obtained herein are compared against several accepted sets of bases. Optical images of real scenes as well as textures are tested. SAR data from real metallic polyhedral objects as well as U.S. military tanks are imaged and tested for automatic target detection performance. It is shown that by appropriate incorporation of a priori information into the superresolution procedure, better reconstructed images result.
CHAPTER 1
INTRODUCTION

The intrinsic resolution limits of a sensor can typically be attributed to its physical properties. One way of increasing this resolution is to physically alter the sensor. Another approach would involve the appropriate signal processing of collected data. The work herein is concerned with the latter within the specific domains of optical and synthetic aperture radar (SAR) images. Superresolution is the term given to the signal processing of collected information which achieves a resolution higher than the one afforded by the physical sensor. Applications of this work include obtaining high quality image prints from digital images, increased detection of targets in military or civilian imaged scenes and detection of small tumors in medical imaging.

Superresolution of optical images is generally viewed as seeking a perfect reconstruction. As such, a comparison to the interpolation of signals is natural and warranted. In sampling theory we can define the resolution of a signal as the minimum number of samples sufficient to completely represent our signal [Vetterli and Kovacevic, 1995]. To this end, one important result regarding the perfect reconstruction of a signal from its samples comes from the Shannon sampling theory [Marks, 1991; Oppenheim and Schafer, 1989]. In this theory, the original signal can theoretically be reconstructed from its samples provided certain conditions regarding the signal and its samples are met – irrespective of the statistics of the signal. If these conditions are violated, then common
engineering knowledge says that we can never recover information beyond half the sampling frequency. In fact, this is not necessarily so. A priori knowledge regarding the original signal can be used to optimally recover a signal from its samples.

Superresolution of SAR images is generally viewed as the problem of detectability of complex sinusoids embedded in nonsinusoidal signals ("noise"). The desire to locate radar scattering elements (triangular reflectors) implies the detection of these complex sinusoids. It is the detection of these reflectors upon which SAR imaging is presently based. Having said this, we can define the resolution of a signal in the SAR domain as the minimum number of samples sufficient to distinguish between the two complex sinusoids in that signal that are closest in 2D frequency. This resolution, or spectral resolution as it is also referred, is established in the frequency domain and is determined by the main lobe of the rectangular window [Kay, 1988; Stoica and Moses, 1997].

In SAR imaging, the superresolution of objects other than triangular reflectors would benefit the automatic target detection and recognition (ATD/R) problem and it is the focus of the superresolution approach presented herein. The phase history of a typical radar illuminated scene contains, among other things, reflected energy from nonmetallic objects such as trees and open field. A priori knowledge of objects within the scene as well as how they ideally manifest themselves in the phase history can be used to enhance the performance of automatic target detection and recognition (ATD/R) systems.

The issue of superresolution is generally viewed as different in the aforementioned image domains. I submit that it can be regarded as treating the same problem – the perfect reconstruction problem – in each image’s respective domain which is space for optical images and 2D frequency for SAR images. Owing to this, an architecture for the
superresolution of images is described in this work which unifies the superresolution problem in both domains. An image is superresolved based on several modules. Each module is “tuned” to the specific character of the local image content at the point to be reconstructed. In order to determine the appropriate module for the point being imaged, a simple clustering on the signal samples is performed.

Chapter 2 introduces and describes the superresolution problem more formally. The goal of the chapter is to motivate the superresolution architecture presented at the chapter’s end. With this in mind, issues regarding the practical limitations of the perfect reconstruction of a signal from a set of samples are addressed and several observations are made regarding the mitigation of the problems faced.

Chapter 3 serves to present methods in common use for the reconstruction of a signal from its samples in both the optical and SAR image domains. It is these methods that the superresolution procedure of this study is compared against.

Chapter 4 concentrates on the details regarding the superresolution of optical images. Here we discuss the reasons for our choice of modules within the architecture presented. It is shown here that the architecture is implementing a convolution with a family of kernels. This is equivalent to convolving with a shift-varying finite impulse response (FIR) filter where the filter weights are determined according to the local image characteristics.

Chapter 5 is dedicated to experimental results regarding the superresolution of optical images. The results obtained are analyzed in detail and comparisons to the results of the reconstruction approaches presented in Chapter 3 are provided. Among the results analyzed are the use of linear associative memories and nonlinear associative memories in
our superresolution methodology as well as a comparison to the performance of the mixture of experts architecture.

Chapter 6 reports on the methodology for superresolving SAR images. The superresolution is based on two models: the point scatter or trihedral reflector model for targets and a noise model for clutter. The superresolution here is also a shift-varying FIR filtering scheme. The filter weights are based on minimum variance estimation, that is, they are dependent on the minimization of expected image power subject to how well the assumed model resided in the SAR phase history samples. Experimental results with this method are reported which increase automatic target detection performance over the conventional methods.

Finally, Chapter 7 concludes this work with comments regarding the material presented herein. The contributions of this work are stated and a commentary on future research directions is given.
CHAPTER 2
AN APPROACH TO SUPERRESOLUTION

The resolution of a signal addresses the ability to perfectly reconstruct a continuous signal from a set of samples. It is dependent on two things: the sampling period for obtaining the samples and the number of samples collected. Accomplishing this perfect reconstruction with a universal set of linear projections defines signal resolution. If the conditions for perfect reconstruction from this universal set are not satisfied, then the reconstruction can only be approximated. Having said this, superresolution is the ability to reconstruct a signal (from a set of samples) beyond the limit imposed by sampling theory.

This chapter introduces a methodology for the superresolution of images. We begin by introducing and defining the concepts of signal resolution. Then comments and observations on the practical limitations associated with perfect reconstruction are presented. The chapter concludes with a motivation for the superresolution of images via the methodology presented in this work as well as a description of the architecture which implements it.

2.1 Introducing the Problem

Webster’s dictionary defines resolution as the process of separating into constituent or elementary parts [Webster, 1989]. Implied in this definition is that, ideally, all of the elementary parts completely describe the object they are derived from. This
work is concerned with the elementary “parts” of signals and the ability to fully describe them from a set of samples. As such, we are necessarily addressing the issue of perfect signal reconstructability. In this sense, resolution is synonymous with reconstruction. This is now elucidated.

Let \( x(\alpha) \), where \( -\infty < \alpha < \infty \), be a continuous signal with maximum frequency content \( \Omega_c \) rad/sec. Thus our analysis is based on bandlimited signals. We can represent \( x(\alpha) \) as a linear combination of an orthogonal set of basis functions as

\[
x(\alpha) = \sum_n y[n]k_\alpha[n]
\]

where the linear weighting is given by the samples in \( y[n] \) and \( k_\alpha[n] \) represents our set of basis functions. Eqn. (2-1) addresses the perfect reconstruction of \( x(\alpha) \) from a set of samples. For the sake of simplicity, our development is performed with 1D signals. The extensions to 2D should be clear.

For a signal of time, \( x(t) \), the perfect reconstruction of bandlimited signals from a set of samples is addressed in the Shannon sampling theory [Marks, 1991; Oppenheim and Schafer, 1989]. Here \( y[n] = x(nT_s) \) for all integers \( n \), that is \( -\infty < n < \infty \), and sampling period \( T_s \). The equation describing this perfect reconstructability is

\[
x(t) = \sum_{n=\infty}^{\infty} x(nT_s)\text{sinc}\left(\frac{t}{T_s} - n\right)
\]

where, by definition, \( \text{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \). We see that our basis functions are given by \( k_\alpha[n] = \text{sinc}(\frac{nT_s}{T}) - n \) and the basis functions in this infinite set are orthogonal [Woodward, 1964], that is,

\[
\int_{-\infty}^{\infty} \text{sinc}(\frac{nT_s}{T} - n) \text{sinc}(\frac{mT_s}{T} - m) dt = T_s \delta[n - m].
\]
The perfect reconstruction can be obtained if the sampling period satisfied $T_s < \frac{T_c}{2}$ where $T_c = \frac{2\pi}{\Omega_c}$. For time signals, $\frac{T_c}{2}$ is the critical sampling rate (also known as the Nyquist rate). Therefore, every sample of $x(t)$ can be exactly resolved with an infinite set of samples provided the sampling period is sufficient.

For a signal in frequency, $x(\Omega)$, the perfect reconstruction of bandlimited signals from a set of samples is addressed by the discrete-time Fourier transform (DTFT). The set of samples $y[n] = y(nT_s)$ comes from the sampling of a superposition of complex exponentials. The equation describing this perfect reconstructability is

$$x(\Omega) = \sum_{n=-\infty}^{\infty} y(nT_s) e^{-j\Omega n} \quad (2-3)$$

Here we see that our basis functions are given by $k_{\Omega}[n] = e^{-j\Omega T_s n}$ and the infinite set of these basis functions are orthogonal, that is,

$$\int_{-\infty}^{\infty} e^{j\Omega n} e^{-j\Omega m} d\Omega = 2\pi \delta[n-m].$$

The perfect reconstruction can be obtained if the sampling period satisfied $T_s < T_c$ where $T_c = \frac{2\pi}{\Omega_c}$ as before. For frequency signals, $T_c$ is the critical sampling rate.

We have seen that every instant of our bandlimited time and frequency signals, $x(t)$ and $x(\Omega)$ respectively, can be exactly recovered (resolved) with an infinite set of samples provided the sampling period is sufficient. Thus the sampling period $T_s$ provides the limit to which our signals can be perfectly reconstructed (resolved) from an orthogonal set of linear projections and an infinite number of samples. It is in this manner that the sinc and Fourier bases are the common and universal set of linear projections capable of perfectly reconstructing bandlimited signals in the time (space) and frequency domains, respectively.
These sets of bases are universal in that all appropriately sampled infinite extent bandlimited signals can be reconstructed with these bases.

For finite extent data, eqns. (2-2) and (2-3) can be expressed respectively as

\[ \hat{x}(t) = \sum_{n=-\infty}^{\infty} [x(nT_s)w[n]] \text{sinc}\left(\frac{t}{T_s} - n\right) \quad (2-4) \]

and

\[ \hat{x}(\Omega) = \sum_{n=-\infty}^{\infty} [y(nT_s)w[n]]e^{-j\Omega T_s n} \quad (2-5) \]

where \(w[n]\) describes samples of our window function

\[ w[n] = \begin{cases} 1; & 0 < n < N - 1 \\ 0; & \text{otherwise} \end{cases} \]

and \(N\) is the extent of our window. Notice that the hat character in \(\hat{x}\) has been used purposely to denote the approximation afforded by a finite number of data samples. The finite number of data reduces the resolvability of the signal. We can see this by examining eqns. (2-4) and (2-5) in the frequency domain. The continuous Fourier transform (FT) of eqn. (2-4) yields

\[ \hat{X}(\Omega) = \frac{T_s}{\Omega_s} [X(\omega) \otimes W(\omega)] \Pi\left(\frac{\Omega}{\Omega_s}\right) \]

and eqn. (2-5) can be equivalently expressed as

\[ \hat{X}(\Omega) = \frac{1}{\Omega_s} [Y(\omega) \otimes W(\omega)] \]

where \(\otimes\) is the periodic convolution operator, \(\omega = \Omega T_s\) is angular frequency in radians, \(Y(\omega)\) is the DTFT of \(y(nT_s)\) and is periodic of period \(2\pi\) in \(\omega\) and \(\frac{2\pi}{T_s}\) in \(\Omega\) (similarly for \(X(\omega)\) and \(W(\omega)\)) and \(\Pi(\Omega) = \{ 1 \mid \Omega < \frac{1}{T_s}; 0 \text{ otherwise} \}\). The effect of the windowing function \(w[n]\) on eqns. (2-4) and (2-5) is to smear (distort) the true frequency...
spectrum $X(\Omega)$ of $x(\alpha)$. This results in a decrease in the ability to properly resolve the signal $x(\alpha)$.

Thus the resolution of a signal is also limited by the number of samples we have to describe the signal. The object of superresolution is to be able to reconstruct $x(\alpha)$ more faithfully than the resolution afforded by eqns. (2-4) and (2-5), that is, the resolution afforded from a finite set of observed data obtained at a sampling rate $T_s$.

### 2.2 Comments and Observations

Practical issues may preclude the satisfaction of the conditions necessary for perfect signal reconstruction as given in the previous section. In particular, (1) there must be no noise associated with the collected samples (e.g. no quantization error), (2) the sampling rate must be less than the critical sampling rate of the signal and (3) the signal must have been sampled to infinite extent.

In the practical sampling of SAR data, the issue concerning sampling rate is usually not critical. This is because the bandwidth and frequencies at which the radar operates are known at the time of data collection. The issue of noise is much more critical. For example, SAR data is typically imaged from a rectangular grid of equispaced radar returns. This data, however, is sometimes processed by interpolating through the actual collected return grid. This is particularly true for inverse SAR and spotlight mode SAR data collection where a typical grid consists of radially equispaced samples swept out through several angular increments [Odendaal et al., 1994; Mensa, 1991]. As such, the accuracy with which frequency information from the imaged scene can be produced is reduced.
In the practical sampling of optical images, the issue concerning quantization error is usually not critical. The standard use of an 8-bit dynamic range usually yields highly acceptable and pleasing images. The issue of sampling frequency is much more critical. The information of a natural scene has typically very high spatial frequency content. The sharp contrast we perceive in order to delineate objects (object boundaries) as well as the textural character of those objects are but some of the attributes inherent to the high frequency information content of optical images. As such, the sampling frequency used in collecting optical images is generally not large enough to fully describe a “continuous image” in the sense of the Shannon sampling theory. An interesting attribute of optical images is their highly structured nature. This structure appears locally and can be used to characterize objects in these images, that is, portions of objects can be described as smooth, edgy, etc. Information such as this is not considered in the sampling theorem.

Let us now make a few observations. Eqns. (2-2) and (2-3) specify a universal set of basis functions which are linearly weighted by a set of collected samples corresponding to signal \( x(\alpha) \). There is an uncountable infinity of basis functions in this universal set (one for each instant of \( \alpha \) in our original signal to reconstruct) and they are given by

\[
 s_t = \{ \text{sinc} \left( \frac{t - n}{T_s} \right) \} \quad \infty < n < \infty, -\infty < t < \infty \}
\]

for spatial signals and, for frequency signals, by

\[
 s_\Omega = \{ e^{-j\Omega T_s n} \} \quad -\infty < n < \infty, -\infty < \Omega < \infty \}
\]

These sets are universal with respect to those signals that are sampled beyond their critical rates and the resolution afforded by these sets is given by this rate. Hence, if samples were collected at a sampling rate \( T_s \) that did not meet the critical sampling rate, the sets of sinc or Fourier basis functions could not be linearly weighted as in eqns. (2-2) and (2-3) to perfectly reconstruct our signal. This does not preclude the existence of other sets of basis functions which could be linearly weighted.
combined by samples collected at a rate below the critical sampling rate to still yield the signal’s perfect reconstruction.

As an example, consider the aliased spectrum of fig. (2-1). Here, the original signal’s scaled spectrum is given by the dashed triangular region centered at the origin. The original spectrum is

$$X(\Omega) = \begin{cases} 1 - \frac{|\Omega|}{\Omega_c}; & |\Omega| \leq \Omega_c \\ 0; & \text{otherwise} \end{cases}$$

The aliased spectrum of fig. (2-1) for $|\Omega| < \Omega_c$ when $\Omega_c < \Omega_s < 2\Omega_c$ is

$$X_s(\Omega) = \begin{cases} \frac{1}{T_s} - \frac{|\Omega|}{T_s \Omega_c}; & |\Omega| < \Omega_s - \Omega_c \\ \frac{2}{T_s} - \frac{\Omega_s}{T_s \Omega_c}; & \Omega_s - \Omega_c < |\Omega| < \Omega_c \end{cases}$$

and is given by the solid line. It can be easily seen that our contrived signal $X(\Omega)$ can be recovered by linear filtering (frequency multiplication) with the filter

$$K(\Omega) = \begin{cases} T_s; & |\Omega| \leq \Omega_s - \Omega_c \\ \frac{T_s \Omega_c}{2\Omega_c - \Omega_s} \left(1 - \frac{|\Omega|}{\Omega_c}\right); & \Omega_s - \Omega_c < |\Omega| \leq \Omega_c \\ 0; & |\Omega| > \Omega_c \end{cases}$$

as pictured in fig. (2-1), that is, $X(\Omega) = X_s(\Omega)K(\Omega)$. This example shows that the original signal can be recovered via a linear filtering even if it has been aliased as long as the aliased spectrum has no null frequencies and the original signal’s spectrum is known. In practice this information is not typically known making the recovery of “improperly” sampled signals usually a difficult task.
Figure 2-1. Recovering a signal from an aliased representation. Aliasing results when sampling below the Nyquist rate. Copies of the true signal spectrum are given by the dashed triangular regions. The aliased spectrum is given by the solid line. Here, the true signal can be recovered from the aliased signal by linear filtering (frequency multiplication) via the filter $K(\Omega)$ designed from a priori knowledge of the true spectrum.

A priori knowledge of the character of a signal, rather than its frequency spectrum, could also be used to appropriately reconstruct the signal from its observed samples. As a simple example, let’s consider the set of piecewise constant continuous time functions where each constant segment in the signals has duration $T$ seconds (e.g. signals quantized in time). An illustration of such a function is provided in fig. (2-2a). Note that this signal has infinite frequency content due to its piecewise constant nature.

If our observed samples were obtained by sampling this function every $T$ seconds, then clearly the convolution of a zero-order-hold kernel with the observed samples would be optimal for recovering the piecewise constant function, that is, it would yield perfect reconstruction even though the function in question was grossly undersampled according to the Shannon sampling theory. This convolution is pictured in fig. (2-2b).

The set of piecewise constant signals is not typically encountered in practice so the basis set resulting from the zero-order-hold kernel is of limited use.
Figure 2-2. Simple example illustrating how perfect reconstruction is possible when a priori knowledge of the relation between signal samples and its corresponding continuous function is known. (a) Piecewise constant continuous function grossly undersampled according to the Nyquist criterion, (b) Recovering the continuous function in (a) from its samples requires a simple convolution of the samples with a zero-order-hold kernel.

However, this very simple example illustrates that superresolution could be based on a priori knowledge about the signal to reconstruct given the observed samples — irrespective of the frequency content of the signal. It also illustrates the fact that perfect reconstruction of a signal according to the Shannon sampling theory only establishes sufficient conditions for the perfect reconstruction of signals characterized by frequency from their samples. If a priori knowledge of the signal corresponding to the observed samples is available, then this can be used to develop a technique for superresolving a signal. Therefore, superresolving images is directly associated with methods to acquire a priori information about the signal of interest.

Recently, optimal reconstruction of signals sampled below the Nyquist rate was proved possible by using a priori knowledge of the signal statistics [Ruderman and Bialek, 1992]. The statistics of the continuous signal were assumed known rather than any specific relation between the signal and its samples. In the results reported, the signal $x(t)$
was assumed to be bandlimited, Gaussian, zero-mean and stationary. The samples $x[n]$ from this signal were obtained via a point spread function $f(t)$ that was not an ideal sampling Dirac delta and these samples were corrupted by zero-mean, additive, independent and identically distributed (i.i.d.) Gaussian noise $\eta[n]$ of power $\sigma^2$ to yield the collected samples $y[n] = x[n] + \eta[n]$.

The authors obtained the same reconstruction filter when using two different optimization approaches: maximum likelihood and conditional average estimation. The optimal signal reconstruction was given by

$$ \hat{X}(\Omega) = K(\Omega)Y(\Omega) $$  \hspace{1cm} (2-6)

where the filter $K(\Omega)$ was

$$ K(\Omega) = \frac{F(\Omega)S(\Omega)}{\sigma^2 + T_\Omega \sum_{n=-\infty}^{\infty} |F(\Omega + n\Omega_s)|^2 S(\Omega + n\Omega_s)} $$

Notice that $Y(\Omega)$ is the DTFT of the sequence $y[n]$ so that $Y(\Omega) = Y(\Omega + n\Omega_s)$ is periodic with period $\Omega_s = \frac{2\pi}{T_s}$ and $T_s$ was the sampling period. $S(\Omega)$ is the power spectrum of $x(t)$ and $F(\Omega)$ is the FT of the point spread function $f(t)$. From eqn. (2-6), the most likely signal $\hat{x}(t)$ is obtained through a linear filtering of the collected samples $y[n]$.

In the case when the noise vanishes, the point spread function becomes an ideal sampler, the signal was bandlimited and the sampling frequency is above the Nyquist rate, that is, $\sigma^2 = 0$, $F(\Omega) = 1$, $X(\Omega) = 0$ for $|\Omega| > \Omega_c$ and $\Omega_s > 2\Omega_c$, the filter $K(\Omega)$ reduces to

$$ K(\Omega) = \begin{cases} 
T_s & |\Omega| \leq \Omega_c \\
0 & \text{otherwise}
\end{cases} $$
which is the FT of the sinc function. This result shows that the signal statistics play no role in perfect reconstruction when the Shannon sampling conditions are met.

More importantly, this result shows that a priori knowledge of the statistics of a signal can be used to superresolve a signal from a collected set of samples. However, this analytic result is only valid for stationary Gaussian signal statistics and, in practice, signals are typically nonstationary and have very complex statistics. Our lack of knowledge of these statistics and the analytical intractability of determining the optimal filters for complicated density functions, as commented by the authors, limits the practical use of this method.

2.3 Motivation and Description of the Superresolution Architecture

The motivation for the superresolution methodology stems from our goal of more efficiently using the information contained in available image samples. This requires projections onto data specific sets of bases instead of the bases established in the reconstruction theories. Naturally, learning or adaptive system theories play a crucial role in this methodology of designing data specific projections. The manner in which these models are realized must be consistent with the information character of images and how this relates to the superresolution problem. These issues are now addressed.

The universality of the perfect reconstruction theories is an amazing result but the price paid is a strict limitation on the required resolution. The practical problem is to do the best we can with the available samples, that is, to superresolve the images. To yield better reconstructed images, we must find alternate sets of projections from which to reconstruct our images. It has been shown that perfect reconstruction can be achieved if a
priori knowledge is available, but in practice this knowledge is absent. So a central problem is how to capture a priori knowledge about the domain. In determining the set of projections to use, we must either make assumptions regarding our data or we must learn this a priori knowledge from the available data using nonparametric models. We are concerned with the latter.

We propose in this dissertation a novel technique of extracting a priori information from an image by performing a down sampling operation on the original image. The high resolution image becomes the desired response to a learning system that receives the low resolution image as input. The highly structured and localized nature of images can be exploited in modeling the relation between low and high resolution versions of an image. It seems reasonable to consider several local models in the imaging process. This is due to the various structures present in images resulting from the objects in the imaged scene. Two particular image traits can be used in this modeling:

- there is much similarity in local structure throughout an image
- this structure is typically maintained across scales

**Local structure.** A direct example of the first trait in optical images can be provided by considering an image of a person’s face. Local neighborhoods about the person’s cheeks and forehead are generally indistinguishable when viewed independently. We have assumed that the effects of lighting and other “special” attributes (scars, moles, birth marks, etc.) are absent in this comparison. An easy method to test this observation is to locate these similar image portions in an image and randomly swap them to form a new image. If the new image resembles the original one then our observation is correct. In this manner, all neighborhoods exhibiting a particular characteristic can be treated in
practically the same manner. These neighborhoods can be considered generated by the same statistical process. It will be shown that the models which we compare our neighborhoods against can be interpreted as the mean of each statistical process – one for each model used.

Examples of the existence of this first trait abound in images. It has recently been exploited to increase gains in compression schemes [Dony and Haykin, 1995a; Kambhatla and Leen, 1997]. The use of image compression has been popular for some time now and is evidenced by the widespread JPEG still image compression standard [Wallace, 1991] and PCA compression schemes [Dony and Haykin, 1995b]. The standard schemes for lossy image compression are based around the highly redundant information generally present in small image blocks which could be described in a more efficient and compact manner. In the case of compression via PCA, a single representation is established for all of the blocks in the image. The recent compression approaches exploit the first image trait by grouping the small blocks of an image into clusters that are most similar (correlated) with one another. In this way, each cluster can be described with an efficient representation of its own – separate from those corresponding to other clusters. Overall, this results in a more efficient image representation than that afforded by the approaches of the standard compression schemes.

The first image trait can also be seen in SAR images – but in the frequency domain. It results from the radar backscattering of objects of similar composition and orientation. SAR imaging typically, however, is based solely on the backscatter model of the trihedral reflector. This reflector is ideally imaged as a bright point in a SAR image. However, the formation of SAR images should be based on a variety of models. In this
manner, objects exhibiting a particular backscatter characteristic can be imaged with the appropriate model.

**Across scale similarity.** The second trait can be used in obtaining a high resolution image from a low resolution one. To accomplish this we must have knowledge of the relation between similar neighborhoods in a low resolution image and their higher resolution counterparts. In optical images, this relation can be established from available data using the down sampling operation. This is done by optimally associating homologous low and high resolution neighborhoods. In SAR images, the neighborhood relation is in the frequency domain. However, establishing the relation between SAR data and the a priori models used in the imaging is performed in the phase history domain. This is done by minimum variance model matching between the phase history data and the model used for the point being imaged.

The manner in which we superresolve images exploits the aforementioned traits. Superresolving based on these two traits necessitates the need for establishing

- which neighborhoods of an image are similar in local structure
- how these neighborhoods relate across scale

The superresolution architecture of fig. (2-3) operates on knowledge of these two premises. It exploits the relations between collected samples in establishing the basis functions for reconstruction since the conditions for perfect reconstruction of the sampling theorem cannot be met in practice. The neighborhood selection is established by a comparison of the distance between the received data and learned local models representative of the data. This is performed by the analysis and decision block of fig.
This comparison is used to select the module most appropriate for the reconstruction of a particular image point from the \( C \) available modules.

The reconstruction modules in the architecture for optical images are associative memories (optimal mappings) which produce the output pattern best associated with the input. They describe the basis functions for the reconstruction of image samples. They are established via training from available data of homologous low and high resolution neighborhoods. As such, the basis functions are highly localized in space. This is because the information most contributing to the character of an image sample resides about that sample. It is in agreement with the support properties of the sinc kernel in the sampling theorem. For SAR images, the modules of the architecture are nonparametric model estimators. The basis functions derived from these modules are global in extent. This affords the most frequency resolution in the reconstructed image. They are also in agreement with the global extent of the basis functions of the FT.

One can expect that this methodology will yield better reconstruction than methods based on the sampling theorem. However, unlike the universal character of the sampling theorem, this superresolution method is specific to the character of images, that is, bases obtained for one class of images may perform poorly when reconstructing another class. Because of this, establishing the appropriate models with which to compare our data against is important to the successful superresolution of an image. With the appropriate models established, the collected data is transformed via the projections corresponding to each model. A recent approach in the neural network literature that is used in modeling the probability density function (PDF) of a set of data conditioned on another set of data is the mixture of experts (MOE) [Jacobs et al., 1991; Jacobs and Jordan, 1991]. The
conditional PDF in the MOE is modeled as a mixture Gaussian distribution. This architecture can be used in modeling the transformation between low and high resolution neighborhoods of images. The MOE will be compared to our superresolution methodology. Comparisons of the theoretical issues between the MOE and our approach as well as superresolution performance will be provided in a later chapter.

Note that the reconstructed images from the architecture of fig. (2-3) are not continuous but are a set of samples at a higher density than the collected samples. Further details regarding this architecture for each image domain will appear in their respective chapters.
Figure 2-3. The general architecture for the superresolution of optical and SAR images.
CHAPTER 3
EXISTING APPROACHES TO RECONSTRUCTION OF OPTICAL AND SYNTHETIC APERTURE RADAR IMAGES

This chapter presents background on common approaches to image reconstruction from collected samples. To address the perfect reconstruction issue, we require knowledge of the ideal image to be formed. The ideal optical image typically describes functionally the reflected light at each spatial location from the camera’s viewpoint and within its field of view. This function describes the imaged environment as perceived by the human eye. The ideal SAR image does not have such a clear description. Among the possibilities, the ideal SAR image could represent the environment as perceived by the human eye (without the effects of atmospheric occlusions such as clouds, fog, precipitation, etc.). It could also describe functionally the electromagnetic reflectivity at each spatial location from the radar’s viewpoint and within its “field of view” such that closely spaced scattering elements in the illuminated scene could be discriminated. In fact, SAR imaging currently attempts to address both of these issues.

A description and formulation of the common approaches for reconstruction of optical and SAR images from collected samples is presented in this chapter. Also included herein are reviews of the recent work on the reconstruction and superresolution of optical and SAR images. The information presented in this chapter is the platform upon which the details of the superresolution methodology for both optical and SAR images will be developed in the appropriate chapters. These chapters will show that the common
approaches presented here are special cases of the superresolution methodology of this work.

3.1 Optical Image Interpolation with a Single Kernel

A common manner in which optical images are reconstructed from a set of collected samples is by interpolation with a single 2D kernel. The process of interpolation was alluded to in the previous chapter when discussing perfect reconstruction via convolution with the sinc kernel. We will formally introduce the problem here for the sake of clarity. The interpolation methods of this section utilize the most commonly encountered 2D kernels. Interpolation is seen to take the form of a convolution between the observed image samples and a 2D kernel. The samples of an analog image \( x(t_1, t_2) \) are denoted by \( x[n_1, n_2] \equiv x(t_1, t_2)|_{t_1=n_1T_1, t_2=n_2T_2} \). Each interpolated point \( \hat{x}(t_1, t_2) \) in these techniques is linearly related to samples of the image in question. The procedures of this section for interpolating continuous images take the form

\[
\hat{x}(t_1, t_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} x[n_1, n_2]k\left(\frac{t_1}{T_1} - n_1, \frac{t_2}{T_2} - n_2\right) \tag{3-1}
\]

where \( k(t_1, t_2) \) is the continuous interpolation kernel that defines the weighting of the signal samples and \( T_1, T_2 \) denote the sampling periods in their respective axes. The kernels that are presented here are prototypes, that is, \( T_1 = T_2 = 1 \). A simple scaling by the appropriate sampling rate, as in eqn. (3-1), yields the proper kernel.

There are two important properties which are common to interpolation. They will be referred to as interpolation properties 1 and 2. These are:
1. \( \hat{x}(t_1, t_2) \) must pass through the image samples when \( t_1 \) is an integer multiple of \( T_1 \) and \( t_2 \) is an integer multiple of \( T_2 \). This is mathematically equivalent to constraining the interpolation kernel as follows

\[
k(t_1, t_2) = \begin{cases} 
1 & t_1 = 0, t_2 = 0 \\
0 & t_1 = m_1 T_1, t_2 = m_2 T_2; m_1, m_2 \in \mathbb{Z}
\end{cases}
\]

where \( \mathbb{Z} \) is the set of integers [Hamming, 1973].

2. Signal sample weighting diminishes as the distance from the point being interpolated increases. This property constrains the kernel’s envelope to be a decreasing function as we move away from its center. If the sample weighting does not diminish, then it is finite within a “small” support region.

Property 1 is a consequence of the definition of interpolation. That is, samples are only constructed between the given sample locations since, according to the Shannon sampling theorem, the observed samples must be free of quantization error and noise for perfect signal reconstruction. Property 2 is usually invoked by drawing a parallel with the Shannon sampling theory. If samples of a function are drawn at small enough intervals, then it is known that any function value between these samples can be reconstructed with the use of an asymptotically decaying function – the sinc function. Finite extent kernels do not necessarily need to adhere to this. Specifying finite sample weighting within a small region of support is sufficient for computational numerical stability.

3.1.1 Common Kernels

We will now present and briefly describe four common kernels particular to image interpolation. These are the zero-order-hold, bilinear, bicubic and sinc kernels.
**Zero-Order-Hold.** In general, this is the simplest of interpolation kernels. Interpolation with this kernel is commonly referred to as nearest neighbor interpolation or replication. Using this kernel, the interpolated point \( \hat{x}(t_1, t_2) \) is chosen as \( x[n_1, n_2] \) at the signal sample closest to \((t_1, t_2)\). The interpolation kernel is a “rectangular box” given by

\[
k(t_1, t_2) = \begin{cases} 
1 & |t_1| < \frac{1}{2}, |t_2| < \frac{1}{2} \\
0 & \text{otherwise}
\end{cases}
\]  

(3-2)

The region of support here is only one sample in extent. The interpolated image usually has a very “blocky” appearance due to the repeating of image samples. This kernel satisfies *property 1* but is not in strict compliance with the decay concept of *property 2*. The signal sample weighting does not diminish with distance, rather it is constant only within a small rectangular region. This interpolation kernel is illustrated in fig. (3-1).

**Bilinear.** Interpolation with this kernel is the 2D counterpart of 1D linear interpolation. The bilinear interpolation kernel is separable and is composed of two 1D linear interpolation kernels as shown in the equation below

\[
k(t_1, t_2) = \begin{cases} 
(1-|t_1|)(1-|t_2|) & |t_1| < 1, |t_2| < 1 \\
0 & \text{otherwise}
\end{cases}
\]  

(3-3)

The region of support for this method is local. Its support region is comprised of the four samples nearest the point of interpolation. It satisfies both interpolation properties listed above. This interpolation kernel is illustrated in fig. (3-1).

**Bicubic.** Unlike the interpolation kernels of eqns. (3-2) and (3-3), this one is smooth. Interpolation with this kernel locally fits a separable polynomial through our sample points. The kernel is constrained so that the polynomial’s first derivative must
agrees at the endpoints of adjacent polynomials. The 1D prototype kernel developed to meet such a criterion for the bicubic kernel is

\[
k(t) = \begin{cases} 
  \frac{1}{2}(t+2)^2(t+1) & -2 \leq t < -1 \\
  -\frac{3}{2}(t+1)(t^2 + \frac{2}{3}t - \frac{2}{3}) & -1 \leq t < 0 \\
  \frac{1}{2}(t-1)(t^2 - \frac{2}{3}t - \frac{2}{3}) & 0 \leq t < 1 \\
  -\frac{1}{2}(t-1)(t-2)^2 & 1 \leq t < 2 \\
  0 & \text{otherwise}
\end{cases}
\]  

(3-4)

The separable 2D kernel can be easily generated by considering the sampling rates \( T_1 \) and \( T_2 \) of each axis in the above equation, scaling appropriately, and multiplying the resulting 1D kernels. Properties 1 and 2 both hold for the above interpolation kernel. The separable 2D kernel is shown in fig. (3-1).

**Sinc.** This kernel was established in the Shannon sampling theory. It is well known that any bandlimited signal can be reconstructed from its samples provided that the signal was sampled at a rate higher than the Nyquist rate. Here the interpolation kernel is given by

\[
k(t_1, t_2) = \text{sinc}(t_1) \text{sinc}(t_2) = \frac{\sin(\pi t_1)}{\pi t_1} \frac{\sin(\pi t_2)}{\pi t_2}
\]  

(3-5)

and the region of support is global. This is given by the extent of the kernel which is seen to be infinite due to the sinusoids in the kernel. Note that the weighting function \( k(t_1, t_2) \) in this case is separable and corresponds to an ideal low pass filter. This ideal kernel is illustrated in fig. (3-1).

Notice that each of these interpolation kernels can be used in the perfect reconstruction of a continuous signal from its samples. Convolving a kernel with the signal samples results in perfect reconstruction when the continuous signal is known (or
assumed) to be locally constant for the zero-order-hold kernel, locally linear between
samples for the bilinear kernel, locally smooth up to the first derivative for the bicubic
kernel and globally smooth up to a given frequency of $\frac{f_s}{2}$ for the sinc kernel.

Figure 3-1. Common interpolation kernels. The kernels depicted here are specified in
eqns. (3-2)-(3-5). All are separable and eightfold symmetric ($T_1 = T_2$). Note that all
kernels are of finite extent except the sinc kernel.

A typical property, though not essential, of common interpolation kernels is the
separability of the kernel itself. This property usually simplifies the task of kernel
composition and analysis and could also be useful for efficient implementation of the
interpolation process. Symmetry is also an attribute usually associated with interpolation
kernels. In the case of 2D interpolation, the symmetry is usually fourfold, that is,
\[ k(t_1, t_2) = k(-t_1, t_2) = k(t_1, -t_2). \]

If the sampling rate is the same in each of the two independent axes, this fourfold symmetry becomes eightfold. In this case, \[ k(t_1, t_2) = k(-t_1, t_2) = k(t_1, -t_2) = k(t_2, t_1). \] These additional properties are shared by the kernels presented in this section. Kernels of finite extent necessarily use local information in the construction of a point whereas kernels of infinite extent are required to use the global content of the given signal samples.

### 3.1.2 Increasing the Sample Density

The continuous image obtained via the linear filtering of eqn. (3-1) must be discretized to view or store it on digital computer. This discrete representation allows interpolation to be viewed as the process of increasing the sampling density of a signal. In this case, the interpolated signal can be obtained by expanding the given signal samples \( x[n_1, n_2] \) and convolving with a sampled interpolation kernel [Schafer and Rabiner, 1973]. This form of interpolation is a linear filtering process used in interpolating discrete images.

For an expansion rate of \( G_1 \times G_2 \), our expanded signal to convolve is given by

\[
x_e[n_1, n_2] = \begin{cases} 
  x\left[\frac{n_1}{G_1}, \frac{n_2}{G_2}\right] & n_1 = 0, \pm G_1, \pm 2G_1, \ldots \\
  0 & \text{otherwise} 
\end{cases}
\]

and the corresponding sampled interpolation kernel, obtained by sampling a continuous kernel, would be \( k[n_1, n_2] = k(t_1, t_2)\big|_{t_1 = \frac{n_1}{G_1}, t_2 = \frac{n_2}{G_2}}. \) \( G_1, G_2 \) are both integers greater than 1.

The interpolated signal \( \tilde{x}[n_1, n_2] \) at our new sampling density is

\[
\tilde{x}[n_1, n_2] = x_e[n_1, n_2] * k[n_1, n_2].
\]

Eqn. (3-7) is the discrete counterpart of eqn. (3-1). It will be shown that eqn. (3-7) is a special case of the superresolution methodology for optical images.
3.1.3 Interpolation as Projections

The interpolation paradigm can be cast in several different forms. The form most common in the signal processing literature is that given in eqns. (3-1) and (3-7). It is cast as the linear filtering (convolution) of the signal samples with an interpolation kernel (FIR filter). An equivalent and convenient form of expressing the interpolation operation is via matrix multiplication. This form is compact and directly accessible by the optical image superresolution methodology.

We define a collection of 2D image samples to be a matrix: \( X = [x_{n_1, n_2}] \) where \( x_{n_1, n_2} = x[n_1, n_2] \equiv x(t_1, t_2) \bigg|_{t_1=n_1 T_1, t_2=n_2 T_2} \). Similarly, \( \hat{X} \) is a matrix of interpolated samples, that is, samples at the new sampling density. Notice that \( \hat{X} = [\hat{x}_{n_1, n_2}] \) where \( \hat{x}_{n_1, n_2} = \hat{x}[n_1, n_2] \) and these are estimates of \( x(t_1, t_2) \big|_{t_1=n_1 T_1', t_2=n_2 T_2'} \) for \( T_1' = T_1 / G_1 \) and \( T_2' = T_2 / G_2 \). Note that \( G_1, G_2 > 1 \) when increasing the sampling density, that is, interpolating. For separable interpolation kernels we define two interpolation kernel matrices, \( K_1 \) and \( K_2 \), as

\[
K_1 = \begin{bmatrix}
\vdots & \vdots & \vdots & \vdots \\
\cdots & k(0,0) & k(-1,0) & k(-2,0) & \cdots \\
\cdots & k(1/G_1, 0) & k(0, -1) & k(0, -2) & \cdots \\
\cdots & k(2/G_1, 0) & k(1, -1) & k(1, -2) & \cdots \\
\vdots & \vdots & \vdots & \vdots
\end{bmatrix} \\
K_2 = \begin{bmatrix}
\vdots & \vdots & \vdots & \vdots \\
\cdots & k(0,0) & k(0, -1) & k(0, -2) & \cdots \\
\cdots & k(1/G_2, 0) & k(1, -1) & k(1, -2) & \cdots \\
\cdots & k(2/G_2, 0) & k(2, -1) & k(2, -2) & \cdots \\
\vdots & \vdots & \vdots & \vdots
\end{bmatrix}
\]

It is readily seen that eqn. (3-1), with \( t_1 = p_1 T_1' \), \( t_2 = p_2 T_2' \) and \( p_1, p_2 \in Z \), is equivalently expressed as

\[
\hat{X} = K_1 X K_2^T.
\] (3-8)

Notice that if \( G_1 = G_2 \), then \( K_1 = K_2^T \). If the interpolation kernel is of finite extent then \( K_1 \) and \( K_2 \) are also finite extent matrices.

Another useful form of expressing eqns. (3-7) and (3-8) results when representing the image samples in \( X \) as a vector rather than as a matrix. In this case, we draw from the
results of Kronecker products in linear algebra to establish our expression [Graham, 1981]. The Kronecker product $A \otimes B$ of any two matrices $A$ and $B$ is the matrix obtained from the matrix $A$ by replacing each entry $a_{ij}$ of $A$ with the matrix $a_{ij}B$. If $A$ is any $M \times N$ matrix and $a_j$ is its $j$th column then $\text{vec}(A)$ is the $MN \times 1$ vector

$$\text{vec}(A) = [a_1^T, a_2^T, \cdots, a_N^T]^T.$$  

Thus the $\text{vec}$ operation transforms a matrix into a column vector by placing the columns of the matrix one underneath the other. Let us also define an operator $u\text{vec}$ which undoes the $\text{vec}$ operation. That is, the $u\text{vec}$ operation transforms a column vector into a matrix by partitioning the vector into adjacent sections of equal length and placing them side by side to form a matrix. A result from linear algebra which is useful to us here is that $\text{vec}(ABC^T) = (C \otimes A)\text{vec}(B)$. With this, eqn. (3-8) is equivalently expressed in vector form as

$$\text{vec}(\hat{X}) = (K_2^T \otimes K_1)\text{vec}(X).$$  

(3-9)

Notice that in this expression the equivalent of one interpolation kernel is being used on a vector of data.

### 3.2 Review of Reconstruction for Optical Images

Work related to the notion of reconstruction has been described in the literature as zooming, magnification, enlargement, resampling, sample expansion, interpolation and superresolution. The foundation of this work, leading up to the present, lies in interpolation theory. Below is a brief exposition of the core of classical interpolation methodologies. This is followed by a comprehensive account of the recent work into the reconstruction and superresolution of optical images of the last decade or so.
Interpolation has an extensive history and has been applied in a variety of fields. An early use of such procedures was to determine function values between mathematically tabulated data [Hamming, 1973; Kreyszig, 1993]. Interpolation has also found use in the design of car and ship bodies. More recently, interpolation has found applications in image coding and video compression [Lim, 1990; Cramer et al., 1996]. The object of classical interpolation is to find polynomials that fit “well” through sample points in the data set. These polynomials generally fall into one of two categories: unconstrained and constrained polynomials. Procedures such as Lagrange interpolation and Newton’s divided difference are classic examples of the former type of interpolating polynomials [Kreyszig, 1993]. These two approaches only make use of the given data samples. The polynomials can be fit through the data locally or globally. In either case, a polynomial of degree \( N - 1 \) must be used in order to fit through \( N \) data samples (this applies to the case of one-dimensional interpolation). Spline interpolation is an example of the latter. The name is derived from thin elastic rods, called splines, which engineers used to fit smooth curves through given sample points. Today it finds use in the design of car and ship bodies – as mentioned earlier. It is a procedure whereby contiguous chunks of data samples are each fit with a polynomial. Further, the derivative(s) at the endpoints of adjacent polynomials are required to agree [Hamming, 1973]. This constraint results in a smooth polynomial fit through the data set. The resulting interpolation is generally not as oscillatory between data samples as compared with an unconstrained polynomial fit [Kreyszig, 1993].

It has been observed that the quality of interpolation is not generally dependent on the polynomial order of the interpolating functions [Kreyszig, 1993]. In many cases, it is
sufficient to impose that the first and/or second derivatives of the interpolation kernel exist. The smoothness of interpolating with splines over polynomials that lack derivative constraints is most noticeable in regions of abrupt change. Splines tend to limit the amount of oscillation following an abrupt change.

An important result in interpolation theory resulted from the works of Borel, Whittaker, Nyquist, Kotel’nikov and Shannon [Marks, 1991]. These men introduced what is known today as the Shannon-Whittaker-Kotel’nikov Sampling Theorem. The fundamental result of the Shannon sampling theory, as it is referred to in this work, is that a bandlimited signal is uniquely specified by its sufficiently close equally spaced samples. In practice, interpolating through a set of samples using the sinc function of the Shannon sampling theory is not possible. Instead, it is necessary to integrate other functions and/or approaches into the interpolation process. This case has been extensively and effectively treated for the case of bandlimited signals sampled at a rate exceeding the Nyquist rate [Oppenheim and Schafer, 1989; Roberts and Mullis, 1987]. It is known that oversampling can be utilized in easing the constraints on the interpolation function and for the recovery of the original signal to highly acceptable levels of accuracy [Hauser, 1991]. This is evidenced by the compact digital audio technology used today.

The interpolation work that is particular to images used the aforementioned polynomials a great deal up into the early 1980s. The 1D interpolation techniques were extended to the two dimensions corresponding to images by simply considering a separable interpolation polynomial. The most used interpolation polynomials, or kernels, consisted of either zero, first or third order polynomials. These were the zero-order-hold (nearest neighbor), bilinear and bicubic kernels [Netravali and Haskell, 1995]. The zero-
order-hold and bilinear kernels do not produce a smooth transition between image samples. The bicubic is better in that its polynomial does have a continuous first derivative. This leads to a smoother transition between interpolated samples. The work of Hou and Andrews [1978] considered the use of a cubic B-spline as the interpolating kernel and Keys [1981] used a cubic spline. An analysis and comparison of these image interpolation methods is given by Parker et al. [1983].

There have been recent studies into the development of the theory of interpolating kernels [Lucke, 1993; Lanzavecchia and Bellon, 1994, 1995; Schaum, 1993; Appledorn, 1996; Unser et al., 1992; Knoll, 1991]. These works focus on the developing of an "optimal" single interpolation kernel as a substitute for the sinc kernel. The kernels to develop are local. The assumption made in these works is that the signal has been sampled, at least, at the Nyquist rate. As mentioned earlier, the observed sample sequence is of finite extent, hence, the sinc kernel cannot be practically implemented. In particular, kernels have been composed of trigonometric functions [Lucke, 1993; Lanzavecchia and Bellon, 1994 and 1995], linear sums of Gaussian functions [Appledorn, 1996], polynomial B-splines [Unser et al., 1992] and trigonometric as well as low order polynomial functions [Schaum, 1993]. It is interesting to note that some of the developed local kernels approach the sinc function in the limit of an infinite data sequence [Unser et al., 1992; Lucke, 1993]. Interpolation filters for restoring highly subsampled images have also been studied [Knoll, 1991]. The method is based on the iterative minimization of an error weighting function in a constrained $m$-dimensional coefficient space.

The use of a single linear interpolation kernel has proven to be very popular for image interpolation for two main reasons: (1) the design of linear kernels (kernels that
linearly combine sample points) is a mathematically tractable problem and (2) their implementation is nothing more than a convolution – for which there exist fast algorithms for their implementation via the fast Fourier transform [Oppenheim and Schafer, 1989]. However, there are important considerations as to why these methods might not be best for reconstructing images. Images are typically characterized by the objects contained within the imaged scene. The figural continuity of objects which leads to an “instantaneous” transition of measured light intensity at objects’ boundaries is probably the most prevalent characteristic particular to optical images. It is this abrupt change in reflected light which conveys the most information about a viewed scene. Other striking characteristics are the textures and light reflections from the objects which comprise the scene. All of this manifests itself as high spatial frequency information which is to be imaged. The observed image samples from a scene typically cannot represent well the scene in the sense of the Shannon sampling theory, that is, the aforementioned high spatial frequencies are aliased due to an insufficient sampling rate – which comes from practical restrictions to the physical sensor.

Recent trends in the literature show an effort in developing algorithms to better cope with this problem. There has been a shift from single kernel approaches, which cannot recover aliased and/or filtered high spatial frequencies – to research effort which attempt to account for them. The methodologies developed thus far for addressing this problem can be roughly categorized as follows: (1) directional/edge based, (2) multiple kernel based, (3) orthogonal transform based and (4) variational/statistically based. The theoretical and practical issues characteristic of these methodologies are by no means mutually exclusive. As such, numerous algorithms cannot strictly be categorized as listed
Directional interpolation techniques recognize that high detail areas in images most often have a definite geometric structure or pattern, such as in edge regions [Lee and Paik, 1993; Algazi et al., 1991; Bayrakeri and Mersereau, 1995; Marsi et al., 1996; Jensen and Anastassiou, 1995]. In such cases, interpolation in the low frequency direction (along the edge) is much better than interpolation in the high frequency direction (across the edge). Thus, a directional interpolation scheme has to perform a local analysis of the image structure first and then base the interpolation on the local structure if a low frequency direction does exist. The algorithm by Lee and Paik uses an adaptive combination of zero-order-hold and moving average filters. By combining these two kernels the authors show that an adaptive version of the fast B-spline algorithm is obtained. The work by Algazi et al. focuses on high contrast directional structures such as edges, thin lines and corners. The authors use rank order statistics to detect the presence of these local image structures. They do this because they state that each of these local image structures is characterized by a bimodal distribution. Jensen and Anastassiou have a method that first detects the presence of a high contrast edge then estimates its location and orientation. If an edge is detected, it fits an ideal step edge in the region. If no edge was detected the algorithm reverts to bilinear interpolation. The algorithm by Bayrakeri and Mersereau considers weighted directional interpolation. The interpolated values along various directions are combined using directional weights which depend on the variation in that direction. The interpolation value for each direction is assigned based on the magnitude of its directional derivative. Marsi et al. create a simple edge-sensitive interpolation filter. It is a separable
filter with one adjustable parameter. The parameter is selected adaptively according to the sample values neighboring the point to be interpolated. The filter is shown to default to a linear interpolator when this adjustable parameter equals zero. It should be noted that these algorithms tend to be computationally cheap.

Multiple kernels for interpolation analyze local image statistics in order to determine an appropriate interpolation kernel for the region under consideration [Wang and Mitra, 1988; Winans and Walowit, 1992; Darwish and Bedair, 1996; Mahmoodi 1993]. The choice of kernels are typically ad hoc and typically two to three classic interpolation kernels are employed by these algorithms. The actual interpolation kernel used in each of these techniques can be interpreted as a shift-varying kernel whose form is dictated by an image’s local structure. The work of Wang and Mitra bases its interpolation on three local models: constant, oriented and irregular. The constant model is used to describe a block whose intensity variation is negligible to the human eye. The oriented model represents a block where the intensity variation is along a particular direction and the irregular model is used on blocks with more than one pronounced direction present (such as corners). Winans and Walowit base their work on a cubic spline kernel with an adjustable parameter. The authors preset this parameter to one of three values, hence resulting in three interpolation kernels. The parameter is chosen according to the edge information in the neighborhood of the interpolated pixels. Darwish and Bedair also use three interpolation kernels: the zero-order-hold, a cubic spline and an additional kernel established by the authors. All three are separable. The zero-order-hold kernel is used in edge regions, the cubic spline is used in homogeneous (smooth) regions and their third kernel is used in regions that are not classified as smooth or edgy.
Mahmoodi describes a method for using two interpolation kernels. A fifth order kernel with a 2×2 region of support is used in high contrast regions while a cubic spline with a 4×4 region of support is used otherwise. In order to select a kernel, a threshold is checked between the ratio of the sample deviation of the 2×2 and the 4×4 block around the point to be interpolated.

The use of orthogonal transforms for interpolation [Koltracht et al., 1996] has focused on using the discrete cosine transform (DCT) [Martucci, 1995; Shinbori and Takagi, 1994; Hong et al., 1996] and the wavelet transform [Keshavmurthy et al., 1996; Chang et al., 1995]. Work involving the wavelet transform utilizes information across image scales to predict information at finer scales. This exploits the fact that evolution across resolution scales characterizes the local regularity of a signal [Mallat and Zhong, 1992]. The work of Martucci performs the interpolation completely in the DCT domain. Previous work by the author has shown that it is possible to use the DCT to implement digital filters [Martucci, 1993]. As such, all spatial domain filtering is equivalently performed in the DCT domain. Shinbori and Takagi use an iterative application of the Gerchberg-Papoulis (GP) algorithm with the DCT to obtain a magnified image. This technique uses the GP algorithm as the basic principle for extending the frequency band. The spatial high frequency components are restored during the forward and inverse iterative transform process for the image by DCT – using two constraints that the spatial extent of the image is finite and correct information is already known for the low frequency components. Hong et al. use the coefficients of a locally performed DCT to extract edge information. Five different edge types are identified and each type has a corresponding interpolation technique. Once the edge type is determined, the zero-order-
hold and bilinear interpolations are performed and combined. Then five different Gaussian low pass filters adaptively reduce the discontinuities which resulted from this aforementioned interpolation combination. The wavelet based work by Keshavmurthy et al. decomposes the image into an overcomplete (or redundant) representation via an autocorrelation shell derived from a Daubechies wavelet. The wavelet detail information across known scales is used to extrapolate the details at finer scales. Chang et al. implement a method that extrapolates the wavelet transform of the higher resolution image based on the evolution of the wavelet transform extrema across scales. By identifying three constraints which the higher resolution information needs to obey, they enhance the reconstructed image through alternating projections onto the sets defined by these constraints. Koltracht outlines a method based on parametric families of orthogonal transforms which arise from the wavelet multiresolution paradigm. The idea’s development considers the original image as a low frequency component of a family of potential enlargements.

Interpolation based on variational and/or statistical principles presents a different view of interpolation than from those previously discussed [Karyiannis and Venetsanopoulos, 1991; Saul and Jordan, 1996; Schultz and Stevenson, 1994; Ruderman and Bialek, 1992]. Despite this, some interesting connections between these recent methodologies and some previously mentioned have been made. In fact, certain interpolation techniques previously discussed have been found to be special cases of these new methodologies. The variational methods formulate the interpolation problem as the constrained minimization of a certain functional. The choice of the functional to be minimized varies and is based on a combination of the restrictions imposed by the
mathematical problem and certain considerations related to the physical problem. The work of Karyiannis and Venetsanopoulos concerns itself with the application of variational principles to the image interpolation problem. They formulated the problem as the constrained minimization of a quadratic functional, with particular emphasis on the class of quadratic functionals whose minimization amounts to 2D generalized L-splines. The specific functionals used were based on certain image models. It is stated in this paper that the previously discussed method of cubic splines results from the minimization of a certain quadratic functional. The work of Saul and Jordan incorporates variational and statistical principles for the interpolation problem. Their approach considers a multidimensional input and a model of its density. They don't specifically concern themselves with images but focus on how to optimally interpolate between two points in space. They do this by assigning a cost to each path through space, based on two competing goals – one to interpolate through regions of high density, the other to minimize arc length. They present an efficient solution (in high dimensions) to this problem for Gaussian, Dirichlet and mixture models. Schultz and Stevenson describe an approach which considers variational and statistical principles for image interpolation. Their methodology results in the optimization of convex functionals (not necessarily quadratic). A statistical model for the image was assumed which incorporates the convex Huber function. This function is quadratic up to a threshold and linear beyond this threshold. Use of this cost function helps maintain discontinuities from an image into its expanded version. The work of Ruderman and Bialek shows that if the probability density of a signal is known a priori, this knowledge can be used to dealias information contained in samples of that signal. In fact, this knowledge can be used to estimate frequencies
beyond the Nyquist limit. They posed the superresolution problem as one of finding the “best” estimate of the original signal given the original signal’s PDF and the observed signal samples. The most useful estimator will depend on the costs for making different kinds of errors. The two natural choices for an estimator are maximum likelihood (ML) and the conditional average. Their observed signal samples are assumed to be obtained by a general point spread function (PSF), not just sampled by a train of impulses, and corrupted by Gaussian noise. The noise for each sample is assumed to be statistically independent. The example presented in their work assumes the original signal is drawn from a Gaussian ensemble. Under these assumptions both estimators yield the same result. They show that the optimal superresolution procedure for a Gaussian signal results in a convolution of the signal samples with an interpolation kernel which is a function of the noise variance, the point spread function, the original signal power spectrum and the sampling period. They obtain a very satisfying result using this interpolation methodology. For the case of zero noise variance, a delta PSF and a bandlimited original signal that was sampled above the Nyquist rate, the interpolation kernel defaults to the sinc kernel.

There is other work that differs in approach from the above stated categories but is nonetheless noteworthy. The work of Cheung and Zeng only uses a single interpolation kernel. However, this kernel is derived using a least squares minimization process. The kernel derived has a four sample diamond like region of support [Cheung and Zeng, 1995]. Zeng and Venetsanopoulos use the median filter with local image features for their interpolation [Zeng and Venetsanopoulos, 1992]. The median filter is known to remove outliers in the local distribution of pixels and to preserve certain types of image structure [Huang and Tang, 1979]. They attempt to exploit this property of the median filter to
produce more accurate interpolation results in edge regions. Fathima and Yegnanarayana have used a maximum entropy approach to interpolation. They specifically address the problem of recovery of the missing samples of a signal from a few of its randomly distributed samples. They also discuss the appropriateness of their method for different distributions of the known samples [Fathima and Yegnanarayana, 1990]. Herodotou et al. have used a simple Gibbs random field model for interpolation. In their paper, a binary based Gibbs random field model is used for image interpolation. Images are interpolated from their downsampled versions along with a number of texture parameters that are estimated within smaller image blocks [Herodotou et al., 1995]. The use of wavelet bases for the interpolation of sparse data has been addressed by Pentland. This work introduces good approximate solutions for this problem in only a few iterative steps – where each step requires O(n) operations and storage locations. The author illustrates the use of this technique for interpolating sparse range data [Pentland, 1994]. Work by DeNatale et al. uses a least squares bilinear interpolation scheme based on the least squares optimization of a spline-like scheme. This work incorporates adaptive image sampling and interpolation for data compression [DeNatale et al., 1994]. Tom and Katsaggelos address the problem of obtaining a high resolution image from multiple low resolution images that have been subsampled and displaced by different amounts of subpixel shifts. The paper poses the low to high resolution problem as an ML problem which is solved by the expectation maximization (EM) algorithm. By exploiting the structure of the matrices involved, the problem can be worked on in the discrete frequency domain [Tom and Katsaggelos, 1995]. Typically, multiple displaced versions of a single image are not available for interpolation. For a video sequence, however, adjacent temporal frames tend
to satisfy this assumption particularly well. A recent book addresses this problem [Tekalp, 1995].

3.3 SAR Imaging via Nonparametric Spectral Estimation

The remainder of this chapter is devoted to the reconstruction of SAR images from the collected samples of the radar backscattering. This set of collected samples is typically termed the phase history corresponding to a SAR image. An introduction regarding the information provided by the phase history data will be presented in this section.

SAR images are commonly formed via spectral estimation techniques. The reason for this is elucidated herein. The "ideal" SAR image is typically assumed to discriminate between radar scattering elements in the illuminated scene. As such, the ideal image is considered a juxtaposition of delta functions in the frequency domain – one for each scattering element. The SAR imaging process is generally performed with this in mind. Owing to this, spectral estimation serves as the basis for SAR imaging.

Spectral estimation methods can be broadly classified as nonparametric or parametric. The nonparametric approaches generally perform a narrowband filtering over a band of frequencies. The parametric approaches postulate a model for the data and the parameters of this model must be estimated. The minimum variance (MV) superresolution approaches, to which our superresolution methodology belongs, postulate a local model for data to produce an optimal filter matched to that model. In a sense, this can be considered a semiparametric approach. Nonetheless, since it does produce a filter and model parameters are not estimated, it is considered in this work as a nonparametric approach.
3.3.1 Phase History Data

The phase history data samples collected from a radar which electromagnetically illuminates a scene are used in the formation of SAR images. This section presents an introduction to the information provided by the phase history samples. It is provided so that the unfamiliar reader might better understand the premises underlying the imaging of SAR data. Interested readers are referred to the books by Hovanessian [1980], Mensa [1991] and Skolnik [1990] for detailed explanations of the SAR data collection process.

The SAR phase history represents the reflected radar information in the downrange and crossrange (Doppler) directions. The downrange data is obtained by transmitting a stepped-frequency waveform in the downrange direction of known frequency step size and bandwidth. The range extent is a function of the frequency step size and the range resolution is a function of the total bandwidth of the stepped-frequency waveform. By varying the frequency step size and total bandwidth, arbitrary range depth and resolution can be achieved.

The reflected radar information in the downrange direction is contained in the composite wave received by the radar. It contains the shifted incident waves which resulted from the objects present in the radar illuminated scene. To obtain data in the crossrange direction, the complete stepped-frequency waveform is transmitted for equidistant intervals along the radar’s flight path. The crossrange resolution is a function of the transmit wavelength and the relative velocity between the radar and the scene. Along the cross range direction, objects also reflect the incident waves with different phase shifts.
The formation of a SAR image from this information involves determining the reflectivity function of objects in the radar illuminated scene in both the downrange and crossrange directions. The reflectivity at a given range is obtained by narrowband filtering at each of the frequencies transmitted by the radar. The filtered signal’s magnitude about the center frequency in the narrowband is an estimate of the reflectivity value we seek. The reflectivity function is typically estimated using the fast Fourier transform (FFT). This is done because, first, it is known that each of the basis functions of the FFT is a narrowband filter [Stoica and Moses, 1997] and, second, the reflectivity function at all frequencies of interest can be computed efficiently via the FFT. The FFT is typically used in the downrange and crossrange directions.

The reflectivity function for points in space, that is, for distances from the radar, are typically computed by taking the magnitude of the aforementioned FFT computation. It is known that this operation is one way to estimate the power spectral density (PSD) of a signal [Stoica and Moses, 1997]. In this manner, the formation of SAR images is a direct result of PSD estimation from phase history samples. The illuminated radar scene can also be assumed to consist of a finite number of prominent scattering elements. In this manner, the phase history is modeled as the superposition of 2D complex exponentials – one for each scattering element. The amplitudes of each complex exponential is proportional to the amount of scattering produced by each element. Spectral estimation techniques can also be used in detecting these scattering elements. Common approaches to spectral estimation (and hence SAR imaging) are now discussed.
3.3.2 Standard Approaches

Spectral estimators are used to detect complex sinusoids embedded in noise from a finite number of observations [Stoica and Moses, 1997]. In 1D, we denote the \( N \) observed signal samples by \( x[n] \), \( n = 0, \ldots, N-1 \); where \( x[n] = m[n] + e[n] \) and \( m_\omega[n] = ae^{j\omega n} \) is the complex sinusoid of frequency \( \omega \) and complex amplitude \( a \) which is additively embedded in noise \( e[n] \). In 2D, the complex sinusoid’s counterpart is the complex plane wave. These are denoted \( m_\omega[n_1, n_2] = ae^{j(\omega_1 n_1 + \omega_2 n_2)} \) where \( \omega = [\omega_1, \omega_2] \).

Here too, the coefficient \( a \) is complex and accounts for the amplitude and phase shift of each wave. The standard methods for detecting these waves (and consequent scattering elements in a SAR image) are based on the FT. This is due to the basis functions of the FT being the waves we seek to detect. The discrete space FT (DSFT) of a finite set of data is defined as

\[
X(\omega_1, \omega_2) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x[n_1, n_2] e^{j(\omega_1 n_1 + \omega_2 n_2)} \tag{3-10}
\]

where \( \omega_1 \) and \( \omega_2 \) are in radians. Thus the DSFT of the noisy signal serves to project this signal onto the complex plane waves we seek. The presence of complex plane waves in the noisy signal would yield large amplitudes at those frequencies of the DSFT corresponding to the waves present in the noisy signal relative to all other frequencies.

The ability to distinguish between several sinusoids spaced closely in frequency is limited by the number of data samples available. This was alluded to in the previous chapter. The resolution afforded is given by \( \Delta \omega_1 = \frac{2\pi}{N_1} \) and \( \Delta \omega_2 = \frac{2\pi}{N_2} \), that is, the approximate half-power (-3dB) width of the main lobe of the 2D rectangular window. In other words, two waves of 2D frequency \( (\omega_1^{(1)}, \omega_2^{(1)}) \) and \( (\omega_1^{(2)}, \omega_2^{(2)}) \), identified by the
superscripts (1) and (2) respectively, can’t be distinguished using a direct application of the DSFT if $\Delta \omega_1 < |\omega_1^{(1)} - \omega_1^{(2)}|$ and $\Delta \omega_2 < |\omega_2^{(1)} - \omega_2^{(2)}|$.

Side lobe artifacts which result from Fourier transforming a finite number of data can affect the ability to accurately detect the waves present in the noisy signal also [Stoica and Moses, 1997]. Side lobes resulting from the presence of several waves can also sum to create the appearance of a wave that is truly not present in the signal. These side lobes can be reduced using special window functions $w[n_i, n_2]$ where $n_i = 0,...,N_i - 1, (i = 1,2)$ [Harris, 1978]. The reduction of these sidelobes comes at the expense of wider main lobe width of the window function thus reducing the ability to distinguish between closely spaced waves. Because spectral estimation methods of this type, such as the Welch, Burg and Daniell method [Stoica and Moses, 1997], are not able to resolve these waves, they will not be explicitly discussed here. Instead, the interested reader is referred to the book by Stoica and Moses or Kay for more details [Stoica and Moses, 1997; Kay, 1988].

3.3.3 Minimum Variance Approaches

It has been discussed that the ability to distinguish (or resolve) between two closely spaced frequencies is limited by the number of observed samples. The MV approaches have the ability to resolve beyond this inherent resolution of our data. Hence, they are superresolution approaches. The MV approaches are so called because they seek to minimize the power from additive noise (typically zero-mean Gaussian noise) present in the observed signal while preserving the sinusoids sought. The power of this type of noise is known to be given by its variance [Stoica and Moses, 1997]. These methods are considered adaptive filtering methods which seek to minimize the output power of the filtered signal with respect to a given model. Specifically, let $X = [x_{n, n_2}] = x[n_i, n_2]$ and
\( \mathbf{W} = [w_{n_1,n_2}] = w[n_1,n_2] \) for \( n_i = 0,...,N_i - 1 \), \( i = 1,2 \), denote the matrix of our 2D phase history samples and filter coefficients, respectively. Let us also define the vectors associated with these samples as \( \mathbf{x} = \text{vec} (\mathbf{X}) \) and \( \mathbf{w} = \text{vec} (\mathbf{W}) \). Each output image sample \( s_q \) is defined by

\[
 s_q = \min_{\mathbf{w}_q} E[ |\mathbf{x}^H \mathbf{w}_q|^2 ] = \min_{\mathbf{w}_q} E[ \mathbf{w}_q^H \mathbf{x} \mathbf{x}^H \mathbf{w}_q ] = \min_{\mathbf{w}_q} \mathbf{w}_q^H \mathbf{R} \mathbf{w}_q
\]

(3-11)

where \( q = [q_1,q_2] \), \( q_i = 0,...,Q_i - 1 \), \( i = 1,2 \), defines the location on the frequency grid from which we form our SAR image and \( \mathbf{R} \) is the autocorrelation matrix of vector \( \mathbf{x} \). From this formulation, the filter \( \mathbf{w}_q \) is dependent on the point to image, hence the subscript \( q \). It must obviously be constrained or the solution to eqn. (3-11) is trivial, that is, \( \mathbf{w}_q = 0 \). This would not be useful to us. Several constraints can be placed in order to arrive at a meaningful solution. The constraints and corresponding solutions of a few of the most used MV superresolution techniques are now discussed.

**Capon's Method.** This is also referred to as the MV method. A single equality constraint is imposed here and it is known as the unit boresight constraint. Specifically \( \mathbf{w}_q^H \mathbf{m}_q = 1 \) where \( \mathbf{w}_q \) is the vector of complex filter coefficients which weights the phase history samples. The point scatterer model is \( \mathbf{m}_q = \text{vec}(\mathbf{M}) \) where the matrix \( \mathbf{M} = [m_{n_1,n_2}] = m_{n_1,n_2} e^{j(\omega_{n_1} n_1 + \omega_{n_2} n_2)} \) for \( n_i = 0,...,N_i - 1 \) and \( \omega_q = [\omega_{q_1}, \omega_{q_2}] \) where \( \omega_{q_i} = \frac{2\pi}{Q_i} q_i \), \( i = 1,2 \). It is the specific wave model we seek at frequency grid coordinates \( q = [q_1,q_2] \). This constraint allows for the model to pass unattenuated if it is indeed present in the noisy observed signal. For each output sample, the optimization problem is
min \( w_q^H R w_q \) subject to \( w_q^H m_q = 1 \). This optimization problem is readily solved with the use of Lagrange multipliers. The optimal filter in Capon’s method is given by [Stoica and Moses, 1997]

\[
\begin{aligned}
w_q &= -\frac{R^{-1}m_q}{m_q^H R^{-1}m_q} \\
&= -\frac{m_q^H R^{-1}m_q}{m_q^H R^{-1}m_q}
\end{aligned}
\quad (3-12)
\]

and the spectral estimate \( S \) which is our superresolved SAR image for phase history \( X \), is composed of the spectral estimates \( s_q \) at each frequency grid location. These are obtained by plugging eqn. (3-12) into eqn. (3-11) to yield

\[
\begin{aligned}
s_q &= \frac{1}{m_q^H R^{-1}m_q} \\
&= \frac{1}{m_q^H R^{-1}m_q}
\end{aligned}
\quad (3-13)
\]

and the resultant SAR image is described by \( S = [s_q] = s[q_1, q_2] \) for \( q_i = 0, \ldots, Q_i - 1 \), \((i = 1, 2)\).

**MUSIC and Eigenvector.** These two approaches follow the same approach taken by Capon’s method. Where they differ is in their use of the autocorrelation matrix. In order to more clearly differentiate sinusoids from the noise they are embedded in, these methods seek to drive the spectral estimate to infinity when there is an exact match between the model vector and its presence in the observed signal. This is accomplished by considering the signal and noise subspaces to be orthogonal. Specifically, if we consider the eigendecomposition of the autocorrelation matrix \( R \), then we can write

\[
R = U \Lambda U^H = \sum_m \lambda_m u_m u_m^H = \sum_{m_{signal}} \lambda_m u_m u_m^H + \sum_{m_{noise}} \lambda_m u_m u_m^H
\]

\[
\quad (3-14)
\]

where \( U = [u_1,...,u_M] \) is the matrix composed of eigenvectors \( u_m \) of matrix \( R \). The eigenvector (EV) method directly uses the noise subspace in eqn. (3-14) for its spectral
estimate. Then by substituting $R^{-1}$ in eqn. (3-13) with the “inverse” of the noise subspace, the EV spectral estimate is written

$$s_q = \frac{1}{m_q^H \left( \sum_{m} \lambda_m^{-1} u_m u_m^H \right) m_q}$$

The MUSIC method exploits the white noise model by replacing the measured noise eigenvalues with the average noise power $\sigma^2$. The spectral estimate for MUSIC is

$$s_q = \frac{1}{m_q^H \left( \sum_{m} \sigma^2 u_m u_m^H \right) m_q}$$

3.3.4 Estimating a Full Rank Autocorrelation Matrix

Many MV related methods require knowledge of the inverse of the autocorrelation matrix $R$ of our phase history samples. This matrix must be estimated from the available samples. In practice, we usually have one phase history observation. Given a matrix of phase history samples we can compute a full rank autocorrelation matrix by averaging subwindows of the given phase history samples.

Specifically, let $X = x[n_1, n_2]$ for $n_1, n_2 = 0, 1, \ldots, N-1$ be an $N \times N$ matrix of our phase history samples. Also, let $X_{ij} = x[i : i + P - 1, j : j + P - 1]$ be the subwindows of size $P \times P$ in matrix $X$. Note that there will be $(N - P + 1)^2$ subwindows of size $P \times P$ in matrix $X$ where $i, j = 0, 1, \ldots, N - P$. The autocorrelation matrix $R$ of the phase history samples in $X$ is defined as

$$R = E[xx^H]$$

where $x = vec(X)$. This matrix is typically estimated by
\[
\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{x}_n \mathbf{x}_n^H
\]

However, in the case of one phase history observation \((N = 1)\), \(\hat{\mathbf{R}}\) is not full rank. With the subwindows we’ve described, the autocorrelation matrix estimate would be

\[
\hat{\mathbf{R}} = \frac{1}{(N-P+1)^2} \sum_{i=0}^{N-P} \sum_{j=0}^{P-P} \mathbf{x}_{ij} \mathbf{x}_{ij}^H
\]  \quad (3-15)

where \(\mathbf{x}_{ij} = \text{vec} (\mathbf{X}_{ij})\).

The size of the subwindows given by \(P\) should be chosen so that they are as large as possible while also ensuring that \(\hat{\mathbf{R}}\) is full rank. It can be shown that the maximum possible \(P\) for which \(\hat{\mathbf{R}}\) can be full rank satisfies \(P < (N+1)/2\), where \(P\) must be an integer. The resolution of the estimated matrix \(\hat{\mathbf{R}}\) could be increased if we had more subwindows to average. This is done by also averaging ‘backward’ versions of the subwindows \(\mathbf{x}_{ij}\). The backward subwindows corresponding to subwindow \(\mathbf{x}_{ij}\) (in vector form) is given by \(\mathbf{Jx}_{ij}^*\) where * denotes conjugation and matrix \(\mathbf{J}\) is a permutation matrix with ones on the antidiagonal. The resultant matrix is the forward-backward averaged autocorrelation matrix. Here we have twice the number of subwindows to average compared with the ‘forward only’ autocorrelation matrix of eqn. (3-15) so the size of \(P\) can increase. The ‘forward-backward’ estimate of the autocorrelation matrix is then

\[
\hat{\mathbf{R}} = \frac{1}{2(N-P+1)^2} \sum_{i=0}^{N-P} \sum_{j=0}^{P-P} \mathbf{x}_{ij} \mathbf{x}_{ij}^H + \mathbf{Jx}_{ij}^* \mathbf{x}_{ij}^T \mathbf{J}
\]  \quad (3-16)

Now the maximum possible \(P\) for which \(\hat{\mathbf{R}}\) can be full rank satisfies \(P < (2 - \sqrt{2})(N+1)\) where \(P\) must be an integer. The use of subwindows necessarily decreases the resolution of our spectral estimate. But, in order to compute an MV spectral estimate, this must be done.
3.4 Review of Reconstruction for Synthetic Aperture Radar Images

SAR images are reconstructed from phase history data. This reconstruction can, in general, be described as model based. These models characterize the back scatter of radar energy reflected from known objects. By far, the trihedral reflector is the most commonly considered model in SAR image formation. This is because the reflected energy of the trihedral reflector manifests itself as a complex plane wave (a 2D complex sinusoid) in the phase history of the radar return and as such, techniques based on the ubiquitous FFT can be used in the imaging process of signals in the SAR domain. SAR imaging based on other models is, to the author’s knowledge, very scarce. Imaging based on trihedral as well as dihedral corner reflectors has been studied [Liu and Li, 1996]. These authors also state that dihedral feature extraction for imaging has, to their knowledge, never been considered. This being the case, the remainder of this section introduces methods for the imaging of SAR data based on the trihedral reflector model. Some of these methods were discussed in the previous section but are included here for a more coherent review.

Ideally, when Fourier transforming the phase history, the trihedral reflector would appear as a point in the image corresponding to the true reflector’s location in space. However, the finite number of collected samples (yielding the truncation of the complex plane wave), tends to smear or distort the actual location of the object [Stoica and Moses, 1997]. Another undesirable effect in SAR image formation via Fourier transform is the presence of side lobes corresponding to each reflector in the image. This is a well-known consequence of Fourier transforming a truncated sinusoidal wave. It is known that these side lobes can be reduced or suppressed through the use of appropriately designed windows at the expense of increased main lobe width [Harris, 1978].
The problem in the superresolution of SAR imagery is one of correctly detecting the presence of the models which the objects in the imaged scene are assumed to be composed of. Typically, the superresolved image has increased in sample density, decreased in main lobe width (relative to the resolution afforded by the number of phase history samples given) and suppressed side lobe artifacts. SAR superresolution techniques can be grouped into parametric and nonparametric approaches. Many spectral estimation books exist which describe well the theory and techniques used in SAR imaging [Kay, 1988; Marple, 1987; Stoica and Moses, 1997]. The parametric approaches essentially perform signal detection and model parameter estimation. The nonparametric techniques are generally based on adaptive filtering schemes. It was mentioned that our superresolution methodology is nonparametric in nature. As such, a review of the most common nonparametric SAR superresolution approaches is presented herein. Before this, a quick review on the use of windows and related schemes will be presented. Also, a brief discussion on parametric SAR superresolution is presented.

The windowing of a signal for side lobe suppression is well established and studied [Harris, 1978; Skolnik, 1990]. The subsequent Fourier transforming of the windowed signal can produce a marked suppression of the side lobes which result from a limited number of collected data. As stated before, this is accomplished at the expense of an increase in main lobe width. This results in a decrease in spectral resolvability between closely spaced sinusoids. Stankwitz et al. have developed "smart" windowing schemes which eliminate the side lobes that would result from the FFT of the collected phase history of a trihedral corner reflector [Stankwitz et al., 1995]. These techniques are referred to as spatially variant apodization (SVA); apodization is a term borrowed from
optics which refers to the suppression of diffraction side lobes. SVA is based on deciding which window from the family of cosine-on-pedestal window functions would be most appropriate to use on the point to image. An adjustable parameter $\alpha$ ranging from 0 to 0.5 is used to describe which of the cosine-on-pedestal functions to select at the point being imaged. At each of the extremes of this family of windows is the rectangular window for $\alpha = 0$ (all pedestal, no cosine) and the Hanning window for $\alpha = 0.5$ (all cosine, no pedestal). Hamming weighting is a special case of cosine-on-pedestal ($\alpha = 0.43$) which nulls the first side lobe.

The advantage of this technique is that it is very computationally efficient. It can be implemented easily with a few FFTs. The drawbacks of this method are that it has only been developed to model trihedral reflectors and the best resolution it can afford is that of the main lobe width of the rectangular window – hence its resolution is dependent on the number of collected samples. As such, it is not a superresolution technique per se. A variant of this technique which involves an iterative procedure of “SVA windowing” and phase history extrapolation has been reported which can potentially increase resolution beyond that afforded by the number of observed samples due to an extrapolation phase which increases the number of overall image samples. [Stankwitz and Kosek, 1996]. This approach has been termed super SVA.

Superresolution can be performed parametrically by estimating the parameters of the models assumed to comprise the image [Bi et al., 1997]. In this case, the number of reflectors for each assumed model must be known a priori or must be estimated using some model order selection strategy such as Akaike information criteria [Soderstrom and Stoica, 1989]. This method tends to be slow in practice because the number of corner
reflectors is generally not known and solutions for several model orders must be obtained and compared before an appropriate model order, that is, number of scatterers, can be decided upon. Parameter estimates for each assumed model are typically obtained using least squares approaches. For the case of a complex plane wave, its complex amplitude and 2D frequency must be estimated. This must be done for each plane wave assumed to comprise the signal.

The nonparametric adaptive filtering schemes are based on determining the optimal linear projection of the observed phase history samples for the point being imaged. Essentially, the image is formed on a grid of equispaced frequency coordinates, thus each imaged point corresponds to the 2D frequency of its associated grid location. The frequency samples lie in the rectangular interval \([0, 2\pi) \times [0, 2\pi)\). The optimal filter is the one that minimizes the output power at the point (frequency) being imaged subject to certain constraints. These techniques, generally referred to as minimum variance techniques, will be discussed here. What differentiates each of them are the constraints imposed and/or assumptions made concerning the autocorrelation matrix of the observed samples.

The first of these SAR imaging methods can be attributed to Capon [Capon, 1969]. This approach is typically referred to as Capon’s method or the minimum variance (MV) method. This method imposes what is known as the unit boresight constraint. Here, the filter should pass a complex plane wave of a specified frequency unimpeded while maximally suppressing signal power at all other frequencies. The disadvantage of this method is that the inverse of the autocorrelation matrix must be computed though there is typically only one phase history observation to work with for the scene being
imaged. What is typically done is that subwindows within the observed phase history are averaged in order to obtain a full rank correlation matrix which can be inverted. This approach sacrifices resolution (when considering subwindows) in order to obtain the minimum variance solution.

A reduced rank minimum variance (RRMVM) method has been proposed which does not sacrifice as much resolution as its full rank counterpart [Benitz, 1993]. Here, the subwindows are larger than with the MV method but the autocorrelation matrix is not full rank. The unit boresight constraint is imposed as well as an inequality constraint on the norm of the filter. It turns out that the optimal filter is based on an optimal diagonal loading of the singular autocorrelation matrix – leading to an autocorrelation matrix which can now be inverted to yield the optimal filter.

Methods which assume orthogonality between the signal and noise subspaces are also popular [Odendaal et al., 1994]. These are the MUSIC and EV methods. In both of these methods, a full rank correlation matrix is needed and the unit boresight constraint is enforced. The spectral estimate though is determined by the noise subspace. The MUSIC algorithm explicitly whitens, or equalizes, the noise subspace eigenvalues while the EV method does not. This whitening tends to destroy the spatial inhomogeneities associated with terrain clutter, hence this method is not generally suitable for SAR imaging. In contrast, the EV method does tend to preserve clutter inhomogeneities.

Autoregressive linear prediction (ARLP) can also be used in determining the optimal filter [Jackson and Chien, 1979]. Here, the optimal filter which linearly combines phase history samples to predict a “future” sample is the one that minimizes average prediction error. Based on the assumption that the predicted error signal is an innovations
process (white noise), the spectral estimate is the square root of the minimized prediction error energy divided by the magnitude squared of the transfer function. The filter need not be causal but the white noise assumption of the predicted error signal is generally invalid – leading to spiky and spurious behavior in the final estimate. Instead, it has been found useful to average images based on several prediction steps. The RMS average of ARLP images based on all possible prediction steps reduces to one of Pisarenko’s spectral estimates [Pisarenko, 1972].

Other recent and noteworthy adaptive filtering approaches is APES and HDVI. The APES is a recent technique which generalizes the MV method [Li and Stoica, 1996]. The optimal filter in this method has the exact functional form as Capon’s filter. The difference is that the autocorrelation matrix in the Capon filter is replaced with a matrix composed of the sum of this same autocorrelation matrix along with an additional matrix which results from the optimization carried out. The approach is shown to yield more accurate spectral estimates overall than Capon’s method. High definition vector imaging (HDVI) is another recent technique developed at MIT Lincoln Labs for superresolution imaging [Benitz, 1997]. A typical effect of superresolving with the adaptive filtering techniques based on the trihedral model is that any radar illuminated object which does not satisfy the model is either greatly suppressed or distorted – hence most background information can be modified in an undesired manner depending on what this information will be used for. The HDVI technique imposes constraints that less distorts the background relative to most common adaptive filtering schemes. Constraints require the optimal filter length (norm) to equal the length of the assumed model’s projection onto the space spanned by the autocorrelation matrix. Also, the optimal filter is constrained to be a
given distance from the assumed model’s projection onto the space spanned by the rows or columns of the autocorrelation matrix.
CHAPTER 4
SUPERRESOLUTION OF OPTICAL IMAGES

The superresolution of optical images requires the establishing of an appropriate set of bases by which the collected image samples can be weighted for the reconstruction of new image points (samples). The bases established are learned from available data and the basis function used in the creation of an image point is dependent on the local character of our image. The need for such local consideration will be discussed throughout this chapter. We will examine the interdependence of local information in optical images and make observations which will serve to validate the approach taken to the superresolution of optical images. The architecture which embodies our approach will be shown to be equivalent to a convolution with a family of kernels. The image reconstruction is formulated as

$$\tilde{x}[n_1, n_2] = x_c[n_1, n_2] * * k_{c,l}[n_1, n_2] \quad (4-1)$$

Let us note that the subscripts $c$ and $l$ are used in distinguishing which basis function to use in the formation of an image point. These subscripts index a kernel based on the local image characteristics about the point to reconstruct. The family of kernels is given by

$$\{k_{c,l}[n_1, n_2] | c = 1, 2, \ldots, C; l = 1, 2, \ldots, L \}.$$  

$C$ represents the number of established local image characteristics (or features) from which to compare local neighborhood information and $L$ is the number of kernels created per feature. In summary, eqn. (4-1) is a convolution with a shift-varying kernel whose form is dictated by local image information.
It is a generalization of eqn. (3-7) and defaults to the standard convolution of eqn. (3-7) when $C,L = 1$. Hence, the common approaches to reconstruction which utilize eqn. (3-7) are a special case of the reconstruction implemented herein which is described by eqn. (4-1).

There are two main issues to address regarding the superresolution of optical images:

- how to design members of the kernel family
- how to choose them once they have been established

These issues will be addressed in this chapter. In approaching the superresolution problem, we first model the image acquisition process. The model which results is used to simulate low resolution representations of available images. In this manner, data at different resolutions can be used in developing the basis functions for reconstruction that we seek. We will examine and make observations concerning the existence of interdependent information in optical images. This is followed by a detailed explanation of a superresolution procedure which exploits the observations made.

### 4.1 Optical Image Acquisition Model

Let the function $x(t,t_1,t_2)$ represent a continuous, time-varying image impinging on a sensor plane. The spatial plane is referenced by the $t_1$, $t_2$ coordinate axes and time is referenced by the variable $t$. The imaging sensor plane is assumed to be a grid of contiguous $N_1 \times N_2$ rectangular sensor elements. These elements serve to sample the spatial plane within the camera’s field of view. Each of these elements is said to have a physical size of $p_1 \times p_2$ units of area. The output of each element is proportional to the
amount of light which impinges on each sensor during a given time interval. The output of each sensor, given by \( x_i[n_1, n_2] \) where \( n_1 = 0,1,\ldots,N_1 - 1 \) and \( n_2 = 0,1,\ldots,N_2 - 1 \), can then be expressed as

\[
x_i[n_1, n_2] = \int_0^1 \int_{t_1}^{t_1 + 1} \int_{t_2}^{t_2 + 1} x(t, t_1, t_2) dt_2 dt_1 dt
\]

where the integration over time is one time-unit in duration. The subscript \( l \) is used to denote a “low” resolution image.

To obtain a higher resolution image, a finer sensor grid encompassing the same field of view used in obtaining \( x_i[n_1, n_2] \) would have to be employed. Let the resolution in each spatial dimension be increased – by a factor of \( G_1 \) and \( G_2 \) in their respective spatial dimensions. The physical size of the sensor elements now becomes \( \frac{G_1}{G_i} \times \frac{G_2}{G_i} \) units of area. The high resolution image is then given by \( x_h[m_1, m_2] \), where \( m_1 = 0,1,\ldots,M_1 - 1 \) and \( m_2 = 0,1,\ldots,M_2 - 1 \) and \( M_i = G_i N_i \) (\( i = 1,2 \)). The output for each of the \( M_1 \times M_2 \) sensor elements for the high resolution image can be described by

\[
x_h[m_1, m_2] = \int_0^{G_1} \int_{t_1}^{G_1 t_1} \int_{t_2}^{G_1 t_2} x(t, t_1, t_2) dt_2 dt_1 dt
\]

Notice that the integration limits over time have been extended from one time-unit to \( G_1 G_2 \) time-units in order to maintain the average intensity value for each pixel in the image.

The superresolution process this work addresses is to estimate the high resolution image \( x_h[m_1, m_2] \) from the low resolution image \( x_i[n_1, n_2] \). One can notice that the process of acquiring \( x_i[n_1, n_2] \) from \( x_h[m_1, m_2] \) is given by

\[
x_i[n_1, n_2] = \frac{1}{G_1 G_2} \sum_{m_1 = G_1 n_1}^{G_1 (n_1 + 1) - 1} \sum_{m_2 = G_2 n_2}^{G_2 (n_2 + 1) - 1} x_h[m_1, m_2]
\]

(4-2)
The decimation model in the above equation produces a low resolution image by averaging the pixels of $G_1 \times G_2$ nonoverlapping neighborhoods in the high resolution image.

### 4.2 Using Local Information

Support for the local characterization of images comes from several sources. Image interpolation kernels typically have very local support regions. This was discussed in Chapter 3. In the Shannon sampling theory, the weighting of samples was seen to be inversely related to distance from the point to interpolate. As such, a high resolution image sample is, for all practical purposes, generated from local low resolution samples. Also, the use of finite support region kernels is attractive due to their numerical stability and computational efficiency.

If we assume an image can be described locally by different statistical models, then we must somehow characterize information concerning each of the models which will be useful. In this section we experimentally study the interdependencies of local information in images and discuss their application to the superresolution problem. Note that we are referring to relations that exist among image neighborhoods when we speak of the local interdependencies of images.

#### 4.2.1 Relation Between Correlation and Euclidean Distance

In the sections to follow we will exploit the locally correlated neighborhoods in images for our superresolution paradigm. This section serves to detail the correlation under study and how it relates to Euclidean distance. Here, we introduce a correspondence between statistical correlations of random variables and inner products of vectors and also between statistical mean-square differences and squared distances.
between vectors. A good development of this topic can be found in the text by Helstrom [1991]. This correspondence will serve to identify the analogous relations between samples of our observed data and the true statistical character from which they were drawn. The relations exploit the principle of orthogonality (as defined for both Cartesian space as well as for random variables) as we shall see.

Let \( \mathbf{u}, \mathbf{v} \) denote two random variables and \( \mathbf{u}, \mathbf{v} \) two vectors in Cartesian space. The inner product of \( \mathbf{u}, \mathbf{v} \) is \( \mathbf{u}^T \mathbf{v} \) and the length of a vector \( \mathbf{u} \) is given by its 2-norm \( \| \mathbf{u} \|_2 = (\mathbf{u}^T \mathbf{u})^{\frac{1}{2}} \). The Cartesian space distance \( D \) between vectors \( \mathbf{u}, \mathbf{v} \) based at the origin is defined by

\[
D^2 = (\mathbf{u} - \mathbf{v})^T (\mathbf{u} - \mathbf{v}) = \mathbf{u}^T \mathbf{u} - 2 \mathbf{u}^T \mathbf{v} + \mathbf{v}^T \mathbf{v}
\]

and in the “space” of our random variables the distance \( D \) is defined correspondingly as

\[
D^2 = \mathbb{E}[(\mathbf{u} - \mathbf{v})^2] = \mathbb{E}[\mathbf{u}^2] - 2\mathbb{E}[\mathbf{u} \mathbf{v}] + \mathbb{E}[\mathbf{v}^2]
\]

where \( \mathbb{E}[\cdot] \) denotes the statistical expectation operator. We will make use of the Schwartz inequality [Helstrom, 1991] shortly. For vectors, this inequality is given by

\[
(\mathbf{u}^T \mathbf{v})^2 \leq \mathbf{u}^T \mathbf{u} \mathbf{v}^T \mathbf{v}
\]

and for expected values of random variables it is

\[
(\mathbb{E}[\mathbf{u} \mathbf{v}])^2 \leq \mathbb{E}[\mathbf{u}^2] \mathbb{E}[\mathbf{v}^2].
\]

In an ordinary vector space, the triangle inequality \( \| \mathbf{u} + \mathbf{v} \|_2 \leq \| \mathbf{u} \|_2 + \| \mathbf{v} \|_2 \) is a consequence of the Schwartz inequality. The triangle inequality can be expressed in the following manner by noting that \( \cos \theta \leq 1 \):

\[
\| \mathbf{u} + \mathbf{v} \|_2^2 = \mathbf{u}^T \mathbf{u} + \mathbf{v}^T \mathbf{v} + 2 \| \mathbf{u} \|_2 \| \mathbf{v} \|_2 \cos \theta \leq \mathbf{u}^T \mathbf{u} + \mathbf{v}^T \mathbf{v} + 2 \| \mathbf{u} \|_2 \| \mathbf{v} \|_2 = \| \mathbf{u} \|_2^2 + 2 \| \mathbf{u} \|_2 \| \mathbf{v} \|_2 + \| \mathbf{v} \|_2^2.
\]
In the space of random variables, the corresponding triangle inequality yields the analogous relation

\[ E[(u + v)^2] \leq E[u^2] + 2(E[u^2])^{\frac{1}{3}}(E[v^2])^{\frac{1}{3}} + E[v^2] \]

We notice above the analogous relations between \( E[u^2] \) and \( \|u\|^2 = u^T u \). These relations justify considering the mean square value \( E[u^2] \) as the square of the "length" of a random variable and the correlation \( E[uv] \) as a scalar product. Also note that the definitions of orthogonality apply for these relations too. Two vectors are said to be orthogonal if \( u^T v = 0 \). Similarly, one says that random variables are orthogonal if \( E[uv] = 0 \).

We now discuss the distance measure previously defined for vectors and random variables. We will show an important relationship between inner products and distance in Cartesian space as well as correlations and distance in the analogous "probability space."

Let us now consider the relationship between minimization of Euclidean distance with the maximization of inner products in a vector space. Consider a set of vectors \( x, f_1, \ldots, f_N \) where \( f_i \) is closest in Euclidean distance to \( x \), that is,

\[ \| f_i - x \|^2 \leq \| f_j - x \|^2, \quad i \neq j. \quad (4-3) \]

We can rewrite this as \( f_i^T f_i - 2f_i^T x \leq f_j^T f_j - 2f_j^T x \). Notice that if all the \( f_k \) vectors are of equal length, we can equivalently express eqn. (4-3) as

\[ f_i^T x \geq f_j^T x, \quad i \neq j. \quad (4-4) \]

If the \( f_k \) vectors are not of equal length, then eqn. (4-3) is equivalent to

\[ f_i^T x - \frac{1}{2} f_i^T f_i > f_j^T x - \frac{1}{2} f_j^T f_j. \]

In either case, minimizing the Euclidean distance between the \( x \) and \( f_k \) vectors is equivalent to maximizing the inner product of those vectors if the length of the \( f_k \) vectors
are all equal. If the $f_k$ vectors are not of equal length, then the square of the vectors’ length must be considered.

A similar result arises when minimizing the mean squared distance of random variables. Consider the random variables $x_1, f_1, \ldots, f_N$ where $f_i$ results in the minimum mean squared difference between the random variables $f_1, \ldots, f_N$ and $x$, that is,

$$E[(f_i - x)^2] \leq E[(f_j - x)^2], \quad i \neq j .$$

(4-5)

We can rewrite this as $E[f_i^2] - 2E[f_i x] \leq E[f_j^2] - 2E[f_j x]$. Notice that if all the random variables $f_k$ have equivalent second order moments or squared “lengths”, then we can equivalently express eqn. (4-5) as

$$E[f_i x] \geq E[f_j x]$$

If the random variables $f_k$ do not have equivalent second order moments, then eqn. (4-5) is equivalent to

$$E[f_i x] - \frac{1}{2} E[f_i^2] > E[f_j x] - \frac{1}{2} E[f_j^2].$$

(4-6)

We see that minimizing the mean squared difference between $x$ and the random variables $f_1, \ldots, f_N$ is equivalent to maximizing their correlation if the second order moments of $f_1, \ldots, f_N$ are equivalent. If the second order moments of the random variables $f_1, \ldots, f_N$ are not all equivalent, then their squared “length” must also be considered.

Let us now note that the minimization of Euclidean distance between vectors is the template matching problem [Duda and Hart, 1973]. Template matching is a scheme used for classification purposes. Let vector $x$ be a portion of an image with support size equal to that of the template. The $f_k$ vectors are the set of $N$ templates from which we attempt to classify image regions. The best match is found to be with the template maximally
correlated with the image region of interest. This correlation measure is nonstatistical and is described in Chapter 7 of the text by Duda and Hart [1973].

A statistical interpretation of the template matching problem can also be developed. Assume a random vector \( \mathbf{x} = [x_1, ..., x_M]^T \) can be described by a mixture model as follows

\[
p_x(\mathbf{x}) = \sum_{k=1}^{N} P(k) p_x(\mathbf{x}|k)
\]  

(4-7)

where \( \sum_{k=1}^{N} P(k) = 1 \) and \( 0 \leq P(k) \leq 1 \) for \( k = 1, ..., N \). Also, let the mean of each conditional density function \( p_x(\mathbf{x}|k) \) be \( \mathbf{f}_k = E[\mathbf{x}_k] \) where \( \mathbf{x}_k = [x_{k_1}, ..., x_{k_m}] \) is the random vector with conditional density \( p_x(\mathbf{x}|k) \). The template matching problem can be posed as finding the mean \( \mathbf{f}_k \) closest in Euclidean distance with an observation \( \mathbf{x} \) from the mixture density function. Specifically, let us define the template matching problem as

\[
J = \arg \min_j \{ tr\{ E[(\mathbf{x}_k - \mathbf{f}_j)(\mathbf{x}_k - \mathbf{f}_j)^T]\} \}
\]

(4-8)

where \( tr\{\cdot\} \) is the trace of a matrix and the best matched template is given by \( \mathbf{f}_j \).

Expanding eqn. (4-8) and evaluating terms, we can equivalently determine the best matched template from

\[
J = \arg \min_j \left( \mathbf{f}_j^T \mathbf{f}_j - 2 \mathbf{f}_j ^T \mathbf{f}_k \right)
\]

(4-9)

from which it is easy to show that \( J = k \) is the solution. Specifically, if we let \( j = k \) in eqn. (4-9), then \( \mathbf{f}_k^T \mathbf{f}_k - 2 \mathbf{f}_k ^T \mathbf{f}_k < \mathbf{f}_j^T \mathbf{f}_j - 2 \mathbf{f}_j ^T \mathbf{f}_k \). Now adding \( \mathbf{f}_k^T \mathbf{f}_k \) to both sides of this inequality yields \( 0 \leq (\mathbf{f}_j - \mathbf{f}_k)^T (\mathbf{f}_j - \mathbf{f}_k) \) where \( j = k \) achieves the equality. So the best matched template corresponds to the mean of the conditional density function which generated the observation \( \mathbf{x} \) from \( p_x(\mathbf{x}) \).
If the templates are now interpreted statistically, then the template matching problem of eqn. (4-8) can be expressed as

$$ J = \arg \min_j \left\{ \text{tr} \left\{ \mathbb{E} \left[ (x - f_j)(x - f_j)^T \right] \right\} \right\} $$

given a random vector $x$. This can be shown to be equivalent to

$$ J = \arg \max_j \left( \mathbb{E}[f_j^T x] - \frac{1}{2} \mathbb{E}[f_j^T f_j] \right) $$

(4-10)

which is seen to be analogous to that of eqn. (4-6). The solution to the template matching problem here then involves the template most correlated with the input.

The two issues to stress here are the analogous relationships between random variables and vectors involving distance metrics in each of their respective spaces. The minimization of a squared distance measure is directly related to the maximization of a correlation measure, that is, the distance measure is inversely related to the correlation measure. In this way, observations from highly correlated random variables or random vectors are close in Euclidean space. In particular, if the random variables and/or random vectors are statistically independent and their covariances are assumed equal, then maximal correlation implies minimal Euclidean distance in the mean sense of those random variables. This can be seen by comparing eqn. (4-10) with eqn. (4-4) under these assumptions.

### 4.2.2 Existence of Locally Correlated Information in Images

We can provide experimental evidence for the existence of different types of local interdependencies in images which can be exploited for the superresolution of images. The optical images for which results will be reported in this chapter are given in fig. (4-1). Interdependent information in real-world images are known to appear locally and have
been exploited in compression schemes [Dony and Haykin, 1995a; Kambhatla and Leen, 1997]. These methods attempt to exploit the correlation amongst nonoverlapping image blocks. Specifically, the blocks of the image are grouped into disjoint sets and compression is performed on each set individually.

The next two sections will serve to illustrate interblock and scale interdependencies that exist within the images of fig (4-1). These interdependencies form the basis from which our superresolution methodology arises. It is the general existence of these interdependencies in real-world optical imagery which substantiates the use of our superresolution methodology. The methodology is described in detail later in this chapter.

4.2.2.1 Interblock correlation

The interblock dependencies of images will be illustrated here. The benefits of the existence of such dependencies is that the samples from highly correlated blocks tend to be very similar – as discussed in section 4.2.1. The similarity between these blocks, which can be located throughout an image, leads us to believe that the local statistics governing their “generation” are also similar. The similar blocks are assumed here to have been generated by the same statistical process. The marriage of interblock correlation and their interdependencies across scales makes the superresolution problem approachable from such local information. In order to substantiate the notion that an image can be described by different local statistical models, a simple experiment was performed.
Figure 4-1. Images used in this chapter. (a) 256×256 images - (top left) Lena, (top right) Peppers, (bottom left) left Pentagon image of a stereo pair, (bottom right) right Pentagon image of a stereo pair; (b) 128×128 images obtained by the image acquisition model - (left to right) Lena, Peppers, left Pentagon, right Pentagon.
All of the overlapping neighborhoods of size $H_1 \times H_2$ for the Lena 128$x$128 image of fig. (4-1b) were extracted. Each neighborhood then had its sample mean subtracted from it—yielding what we refer to as a “structural” neighborhood. It is the local structure of images that we seek to statistically describe. By considering structural neighborhoods, an image $x[n_1, n_2]$ is then composed of two sets of ordered neighborhoods: one set contains the neighborhoods’ mean $m[n_1, n_2]$ and the other set the neighborhoods’ structure $s[n_1, n_2]$ such that $x[n_1, n_2] = m[n_1, n_2] + s[n_1, n_2]$. The extracted structural neighborhoods which comprise $s[n_1, n_2]$ were then grouped into $K$ disjoint clusters following a vector quantization (VQ) of these neighborhoods. The clustering was done à la eqn. (4-3) where $x$ now denotes a structural neighborhood and $f_1, ..., f_K$ are the feature (codebook) vectors or Voronoi centers resulting from the VQ. The feature vectors can each be interpreted as corresponding to the mean of various statistical models which describe the structural neighborhoods. This was discussed in eqn. (4-7). After the structural neighborhoods were clustered, the neighborhoods within each cluster were randomly shuffled thus yielding a new set of ordered neighborhoods’ structure $\hat{s}[n_1, n_2]$. If the clustered neighborhoods are truly similar in structure then their random shuffling should still produce a pleasing image. Finally, the composition of a new image results from $\hat{x}[n_1, n_2] = m[n_1, n_2] + \hat{s}[n_1, n_2]$. This type of operation could consider overlapping and nonoverlapping neighborhoods. When overlapping neighborhoods are considered, the overlapping samples from each neighborhood are averaged to yield the new image samples. Note that the “reconstruction” performance obtained using this approach is analogous to a test set performance for validating the existence of interblock correlations.
Fig. (4-2) illustrates the results of performing such a test with overlapping $3 \times 3$ ($H_1=H_2=3$) neighborhoods on the Lena ($N_1=128 \times (N_2=128)$) image.

Figure 4-2. Illustration of the effect of characterizing the Lena image based on the interblock correlation of $3 \times 3$ neighborhoods. The image used was $128 \times 128$. Increasing the number of clusters leads to increased correlation within the neighborhoods of a cluster — thus yielding a more accurate local representation of the image. (a) $K=1$ cluster, (b) $K=15$ clusters, (c) $K=30$ clusters.

With $K=1$ in fig. (4-2a), no clustering is being performed since all neighborhoods from the image are assumed to belong to the same cluster. The result of creating an image in this manner yields a poor (and noisy looking) image when compared to the original. In fig. (4-2b), $K=15$ clusters were formed from the structural neighborhoods of the image. The resulting image produces a better image representation. Fig. (4-2c) illustrates the result with $K=30$ clusters. This resulted in the most faithfully recomposed image upon comparison of their peak signal-to-noise ratios (PSNRs). The PSNR is defined as

$$
\text{PSNR} = -10 \log_{10} \left( e_{rms}^2 \right)
$$

where

$$
e_{rms}^2 = \frac{1}{N_1 N_2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \left( x[n_1, n_2] - \hat{x}[n_1, n_2] \right)^2
$$

and $x$ and $\hat{x}$ take values in $[0,1]$ and are $N_1 \times N_2$ in size. Fig. (4-3) illustrates the recomposition performance for the Lena $128 \times 128$ image using regions of support (ROS).
of size 3×3 and 5×5 for K=1, ..., 30 clusters. We can see that generally the PSNR of the new images increases with increasing number of clusters. The smaller the neighborhood, the more accurate we can consider our image model when based on interblock correlations. Though we only report these results on one image, it is generally appropriate to characterize optical images in this manner.

4.2.2.2 Scale interdependencies

We have just seen how an image can be segmented locally into pieces that are most correlated with each other. Naturally, we can inquire about the existence of similar information across image scales. If a strong similarity exists between homologous and highly correlated regions of the low and high resolution images, then it seems feasible that a simple transformation of the low resolution image neighborhoods within each cluster can yield a corresponding high resolution neighborhood very similar to the low resolution neighborhood’s homologous counterpart.

This transformation is the link in associating low and high resolution neighborhoods. In this section, we wish to analyze experimentally whether such highly correlated information across image scales exists. That is, we wish to examine the correspondence of highly correlated neighborhood information between a low resolution image and its high resolution counterpart. We will show examples using the images shown in fig. (4-1). The experiment will report on the percentage of homologous neighborhoods from the low and high resolution counterparts of these images that were similarly clustered. If a high percentage of these neighborhoods are found, then a strong scale interdependence among neighborhoods is said to exist. We now explain the details for establishing this percentage.
Figure 4-3. Quantitative illustration of the effect of interblock correlation on the recomposition of the Lena 128×128 image. The structural neighborhoods in each of the clusters from this image were randomly shuffled and the image was “pieced” back together. A very high interblock correlation results in an image very similar to the original.
Given a low and high resolution version of an image, \( x_l[n_1, n_2] \) and \( x_h[n_1, n_2] \) respectively, the homologous structural neighborhoods in each of these images were clustered to form \( K \) independent groups. The \( H_1 \times H_2 \) neighborhoods in the \( N_1 \times N_2 \) image \( x_l[n_1, n_2] \) forms the set of neighborhoods

\[
X = \left\{ x_l[m_1 : m_1 + H_1 - 1, m_2 : m_2 + H_2 - 1] \right\}_{m_1=0, \ldots, N_1-H_1, m_2=0, \ldots, N_2-H_2}.
\]

The homologous neighborhoods in the high resolution image are defined as the \( G_1 H_1 \times G_2 H_2 \) neighborhoods in the \((M_1 = G_1 N_1) \times (M_2 = G_2 N_2)\) image \( x_h[n_1, n_2] \) which forms the set

\[
D = \left\{ x_h[G_1 m_1 : G_1 m_1 + G_1 H_1 - 1, G_2 m_2 : G_2 m_2 + G_2 H_2 - 1] \right\}_{m_1=0, \ldots, N_1-H_1, m_2=0, \ldots, N_2-H_2}
\]

where recall that \( x_l[n_1, n_2] \) was obtained from \( x_h[n_1, n_2] \) through decimation by a factor of \( G_1 \times G_2 \) according to our image acquisition model.

Now the neighborhoods in \( X \) are clustered to form \( K \) disjoint groups \( X_1, \ldots, X_K \) and the neighborhoods in \( D \) are separately clustered to form \( K \) disjoint groups \( D_1, \ldots, D_K \). If the homologous neighborhoods in \( X \) and \( D \) form similar clusters in their respective images, then the information content of the low and high resolution images must be similar in some sense. As alluded to earlier, the image is said to exhibit scale interdependence. To determine how well clustered information from the same image relates across scales, we form a confusion matrix as shown in fig. (4-4).

The entry in location \((j,k)\) of the matrix is the number of homologous neighborhoods common to cluster \( X_j \) and \( D_k \), \( j,k = 1, \ldots, K \). The interdependence across scales is determined as the maximum number of homologous neighborhoods common to the clusters formed. Since the ordering of the true clusters or “classes”
between \( X_j \) and \( D_k \) is not known, we can’t just examine the contents of the diagonal of the confusion matrix.

![Confusion Matrix Diagram](image)

Figure 4-4. Confusion matrix for clustered homologous neighborhoods within their low and high resolution images. The \( X_j \) are the disjoint sets of clustered neighborhoods in the low resolution image and the \( D_k \) are the disjoint sets of clustered neighborhoods in the high resolution image; \( j,k = 1,...,K \).

Instead, we must search for the most likely ordering. This in turn yields the number which reveals a measure of how similarly homologous information in the low and high resolution images was clustered. This number is easily found with the following simple algorithm:

Step 1: Initialize \( N=0 \);

Step 2: Find the largest number \( L \) in the confusion matrix and save its row and column coordinates \((r,c)\)

Step 3: Perform \( N\leftarrow N+L \);

Step 4: Remove row \( r \) and col \( c \) from the confusion matrix to form a new confusion matrix with one less row and column.
Step 5: If confusion matrix has no more rows and cols: STOP else Go to step 2

The variable N represents the number of homologous neighborhoods common to similar clusters from the low and high resolution images. The percentage of such clustered neighborhoods is \( P = \frac{N}{(N_1 - H_1 + 1)(N_2 - H_2 + 1)} \) since there are a total of \((N_1 - H_1 + 1)(N_2 - H_2 + 1)\) homologous neighborhoods to cluster in each image.

In the plots below, this percentage is plotted as a function of the number of clusters. This is done for three images: Lena, Peppers and the Left Pentagon. The high resolution images clustered are 256×256 and their low resolution counterparts which were obtained through \((G_1 = 2) \times (G_2 = 2)\) decimation were 128×128. The plots also report on two different neighborhood sizes tested: \(H_1 \times H_2 = 3 \times 3\) and \(H_1 \times H_2 = 5 \times 5\). Figs. (4-5)-(4-7) illustrate that there is a very strong interdependency of homologous neighborhoods across image scales for these three images – even as the number of clusters increases towards \(K=30\). The case of \(K=1\) always yields an interdependency of 1. This is because for \(K=1\), no clustering is actually being performed, that is, all neighborhoods are assumed to belong to the same cluster. As such, the “disjoint” sets (or single set in this case) have all the homologous neighborhoods in common.

The interdependency in general decreases as the number of clusters increases. This is intuitively expected as an increase in the number of clusters results in the clustering of information increasingly specific to a particular image and scale. Because the frequency content between the low and high resolution counterpart images differ, the greater specialization of information within an image is expected to result in less interdependency among them.
Scale interdependency plot for the Lena image. The interdependency of homologous neighborhoods across image scales is reported. The high resolution image was 256×256 and the corresponding low resolution image was 128×128. Two ROSs are reported for computing the interdependency: 3×3 and 5×5. A very high interdependency among homologous neighborhoods is seen for this image even when considering $K=30$ clusters.
Figure 4-6. Scale interdependency plot for the Peppers image. The interdependency of homologous neighborhoods across image scales is reported. The high resolution image was 256×256 and the corresponding low resolution image was 128×128. Two ROSs are reported for computing the interdependency: 3×3 and 5×5. A very high interdependency among homologous neighborhoods is seen for this image even when considering K=30 clusters.
Figure 4-7. Scale interdependency plot for the Left Pentagon image. The interdependency of homologous neighborhoods across image scales is reported. The high resolution image was 256×256 and the corresponding low resolution image was 128×128. Two ROSs are reported for computing the interdependency: 3×3 and 5×5. The interdependencies among homologous neighborhoods are not as strong in this image as compared with the Lena and Peppers images but still remain above 50% even when considering $K=30$ clusters.
Fig. (4-5) illustrates the scale interdependencies for the Lena image. These remain very high even when considering \( K=30 \) clusters. In this case, the interdependencies achieved were all above 75%. Fig. (4-6) reports on the Peppers image. This image yields interdependency percentages very much like those of the Lena image. Fig. (4-7) illustrates the interdependencies of the Left Pentagon image. Here, the interdependencies are seen to drop more rapidly as the number of clusters formed increased than with the previous two reported images. Further, there is roughly half as much interdependency with the Left Pentagon image as with the Lena and Peppers images — going down to about 55% for the case of \( K=30 \) as compared to 75% with the Lena and Peppers images.

4.3 Architecture Specifics

The modular design of the superresolution architecture to be presented is supported by the observations of the local interdependencies in images that have been discussed. Portions of images exhibiting different statistical characteristics are treated differently in order to fully utilize the local image information and provide for a reconstruction not afforded by the sampling theorem. The procedure for learning the a priori information needed for the superresolution of optical images is shown in fig. (4-8).

This procedure, which will be discussed in detail in the next section, establishes the basis functions for reconstruction from the available image data. The information extracted is used in the superresolution of images that are similar in class to those from which our basis functions were derived.
Figure 4-8. Training process for the superresolution of optical images.
Fig. (4-9) illustrates the architecture used for superresolving images once the training is complete, that is, once the necessary information has been established. It is equivalent to a convolution with a family of kernels as presented in eqn. (4-1). Here, we wish to initially present this paradigm in a fashion which facilitates the explanation of our family of kernels later on. As we proceed, we will explain any information not cast in a manner directly conforming to eqn. (4-1) and relate it to that equation's form.

There are two fundamental steps specific to the superresolution of optical images:

- clustering of local low resolution image data
- construction of local high resolution data.

The first step can be viewed as the process of determining the neighborhoods most correlated with the mean (or “template” as discussed in section 4.2.1) of the underlying distribution from which those neighborhoods were drawn. It encompasses the operations of extracting all of the image neighborhoods of a predetermined size from the low resolution image and then clustering them. This clustering addresses the main issue of choosing a kernel. The predetermined neighborhood size describes the region of support of our family of kernels, the location of each neighborhood relates to the location of the convolution output and the cluster associated with each neighborhood will be used to select the kernel with which to compute superresolved neighborhoods.

The second step exploits the interdependent information across scales which has been seen to exist in images. It involves the depicted linear associative memories (LAMs) and proper arranging of superresolved neighborhoods. Here we have addressed the main issue of designing the members of the kernel family.
Figure 4-9. The superresolution architecture for optical images. This paradigm performs the equivalent operation of a convolution with a family of kernels.
Each LAM is encoded with the mapping that locally creates high resolution neighborhoods from low resolution neighborhoods via $L$ transformation kernels. There are $C$ LAMs and each is tuned to specific local image characteristics. Note that each LAM superresolves a neighborhood rather than just a single pixel.

4.4 Training Phase

The basis functions from which to project our image samples are specified by the LAMs in fig. (4-9). They are determined via a training process from available data. The training procedure for the superresolution paradigm was pictured in fig. (4-8). It will now be discussed in detail. This procedure can be broken into two main sections: a data preprocessing section and a high resolution construction section. The preprocessing takes all of the local neighborhoods of a fixed size in the low resolution image and clusters them. Each low resolution neighborhood has a corresponding homologous high resolution neighborhood it is associated with. Each cluster consists of a training and a corresponding desired data set which is used to train a LAM. The high resolution construction is responsible for associating the training and desired sets in each cluster independent of the others. This information is now detailed.

4.4.1 The Preprocessing

Two preprocessing steps are performed prior to the construction of high resolution images from low resolution information. The preprocessing consists of simulating low resolution images followed by neighborhood extraction for pairing homologous neighborhoods.
4.4.1.1 Decimation

Ideally, the low and high resolution data sets used to train the aforementioned LAMs would each encompass the same scene and have been physically obtained by hardware with different, but known, resolution settings. Such data collection is not very common. Instead, the low resolution counterparts of the given signals are simulated. Decimation is only needed to simulate the low resolution counterparts to our high resolution images. If we had, a priori, the necessary two sets of images, this block would be removed from the training process depicted in fig. (4-8). Here, the simulated low resolution images will be obtained following the image acquisition model discussed earlier. In other words, decimation will be referred to as the process of averaging nonoverlapping image neighborhoods in order to obtain a lower resolution image, à la eqn. (4-2).

4.4.1.2 Neighborhood extraction

Each $H_1 \times H_2$ neighborhood of our low resolution image $x_i[n_1, n_2]$ will be used in training our superresolution system. A low resolution image of size $N_1 \times N_2$ will then have $(N_1 - H_1 + 1)(N_2 - H_2 + 1)$ neighborhoods. Note that $N_i = M_i / G_i$ and $M_i$, $(i = 1, 2)$, describes the size of our original image. Each low resolution neighborhood will be associated with its $(2G_1 - 1) \times (2G_2 - 1)$ homologous high resolution neighborhood. We have chosen to associate the low and high resolution neighborhood information in a symmetrical fashion. The homologous neighborhoods we are associating are defined as follows. The set of $(N_1 - H_1 + 1)(N_2 - H_2 + 1)$ neighborhoods in the low resolution image of size $H_1 \times H_2$ are given by

$$X = \left\{ x_i[m_1 : m_1 + H_1 - 1, m_2 : m_2 + H_2 - 1] \right\} \bigg|_{m_1 = 0, \ldots, N_1 - H_1, m_2 = 0, \ldots, N_2 - H_2}$$

(4-12)

The homologous neighborhoods to associate with these are described by
\[ D = \left\{ x_n [G_1 m_1 + \phi_1 + 1 : G_1 (m_1 + 2) + \phi_1 - 1, \right. \\
\left. G_2 m_2 + \phi_2 + 1 : G_2 (m_2 + 2) + \phi_2 - 1] \right\} \]
\[ m_1 = 0, ..., H_1 - 1, m_2 = 0, ..., H_2 - 1 \]

where \( \phi_i = \frac{G_i (H_i - 3)}{2} \) and \( i = 1, 2 \). \( H_1 \) and \( H_2 \) are restricted to being odd numbers so that we can construct about each low resolution neighborhood’s center. We also only construct high resolution information that is contained within the boundary of a low resolution neighborhood. The association of homologous neighborhood information described by eqns. (4-12) and (4-13) is depicted in fig. (4-10).

Figure 4-10. Local image neighborhoods and the pixels they reconstruct. Each circle represents a 2D high resolution image pixel. The shaded circles are the low resolution image pixels obtained via decimation of the high resolution image. The gray pixel is the center of the low resolution neighborhood. Each \( H_1 \times H_2 \) low resolution neighborhood constructs a \((2G_1 - 1) \times (2G_2 - 1)\) high resolution neighborhood about the low resolution neighborhood’s center – these are depicted by the crossed circles. The numbers are a convention used to distinguish between constructed pixels in this neighborhood. (a) Decimation factor \( G_1 = G_2 = 2 \). (b) Decimation factor \( G_1 = G_2 = 3 \).

Consider the center pixel of an \( H_1 \times H_2 \) low resolution neighborhood. For the case of \( G_1 = G_2 = 2 \) in fig. (4-10a), we are constructing the crossed circles about our center (gray) pixel. By not constructing the center pixel, we are strictly interpolating about our observed image samples. If we allow for construction of the center pixel, we
can change a 'noisy' observed sample. Fig. (4-10b) similarly illustrates this for the case of
$G_1 = G_2 = 3$.

The selection of the low resolution neighborhood size is a trade-off between the
amount of local support to consider and the amount of information needed to construct.
As a rule of thumb, one should set $H_i \geq 2G_i - 1$ but should keep $H_i$, ($i = 1, 2$), reasonably
small to preserve sufficient low resolution support associated with the high resolution
neighborhood. As $G_i$ increases, there is more missing information to construct, hence
more low resolution sample support is needed. $H_i$ should not be set too high as we want
to keep the support as local as possible. We will see the effects of this later when the
superresolution results are presented. Also, the smaller the $H_i$, the smaller the number of
free parameters needed in the LAMs.

An important note to consider here is that overlapping neighborhoods construct
some of the same high resolution samples. We must remember that our high resolution
sample construction is a procedure whereby we estimate the samples of a high resolution
image from a low resolution representation. In essence, we would like to obtain more
than one estimate of each high resolution sample point. By averaging several high
resolution sample estimates we believe the final high resolution sample estimated is more
reliable.

4.4.2 Hard Partitioning the Input Space

The input space consists of the neighborhoods in our low resolution image to be
superresolved. Because we assume no prior information about our image, we cluster (or
hard partition) our input space based solely on these neighborhoods. The segmentation is
based on the interblock correlation among neighborhoods of our image. In order to group
(cluster) the image neighborhoods that are most correlated with one another, we interpret each $H_1 \times H_2$ sized neighborhood as a point in $H_1 H_2$ dimensional Euclidean space, that is, as a vector from the origin of our space to the coordinates given by the neighborhood values. As previously discussed, the most correlated neighborhoods will be closest to each other in our $H_1 H_2$ dimensional vector space.

The hard segmentation we seek is equivalent to a vector quantization (VQ) of our input space based on minimizing an $L_2$ norm criteria. Techniques which have been developed to perform such VQ include the LGB [Linde, et al., 1980] and Neural Gas [Martinez, et al., 1993] algorithms as well as the topologically ordered Kohonen self-organizing feature map (SOFM) approach [Kohonen, 1990]. These approaches are considered to be unsupervised learning techniques which produce a set of feature vectors also referred to in the literature as quantization nodes or codebook vectors. The results reported in this work utilize Kohonen’s approach. This is now described.

4.4.2.1 Kohonen’s self-organizing feature map

The principal goal of Kohonen’s self-organizing feature map (SOFM) is to transform a set of input exemplars to a discrete set of representative vectors. The algorithm is a topologically ordered vector quantization (VQ) on the set $\{x_r\}_{r=1}^N$ of exemplars. The representative vectors are typically termed codebook vectors, feature vectors and/or quantization nodes. This ordering serves two purposes: during training it encourages a full utilization of the codebook vectors while they’re evolving and once the vectors have been established, it allows for “close” input exemplars to also be “close” in the representative set.
The feature vectors in the SOFM can be obtained using a simple iterative scheme. After randomly initializing a set of feature vectors \( f_c, c = 1, ..., C \), the input exemplars are compared with the feature vectors one at a time. The best matching feature vector, in the Euclidean sense, is sought. This is given by eqn. (4-14) as

\[
\gamma(x_r) = \arg \min_c \left( \| x_r - f_c \|^2 \right); \quad c = 1, 2, ..., C
\]  

(4-14)

This “winning” vector and its neighbors are then updated by

\[
f_c(n + 1) = \begin{cases} 
    f_c(n) + \mu(n)[x_r - f_c(n)]; & c \in \Lambda_{\gamma(x_r)}(n) \\
    f_c(n) & \text{otherwise}
\end{cases}
\]

where \( \mu \) is the learning rate parameter, \( \Lambda_{\gamma(x_r)} \) is the neighborhood function centered around the winning feature vector \( \gamma(x_r) \) and \( n \) describes the discrete time step (iteration).

The learning rate and neighborhood function are gradually decreased during training for best results.

**4.4.2.2 Cluster formation**

The feature vectors completely define the hard partitioning of our input space. All image neighborhoods closest in Euclidean distance to one of the feature vectors are said to belong to a cluster. Mathematically, given a set of \( N \) vectors \( \{x_r\}_{r=1}^N \) from which \( C \) feature vectors \( \{f_c\}_{c=1}^C \) have been extracted, the \( C \) clusters \( K_c, c = 1, 2, ..., C \) formed by the vectors are given by

\[
K_c = \{ x_r : \gamma(x_r) = c; r = 1, 2, ..., N \}
\]

and the hard partitioning function spoken of earlier is \( \gamma(x_r) \) of eqn. (4-14).

Generally, the feature vectors established from the available data reflect the statistics of the distribution of the data vectors. Regions in the input space of the input
data vectors with high probability will be represented by a subset of the feature vectors. This leads to a natural yet compressed representation of our data with which to compare test data sets against. In this work, our feature vectors are obtained by performing a VQ on the local structure of neighborhoods \( x_r \) of our low resolution training images using the SOFM approach.

### 4.4.3 Associative Memories

The associative memories considered here are essentially parametric mappings that optimally associate between a set of input exemplars \( X = \{ x_r \}_{r=1}^{N} \) and a set of desired exemplars \( D = \{ d_r \}_{r=1}^{N} \) [Haykin, 1994]. These sets of exemplars can be represented as the columns of a matrix, \( X \) and \( D \) respectively. The optimal association is typically defined in the least squares or mean squared sense. The process of arriving at the optimal parameters is accomplished iteratively when no closed form solution is available. This iterative process for optimal association between data sets (matrices) \( X \) and \( D \) is known as supervised learning in neural networks. The two supervised learning topologies discussed in this section are the linear network and the multilayer perceptron. Results using these topologies will be presented later.

#### 4.4.3.1 Least mean squares

Consider the minimization of the following cost function

\[
J(W) = \| D - WX \|^2_F
\]  

where \( D \in \mathbb{R}^{P \times N} \), \( W \in \mathbb{R}^{P \times Q} \), \( X \in \mathbb{R}^{O \times N} \) and \( \| \cdot \|^2_F \) is the Frobenius matrix norm [Golub and Van Loan, 1989]. We are trying to find the matrix \( W \) such that matrices \( X \) and \( D \) are associated optimally in the least squares sense. Let us denote the \( r^{th} \) column of matrices \( D \)
and $X$ by $d_r$ and $x_r$, respectively, where $r = 1, \ldots, N$. Also, let $Y = WX$. The $r$th column of $Y$ is then $y_r = WX_r$. We will now solve for $W$. Rewriting eqn. (4-15), we have

$$J = \sum_{r=1}^{N} \|d_r - y_r\|_2^2$$

$$= \sum_{r=1}^{N} (d_r - y_r)^T (d_r - y_r)$$

$$= \sum_{r=1}^{N} (d_r^T d_r - 2d_r^T y_r - y_r^T y_r)$$

$$= \sum_{r=1}^{N} (d_r^T d_r - 2d_r^T W x_r - x_r^T W^T W x_r)$$

(4-16)

where the dependence of $J$ on $W$ is understood. To solve for $W$, we must take the derivative of $J$ with respect to $W$ and set the resulting equation to 0, that is,

$$\frac{\partial J}{\partial W} = \begin{bmatrix} \frac{\partial J}{\partial w_{11}} & \cdots & \frac{\partial J}{\partial w_{1Q}} \\ \vdots & \ddots & \vdots \\ \frac{\partial J}{\partial w_{P1}} & \cdots & \frac{\partial J}{\partial w_{PW}} \end{bmatrix} = 0$$

(4-17)

Notice that $w_{pq}$, where $p = 1, \ldots, P$ and $q = 1, \ldots, Q$, is an element of matrix $W$, that is, $W = [w_{pq}]$. Then, from eqns. (4-16) and (4-17)

$$\frac{\partial J}{\partial W} = \frac{\partial}{\partial W} \left( \sum_{r=1}^{N} (d_r^T d_r - 2d_r^T W x_r + x_r^T W^T W x_r) \right)$$

$$= \sum_{r=1}^{N} \left( -2 \frac{\partial}{\partial W} (d_r^T W x_r) + \frac{\partial}{\partial W} (x_r^T W^T W x_r) \right)$$

$$= \sum_{r=1}^{N} \left( -2d_r^T x_r^T + 2W x_r x_r^T \right)$$

(4-18)

where it can be shown that $\frac{\partial}{\partial W} (a^T W b) = a^T b$ and $\frac{\partial}{\partial W} (a^T W^T W a) = 2Waa^T$. Setting eqn. (4-18) to 0 and multiplying both sides by $\frac{1}{N}$ we obtain

$$W \left( \frac{1}{N} \sum_{r=1}^{N} x_r x_r^T \right) = \left( \frac{1}{N} \sum_{r=1}^{N} d_r d_r^T \right).$$

(4-19)
The summations of the outer products in eqn. (4-19) are equivalent to matrix multiplications. Eqn. (4-19) is rewritten as

$$W\left(\frac{1}{N} XX^T\right) = \left(\frac{1}{N} DX^T\right).$$

(4-20)

Solving for $W$ we obtain our solution as

$$W = DX^T (XX^T)^{-1}.$$  

(4-21)

A more commonly used approach to obtain the closed form solution of eqn. (4-21) makes use of the singular value decomposition (SVD) of $X^T$ [Golub and Van Loan, 1989]. This approach is more stable since it does not require finding the numerical inverse of a matrix. In this case,

$$W = U \Sigma^+ V^T$$

(4-22)

where $X^T = U \Sigma V^T$ is the SVD of $X^T$, $\Sigma^+ = \text{diag}(\frac{1}{\sigma_1}, \cdots, \frac{1}{\sigma_s}, 0, \cdots, 0)$ and the nonzero singular values of $X^T$ are denoted by $\sigma_i$, $i = 1, \ldots, s$.

The solution to the least squares problem can be arrived at iteratively via the least mean squares (LMS) algorithm [Haykin, 1994]. The weights are iteratively updated as follows

$$W(n + 1) = W(n) + \mu (D - Y)X^T$$

(4-23)

where $\mu$ is the learning rate which is a small positive constant and $n$ is the discrete time step. It is easy to see that this iterative scheme yields the solution of eqn. (4-21). When eqn. (4-23) reaches its global (and only) minimum, then at "steady state" $W(n + 1) = W(n)$ which implies that $(D - Y)X^T = 0$. Now by substituting $Y = WX$ as defined earlier, we can solve for $W$ which yields eqn. (4-21).
Let us make a few observations concerning eqns. (4-20) and (4-21). Notice that \( \frac{1}{N} \mathbf{X}\mathbf{X}^T \) is the sample autocorrelation matrix of the \( \mathbf{x}_r \) vectors. Similarly, \( \frac{1}{N} \mathbf{D}\mathbf{X}^T \) is the sample crosscorrelation matrix between the \( \mathbf{x}_r \) and \( \mathbf{d}_r \) vectors. As the number of sample vectors \( N \) goes to infinity, the sample correlation matrices become the statistical correlation matrices and eqn. (4-21) is the solution to the Wiener-Hopf equations [Haykin, 1994]. Thus, \( \mathbf{W} \) is the Wiener filter.

In neural networks, the matrix \( \mathbf{W} \) of eqn. (4-21) is the least squares solution to the pattern association problem. This matrix embodies a linear single-layer network which transforms the system input via \( \mathbf{Y} = \mathbf{WX} \). It is a linear associative memory (LAM) which is also referred to as the pseudoinverse memory [Haykin, 1994]. This is because the matrix \( \mathbf{X}^T (\mathbf{X}\mathbf{X}^T)^{-1} \) in eqn. (4-21) is commonly termed the pseudoinverse of \( \mathbf{X}^T \) [Golub and Van Loan, 1989].

We can relate the least squares solution of eqn. (4-21) to probability theory. Previously, we alluded to the fact that the least squares solution minimizes the mean squared error when the number of observations \( N \) of a random vector approaches infinity. It is known that for any distribution, the estimator which minimizes the sum of the mean squared error \( J = \sum_{i=1}^{N} E[(d_i - \overline{d}_i)^2] = E[(\mathbf{d} - \overline{\mathbf{d}})^T (\mathbf{d} - \overline{\mathbf{d}})] \) is the conditional expected value \[=\] [Helstrom, 1991] and \( J \) is minimized for

\[
\mathbf{d} = E[\mathbf{d} | \mathbf{x}] \tag{4-24}
\]

where \( \mathbf{d} \) is the conditional expectation of the random vector \( \mathbf{d} = [d_1, d_2, \ldots, d_p]^T \) given an observation \( \mathbf{x} = [x_i, x_2, x_3, \ldots, x_Q = x_Q]^T \) of the random vector \( \mathbf{x} = [x_1, x_2, \ldots, x_Q]^T \) and \( \mathbf{d} = [d_1, \ldots, d_p] \). If the joint conditional PDF \( p_{\mathbf{d}\mathbf{x}}(\mathbf{d}, \mathbf{x}) \) has a multivariate Gaussian form,
then the conditional expected value of eqn. (4-24) is known to be a linear transformation of the observed data given by \( \mathbf{x} \). In particular, the conditional expected value is

\[
\mathbf{d}_i = E[\mathbf{d} | \mathbf{x}] = \sum_{j=1}^{Q} b_{ij} x_j = \mathbf{Bx}, \quad i = 1, \ldots, P
\]

where the \( \mathbf{d}_i \) are elements of vector \( \mathbf{d} \) and \( \mathbf{B} = C_{\mathbf{d}_x} C_{\mathbf{x}\mathbf{x}}^{-1} \) [Helstrom, 1991]. The linear transformation \( \mathbf{B} \) is comprised of the multiplication of two covariance matrices. Notice that \( \mathbf{B} \) is the Wiener filter for zero mean random vectors \( \mathbf{x} \) and \( \mathbf{d} \). It is the same linear transformation as eqn. (4-21) for an infinite amount of data samples. Thus, the optimal association of observations \( \mathbf{x}_r \) and \( \mathbf{d}_r \) from a distribution whose joint density \( p_{\mathbf{d}_x}(\mathbf{d}, \mathbf{x}) \) is multivariate Gaussian and zero mean is the linear transformation given by the Wiener filter. For a finite number of data, this filter can be estimated using eqn. (4-21), (4-22) or (4-23).

### 4.4.3.2 Multilayer perceptrons

The multilayer perceptron (MLP) provides for a nonlinear mapping of an input exemplar \( \mathbf{x} \) [Haykin, 1994]. It is a feedforward network that generalizes the transformations afforded by linear networks on exemplar \( \mathbf{x} \). Note that an MLP can also be referred to as a nonlinear associative memory (NLAM) following the parallel that LAMs are linear networks (as mentioned in the previous section).

The training algorithm for MLPs generalizes the LMS training algorithm of linear networks to MLPs. The on-line weight update is similar to that of the LMS algorithm and is typically termed the error backpropagation algorithm. At each discrete time step \( n \) the weight matrix \( \mathbf{W}_k \), for layer \( k \) of the MLP, is updated as follows

\[
\mathbf{W}_k(n + 1) = \mathbf{W}_k(n) + \mu\mathbf{g}_k(n)\mathbf{y}^{T}_{r,k-1}
\]
where \( g_k(n) \) is the local gradient vector and \( y_{r,k-1} \) is the post neural activity vector of the previous layer (hence the \( k-1 \)) due to input vector \( x_r \). Please note here that the subscript \( k \) describes the weight layer of a feedforward NLAM with several layers. The post neural activity of an NLAM at a given layer is recursively defined as \( y_{r,k} = \varphi(W_k(n)y_{r,k-1}) \). This is because the output of one layer serves as the input to the next layer in the feedforward configuration. The function \( \varphi(\cdot) \) is bounded and differentiable and is termed a “squashing” function. A typical function is \( \varphi(x) = \tanh(x) \). Note that the post neural activity for layer 0 at time step \( n \) is defined as just the input vector at that time step, that is, \( y_{r,0} = x_r \).

If layer \( k \) is the output layer then

\[
g_k(n) = (d_r - y_{r,k}) \bullet \varphi'(y_{r,k})
\]

where \( \bullet \) represents the element-by-element multiplication or Schur product of two matrices (or vectors) and \( \varphi'(\cdot) \) is the first derivative of \( \varphi(\cdot) \). If layer \( k \) is other than the output layer then

\[
g_k(n) = (W^T_{k+1}(n)g_{k+1}(n)) \bullet \varphi'(y_{r,k}).
\]

### 4.4.3.3 Association of neighborhoods

The information to associate is chosen to be the local structure of images. The local structure of an image is defined here as an image neighborhood with its mean subtracted out; each image neighborhood thus becomes a vector whose component mean is zero. This can be obtained by lexicographically ordering an \( H_1 \times H_2 \) neighborhood as a column vector and multiplying with a certain Toeplitz matrix. Specifically, if \( x \) is a 2D neighborhood, its structure \( s \), in column form, is obtained by \( s = Zvec(x) \) where
\[
Z = \frac{1}{H_1H_2} \begin{pmatrix}
H_1H_2 -1 & -1 & \cdots & -1 \\
-1 & H_1H_2 -1 & \ddots & \\
\vdots & \ddots & \ddots & -1 \\
-1 & \cdots & -1 & H_1H_2 -1
\end{pmatrix}
\]

and where \( \text{vec}(\cdot) \) is the lexicographic ordering operation which stacks the columns of a matrix one on top of the other to form a vector. Let us represent the set of input vectors as a matrix \( \mathbf{X} \) and let each column \( \mathbf{x}_r \) of this matrix be reshaped 2D neighborhoods from a low resolution image \( x_i[n_1,n_2] \). The corresponding output vectors \( \mathbf{y}_r \) which will be the local high resolution image structure created by the LAMs will similarly be contained in a matrix \( \mathbf{Y} \). The desired output vectors are also similarly defined and are contained in a matrix \( \mathbf{D} \). Each column in \( \mathbf{D} \), which is the desired exemplar \( \mathbf{d}_r \), is obtained by subtracting the mean of \( \mathbf{x}_r \) from its corresponding homologous high resolution neighborhood. This is done to compensate for the low resolution neighborhood mean which will be subtracted from \( \mathbf{x}_r \) and must be added back after the high resolution neighborhood structure is created. Specifically, \( \mathbf{D} = \mathbf{S} - \mathbf{A} \mathbf{X} \) where \( \mathbf{S} \) is a matrix whose columns \( \mathbf{s}_r \) are reshaped \((2G_1 - 1) \times (2G_2 - 1)\) sized neighborhoods of the high resolution image \( x_h[n_1,n_2] \) and \( \mathbf{A} \) is a \((2G_1 - 1)(2G_2 - 1) \times H_1H_2\) constant matrix with elements \( \frac{1}{H_1H_2} \). Note that fig. (4-10) illustrates the neighborhoods in \( \mathbf{X} \) and \( \mathbf{S} \) which have been reshaped into column vectors. For a given neighborhood (and decimation factor \( G_1 \times G_2 \)), \( \mathbf{x}_r \) corresponds to the shaded circles and \( \mathbf{s}_r \) corresponds to the crossed circles. For proper association it is generally required that the number of input vectors be much larger than the dimensionality of the input space hence resulting in the solution of an overdetermined problem rather than an underdetermined one.

The input-output relationship of a LAM is described by
\[ y_r = WZx_r + b \]  \hspace{1cm} (4-25)

where \( W \) is a weight matrix that specifies the network connectivity of the LAM, \( b \) is a bias vector and \( Z \) extracts the neighborhood structure from the neighborhood represented by \( x_r \). \( y_r \) contains a vector representation of a superresolved 2D neighborhood structure and \( WZ \) plays the role of a kernel à la eqn. (3-9) with an offset of \( b \).

The optimal association for each LAM is defined here as minimizing the mean squared error. That is, if \( d_r \) is the desired output and \( y_r \) the actual output for a given input \( x_r \), then the cost function to be minimized is

\[ E = \sum_r (d_r - y_r)^T (d_r - y_r) \]

Let us briefly pause here and establish the relation between eqn. (4-1) and what has been presented thus far. Eqn. (4-25) computes the output of a single layer feedforward linear net. Each \( x_r \), which we have established as a local neighborhood, is transformed by \( Z \) and then weighted by the rows of a matrix \( W \) and offset by a column vector \( b \). There are \( C \) LAMs and thus \( C \) weight matrices \( W_c \) and column vectors \( b_c, c = 1, \ldots, C \). Each row of \( W_cZ \) and \( b_c \) linearly combines the neighborhood samples in \( x_r \), so these rows constitute the \( L \) kernels associated with each LAM. The \( c \) and \( l \) subscripts in \( k_{c,l} \) indicate the kernel to be used, that is, what row of \( W_cZ \) and \( b_c \) is being referenced.

The optimal weight matrix of a LAM can be obtained recursively via the LMS algorithm or in closed form via the pseudo-inverse as discussed earlier. We have assumed that \( W \) is actually the augmented matrix \([W | b]\) and \( X \) is the augmented matrix \([X^T Z^T | v]^T\) where \( v \) is a column vector of ones. This is done to provide for an affine mapping of the input. Recall that \( ZX \) is the local structure of each low resolution neighborhood.
When considering nonlinear associative memories (NLAMs), we see that an NLAM has an input-output relation of the form

\[ y_r = \kappa(\{W_k\}, \{b_k\}, Z_r) \]

where, in general, \( \kappa(\cdot) \) is a nonlinear function of a set of weight matrices, bias vectors and the local neighborhood structure. Here, an NLAM is a multilayer perceptron (MLP) trained for association. In the nonlinear case, the weight matrices are updated recursively using the backpropagation algorithm. This algorithm is well established as a supervised training method for neural networks and was discussed earlier. Here, there are \( C \) NLAMs to be trained. Each corresponds to a particular cluster of the input data. NLAM \( c \) associates the neighborhoods \( x_r \in K_c \) with its corresponding samples in \( d_r \) as was done with the LAMs.

4.5 Reconstruction Phase

Once the LAMs have been trained, the construction of high resolution images is a simple matter of feeding the local low resolution neighborhoods into the proper LAMs. These LAMs were depicted in fig. (4-8).

All system parameters for the superresolution phase are now fixed. Low resolution neighborhoods of size \( H_1 \times H_2 \) from images not previously seen before by the network are first extracted. Their local structure is obtained and then compared against features that have been established. These are the feature vectors obtained from the SOFM. \( C \) clusters \( K_c, c = 1,2,\ldots,C \) are then formed by assigning each neighborhood to the cluster whose feature vector is closest in Euclidean distance to its local structure.
The construction of high resolution images is then a straightforward transformation of a neighborhood’s local structure. The mean of the neighborhood is subsequently added back to the transformation. In the case of a LAM, the superresolution of a low resolution neighborhood \( x_r \) can be expressed as

\[
\hat{s}_r = uvec(\mathbf{W}_c \mathbf{Z} x_r + \mathbf{b}_c + \mathbf{A} x_r) \quad \text{for} \quad x_r \in K_c
\]  

(4-26)

where \( \mathbf{W}_c \) is the weight matrix associated with the \( c^{th} \) LAM, \( \mathbf{A} \) is the matrix previously mentioned which is responsible for adding back the mean that was subtracted out and \( uvec(\cdot) \) is the operation that undoes the \( vec(\cdot) \) operation, that is, it makes a 2D neighborhood from a column vector. As discussed before, there is a direct relation between eqn. (4-26) and eqn. (4-1). Eqn. (4-26) constructs all the high resolution neighborhoods \( \hat{s}_r \). The subscript \( r \) references the neighborhood to be reconstructed. The reconstructed pixels that overlap are then averaged and the high resolution image is thus constructed. Eqn. (4-26), following the form of eqn. (4-1), can be equivalently expressed as

\[
\hat{s}_h[n_1, n_2] = \sum_{l=1}^{L} [x_e[n_1, n_2] * (k_{c,l}[n_1, n_2] + a[n_1, n_2])] \cdot b[n_1, n_2]
\]

where \( L = (2G_1 - 1)(2G_2 - 1) \), \( x_e \) is the low resolution image that has been expanded by zero insertion as in eqn. (3-6), the kernel was created with the values of \( \mathbf{W}_c \mathbf{Z}(l,:) \) and \( \mathbf{b}_c(l) \), that is, row \( l \) of \( \mathbf{W}_c \mathbf{Z} \) and \( \mathbf{b}_c \), \( a \) is a constant kernel with the same extent as \( k_{c,l} \) that averages a low resolution neighborhood (its impulse response samples equal \( \frac{1}{n_1 n_2} \)) and \( b[n_1, n_2] = b_1[n_1]b_2[n_2] \) is responsible for averaging multiple estimates of reconstructed samples. Specifically,
\[
b_i[n_i] = \begin{cases} 1 & n_i \mod G_i = 0 \\ \frac{1}{2} & \text{otherwise} \end{cases}
\]
for \( i = 1,2 \). Notice that the index \( i \) refers to a specific convolution pass that is constructing the corresponding enumerated crossed circle associated with each low resolution neighborhood in that pass. Please refer to fig. (4-10a) for the case of \( G_1 = G_2 = 2 \). 

The NLAM case only differs by the presence of the nested nonlinearities. The construction, for an \( M \) layer MLP topology, is expressed as 

\[
\hat{s}_r = uvec(\varphi(W_{c,M}\varphi(\cdots \varphi(W_{c,1}Zx + b_{c,1}) + b_{c,M}) + Ax)) \quad \text{for } x_r \in K_c
\]

where \( W_{c,k} \) is the weight matrix at layer \( k \) of the \( c^{th} \) NLAM. Note that there are as many nested nonlinear transformations as there are layers in the feedforward network.

### 4.6 Relation to the Mixture of Experts

Our superresolution methodology for optical images is essentially attempting to find the proper association between low and high resolution images locally. It was mentioned in Chapter 2 that the mixture of experts (MOE) could be used in this modeling. We now introduce the MOE methodology and compare it to our approach.

The MOE is a topology introduced in the neural network literature for estimating the conditional probability of one data set given another [Jacobs et al., 1991; Jacobs and Jordan, 1991]. It is a structure that consists of \( C \) supervised modules called expert networks and an integrating unit called a gating network which mediates among the expert networks. It is a modular architecture that combines associative and competitive learning in order to allow description of different intrinsic signal characteristics by separate
modules. It associates between two sets of data, \( x \) and \( d \), by modeling the density of \( d \) given \( x \). The conditional PDF modeled by the MOE is a mixture Gaussian distribution

\[
p_{dx}(d \mid x) = \sum_{c=1}^{C} P(c \mid x)p_{dx}(d \mid x, c)
\]

(4-27)

where \( C \) represents the total number of local PDF models used in characterizing the true conditional density, the mixing parameters are given by the probability \( P(c \mid x) \) and the local Gaussian model is given by

\[
p_{dx}(d \mid x, c) = \frac{1}{(2\pi)^{P/2}|C_c|^{rac{1}{2}}} \exp\left(-\frac{1}{2}(d - y_c)^T C_c^{-1}(d - y_c)\right)
\]

where \( P \) is the dimensionality of \( d \) and \( C_c \) is the covariance matrix associated with model (or “class”) \( c \). Each of the \( C \) experts in the MOE is responsible for modeling \( p_{dx}(d \mid x, c) \), \( c = 1, \ldots, C \). Their roles are analogous to that of an associative memory. If the experts’ transformation is linear or affine, then we are dealing with LAMs. If they are nonlinear, then we are dealing with NLAMs.

The gating network is responsible for modeling \( P(c \mid x) \). It is a differentiable function whose role is to provide a “degree of confidence” as to how much the \( c^{th} \) expert will contribute to the association of \( x \) with \( d \). Because this confidence measure results from a differentiable function that only depends on the input \( x \), we say that it provides a “soft” segmenting of the input space. This is in sharp contrast to the “hard” segmenting discussed earlier for our superresolution methodology. Two gating network structures are reported in the literature: a single layer gating network [Jacobs et al., 1991; Jacobs and Jordan, 1991] and a hierarchical gating network [Jordan and Jacobs, 1992, 1994]. Both
of these structures model \( P(c|x) \) as previously discussed but the hierarchical gating structure provides a greater flexibility into how the input space of \( x \) is “soft” segmented.

The gating function, in our methodology, would be expressed as

\[
P(c \mid x) = \begin{cases} 
1 & \gamma(x) = c \\
0 & \text{otherwise}
\end{cases}
\]

The function \( \gamma \) is responsible for clustering (classifying) \( x \) into one of \( C \) mutually exclusive groups, that is, \( \gamma : x \rightarrow \{1, \ldots, C\} \). This would lead to expressing eqn. (4-27) as

\[
p_{dx}(d \mid x) = p_{dx}(d \mid x, c) \text{ for all } \gamma(x) = c
\]

This formulation can be viewed as a classification problem where the expert responsible for generating \( d \) given \( x \) is assumed known. In this approach the function \( \gamma(x) \) is non-differentiable. It can be seen as a winner-take-all type of function. In this case, it is known a priori which expert was responsible for generating the desired vector \( d \). Each expert could be trained independent of the others. Using a mean squared error training criteria and modeling each expert as an affine transformation on the input data as the MOE does, the training of each expert could be performed on-line using the LMS algorithm or the experts could be computed in closed form via the pseudo-inverse. This is how each LAM was trained in our superresolution approach. The classification function \( \gamma(x) \) was obtained in eqn. (4-14). We were able to establish such a function from the interdependent information that was studied earlier in this chapter. From such interdependencies, the neighborhoods in \( x \) and \( d \) were then associated using LAMs.

From a modeling perspective, the MOE is achieving a model of the input space as a mixture Gaussian model which is the natural extension to the simple Gaussian assumption. This model, when used for reconstruction, is an extension of Ruderman and
Bialek’s work on optimal reconstruction. However, the theoretical work to show this formal extension is left for future research.
CHAPTER 5
EXPERIMENTAL RESULTS FOR OPTICAL IMAGES

5.1 Analysis and Comparison

The results reported in this section are a sampling of the tests that have been conducted on optical images. Results are reported using the four images of fig. (4-1) as well as the texture images of fig. (5-1). The results obtained with the superresolution architecture are analyzed and compared against several common techniques for image interpolation. A comparison was also made against the subpixel edge localization and interpolation (SEL) technique which fits an ideal step edge through those image regions where an edge was deemed to exist and otherwise uses a bilinear interpolation between samples [Jensen and Anastassiou, 1995]. The parameters for the SEL technique we used were the same as those reported by Jensen and Anastassiou.

Table 5-1 reports on the PSNR resulting from interpolating the 128x128 images of fig. (4-1) by a factor of 2 in each dimension using the common techniques. The PSNR was defined in eqn. (4-11). These numbers are provided as a quantitative basis for comparison between our results and the interpolation techniques mentioned in this study. The PSNRs are very similar for each of the techniques listed in table 5-1. Even though the PSNR measure is not comparable to the fidelity measure human viewers use in judging image quality, it serves to provide insight into the performance of the algorithm.
Figure 5-1. Texture images used to test superresolution performance. (a) 256×256 images: Asphalt_1 (top left), Asphalt_2 (top right), Grass_1 (bottom left), Grass_2 (bottom right); (b) 128×128 images obtained by the image acquisition model - (left to right) Asphalt_1, Asphalt_2, Grass_1 and Grass_2.
The bicubic and cubic B-spline approaches generally outperform the bilinear method due to their assumptions about a smoothly varying image structure. The SEL approach typically works well in edgy image regions as long as there is sufficient resolution about the edge being modeled. A numerical comparison of PSNR using the zero-order-hold kernel has been excluded as the blockiness produced by this algorithm is unnatural to the reconstruction of real-world images and typically results in relatively lower PSNR.

Table 5-1. PSNR for several images using various interpolation methods. The interpolation factor was 2 in each direction from the listed 128x128 images. These results are compared with the results from our superresolution methodology in figs. (5-2) - (5-4).

<table>
<thead>
<tr>
<th></th>
<th>Lena</th>
<th>Peppers</th>
<th>Pentagon (left)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilinear</td>
<td>27.26</td>
<td>27.51</td>
<td>27.11</td>
</tr>
<tr>
<td>Bicubic</td>
<td>27.45</td>
<td>27.74</td>
<td>27.31</td>
</tr>
<tr>
<td>Cubic-B</td>
<td>27.43</td>
<td>27.74</td>
<td>27.33</td>
</tr>
<tr>
<td>SEL</td>
<td>27.48</td>
<td>27.99</td>
<td>27.02</td>
</tr>
</tbody>
</table>

The two texture images considered were of patches of asphalt and grass. Two separate patches from the same piece of asphalt and grass, respectively, are used for our testing. As such the texture images are denoted Asphalt_1, Asphalt_2, Grass_1 and Grass_2. Table 5-2 reports on the reconstruction of the second patch of the texture images of fig. (5-1) using the common methods.

We have chosen not to include the results of the common techniques within the plots of our superresolution methodology because the change in plot scaling resulting from their inclusion would obscure certain information we wish to convey about our approach.
Table 5-2. PSNR for texture images using various interpolation methods. The interpolation factor was 2 in each direction from the listed 128x128 images. These results are compared with the results from our superresolution methodology in figs. (5-11) and (5-12).

<table>
<thead>
<tr>
<th>Method</th>
<th>Asphalt_2</th>
<th>Grass_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilinear</td>
<td>31.00</td>
<td>26.46</td>
</tr>
<tr>
<td>Bicubic</td>
<td>31.61</td>
<td>26.84</td>
</tr>
<tr>
<td>Cubic-B</td>
<td>31.64</td>
<td>26.92</td>
</tr>
<tr>
<td>SEL</td>
<td>30.70</td>
<td>26.45</td>
</tr>
</tbody>
</table>

5.1.1 Linear Associative Memories

We now turn our attention to analyzing the results obtained with our superresolution procedure using LAMs. The PSNRs using the common techniques have been listed separately in tables 5-1 and 5-2.

5.1.1.1 Standard images

Superresolution results with the standard images of fig. (4-1) are reported here. The plot in fig. (5-2) illustrates the PSNR when superresolving the left Pentagon 128x128 image by a factor of 2 in each dimension with varying number of LAMs. The images used in training our system come from the left Pentagon 256x256 image. The trained system was then used to superresolve the right Pentagon 128x128 image. In this case, the similarity in image information between the left and right stereo images is intuitively expected to yield good generalization performance on the test image of the right pentagon. This was indeed the case.
Figure 5-2. Training and testing superresolution results for the left Pentagon image considering two different regions of support (ROS). The solid lines correspond to training set results and the dashed lines are test set results. The curves related to the training data result from superresolving the left Pentagon 128×128 image with the systems (features and LAMs) established to go from the left Pentagon 128×128 image to the left Pentagon 256×256 image. The curves related to testing result from superresolving the left Pentagon 128×128 image with the systems established to go from the right Pentagon 128×128 image to the right Pentagon 256×256 image. Superresolved images corresponding to the two circled points ‘o’ are shown in fig. (5-5).
The solid lines in fig. (5-2) denote training set reconstruction performance using two different regions of support (ROS): 3×3 and 5×5. The dashed lines illustrate test set performance using these regions of support. There is a general trend of increasing PSNR in the training set as the number of LAMs increased. This is intuitively expected as an increase in the number of LAMs yields a greater specialization to particular image features, hence a more accurate image reconstruction. The feature set extracted using a 5×5 ROS yields more macroscopic image characteristics than does a 3×3 ROS. This results in greater specialization of the characteristics particular to the image of interest and generally to a more faithful image reconstruction of the training set. In the test set, however, the larger ROS tended to show a drop in PSNR performance as the degree of specialization increased. This result was generally encountered in all the tests we have run on nontextured imagery. It suggests that the similarity between features, as we specialize more (use more LAMs), tends to occur at a more microscopic level. It can also be observed that our superresolution approach yielded higher PSNR than those methods listed in table 5-1.

In other tests reported here, we trained a set of systems (with varying numbers of LAMs) to reconstruct from the Lena 128×128 image to the Lena 256×256 image as well as training a separate set of systems to reconstruct from the Peppers 128×128 image to the Peppers 256×256. Training and test set performance are illustrated in figs (5-3) and (5-4). Fig. (5-3) reports on the superresolution of the Lena 128×128 image by a factor of two. The training set performance illustrates similar behavior to that discussed in fig. (5-2) for the left Pentagon image.
Figure 5-3. Training and testing superresolution results for the Lena image considering two different regions of support (ROS). The solid lines correspond to training set results and the dashed lines are test set results. The curves related to the training data result from superresolving the Lena 128×128 image with the systems (features and LAMs) established to go from the Lena 128×128 image to the Lena 256×256 image. The curves related to testing result from superresolving the Lena 128×128 image with the systems established to go from the Peppers 128×128 image to the Peppers 256×256 image. Superresolved images corresponding to the two circled points ‘o’ are shown in fig. (5-6).
Figure 5-4. Training and testing superresolution results for the Peppers image considering two different regions of support (ROS). The solid lines correspond to training set results and the dashed lines are test set results. The curves related to the training data result from superresolving the Peppers 128×128 image with the systems (features and LAMs) established to go from the Peppers 128×128 image to the Peppers 256×256 image. The curves related to testing result from superresolving the Peppers 128×128 image with the systems established to go from the Lena 128×128 image to the Lena 256×256 image. Superresolved images corresponding to the two circled points ‘o’ are shown in fig. (5-7).
In the test set curves of fig. (5-3), the Lena 128×128 image was superresolved using the parameters (features and LAMs) established by training to superresolve the Peppers 128×128 image by a factor of 2. Again, similar test set performance to that reported earlier for the Pentagon image can be seen. Similar results to those discussed are also observed in fig. (5-4) on the superresolution of the Peppers 128×128 image by a factor of two. The training curves (solid lines) of this figure were obtained by reconstructing with the systems trained to go from the Peppers 128×128 to the Peppers 256×256. The test curves report on the superresolution of the Peppers 128×128 image with the systems trained to go from the Lena 128×128 image to the Lena 256×256 image. The results of figs. (5-3) and (5-4) suggest that there is much similarity in the local structure of the Lena and Peppers images.

A visual comparison of the results utilizing the common approaches in table 5-1 and our superresolution methodology for the left Pentagon, Lena and Peppers images can be seen in figs (5-5)-(5-7), respectively. The training and testing images shown in each of these figures were created using 30 LAMs and a ROS of 3×3. They correspond to those points on the PSNR figures marked by a circled point ‘o’. The superresolved training and testing images were of similar quality. Notice that, in general, the PSNR for the reconstructed images using the LAM approach is higher than with the common approaches listed in table 5-1. The PSNR measure largely reflects reconstruction performance over the low frequency portions of our images. This is because the proportion of high frequency to low frequency regions in real images is typically small, that is, there are typically many more locally smooth image portions than there are edgy or textured portions. As such, the LAM approach is performing well in reconstructing the
smooth image regions (this is difficult for us to visually perceive) as well as the high frequency regions (which we can perceive more readily). In figs. (5-8)-(5-10) we have shown the magnitude spectra of the reconstructed images corresponding to figs. (5-5)-(5-7), respectively. The spectra here are for the full reconstructed images in figs. (5-5)-(5-7), not just the portions shown in these figures. It is evident from viewing these spectra that the LAM based approach is reproducing better the high frequency information characteristic of the original images. The SEL approach is also able to reproduce high frequency information. This is because, as mentioned earlier, the SEL approach fits an ideal step edge through those image regions it deems are edges. The low PSNR of the SEL approach can be attributed to its performance in smoothly varying image regions. This is because, in these regions, the SEL approach uses the bilinear kernel for its interpolation. In summary, the superresolved images of this work generally appear more crisp than those obtained with the other approaches presented here. Edges seemed to be preserved well with our approach and the higher PSNR obtained with our methodology is evidence of the accuracy of reconstruction in smoothly varying image regions relative to the other approaches reported herein.
Figure 5-5. Visual comparison of the reconstruction results for the left Pentagon image. This image is part of a stereo pair. The 128×128 image was reconstructed to 256×256. A zoomed section of the reconstructed results is displayed. The ‘training’ reconstruction utilized the 30 features and corresponding LAMs obtained in going from the left Pentagon 128×128 image to the left Pentagon 256×256 image with ROS of 3×3. The ‘testing’ reconstruction utilized the 30 features and corresponding LAMs obtained in going from the right Pentagon 128×128 image to the right Pentagon 256×256 image with ROS of 3×3.
Figure 5-6. Visual comparison of the reconstruction results for the Lena image. The $128\times128$ image was reconstructed to $256\times256$. A zoomed section of the reconstructed results is displayed. The ‘training’ reconstruction utilized the 30 features and corresponding LAMs obtained in going from the Lena $128\times128$ image to the Lena $256\times256$ image with ROS of $3\times3$. The ‘testing’ reconstruction utilized the 30 features and corresponding LAMs obtained in going from the Peppers $128\times128$ image to the Peppers $256\times256$ image with ROS $3\times3$. 
Figure 5-7. Visual comparison of the reconstruction results for the Peppers image. The 128×128 image was reconstructed to 256×256. A zoomed section of the reconstructed results is displayed. The ‘training’ reconstruction utilized the 30 features and corresponding LAMs obtained in going from the Peppers 128×128 image to the Peppers 256×256 image with ROS of 3×3. The ‘testing’ reconstruction utilized the 30 features and corresponding LAMs obtained in going from the Lena 128×128 image to the Lena 256×256 image with ROS of 3×3.
Figure 5-8. Magnitude spectra of the reconstructed Left Pentagon images in fig. (5-5). The spectra here are for the corresponding full reconstructed images, not just the zoomed sections pictured in fig. (5-5). The spectra $\mathcal{F}[n_1, n_2]$ of each image has been enhanced via the log scaling $rac{2}{\log_{10}(\|\mathcal{F}\|_0)} \log_{10}(\|\mathcal{F}[n_1, n_2]\|)$. The LAM and SEL approaches are better able to produce higher frequency information relative to the other methods compared.
Figure 5-9. Magnitude spectra of the reconstructed Lena images in fig. (5-6). The spectra here are for the corresponding full reconstructed images, not just the zoomed sections pictured in fig. (5-6). The spectra $\mathcal{F}[n_1, n_2]$ of each image has been enhanced via the log scaling $\log_{10}(\mathcal{F}(0,0))\log_{10}(\mathcal{F}[n_1, n_2])$. The LAM and SEL approaches are better able to produce higher frequency information relative to the other methods compared.
Figure 5-10. Magnitude spectra of the reconstructed Peppers images in fig. (5-7). The spectra here are for the corresponding full reconstructed images, not just the zoomed sections pictured in fig. (5-7). The spectra $\mathcal{S}[n_1, n_2]$ of each image has been enhanced via the log scaling $\frac{1}{\log_{10}(\mathcal{S}(0,0))} \log_{10}(\mathcal{S}[n_1, n_2])$. The LAM and SEL approaches are better able to produce higher frequency information relative to the other methods compared.
5.1.1.2 Texture images

The results pertaining to the superresolution of the texture images of fig. (5-1) are illustrated in figs. (5-11) and (5-12). These figures illustrate, respectively, the training and testing results of reconstructing the Asphalt_2 and Grass_2 128×128 images by a factor of 2 in each axis. In fig. (5-11), the training results are obtained by reconstructing the Asphalt_2 128×128 image with the features and LAMs trained to reconstruct the Asphalt_2 256×256 image from the Asphalt_2 128×128 image. The testing results are obtained by reconstructing the Asphalt_2 128×128 image with the features and LAMs trained to reconstruct the Asphalt_1 256×256 image from the Asphalt_1 128×128 image. In fig. (5-12), the training results are obtained by reconstructing the Grass_2 128×128 image with the features and LAMs trained to reconstruct the Grass_2 256×256 image from the Grass_2 128×128 image. The testing results are obtained by reconstructing the Grass_2 128×128 image with the features and LAMs trained to reconstruct the Grass_1 256×256 image from the Grass_1 128×128 image. A visual comparison of the superresolution results for the texture images presented is given in figs. (5-13) and (5-14). Comparisons using the same common methods as with the standard images are shown. Notice that the superresolution results in these two figures correspond to the circled points ‘o’ in the plots of figs. (5-11) and (5-12).

We can notice in these plots that the superresolution results yielded higher PSNR for the 5×5 ROS as compared with the 3×3 ROS — in both training and test set performance. This is in contrast to the results for the nontexture images of fig. (4-1).
Figure 5-11. Training and testing superresolution results for the Asphalt_2 texture image considering two different regions of support (ROS). The solid lines correspond to training set results and the dashed lines are test set results. The curves related to the training data result from superresolving the Asphalt_2 128×128 image with the systems (features and LAMs) established to go from the Asphalt_2 128×128 image to the Asphalt_2 256×256 image. The curves related to testing result from superresolving the Asphalt_2 128×128 image with the systems established to go from the Asphalt_1 128×128 image to the Asphalt_1 256×256 image. Superresolved images corresponding to the two circled points ‘o’ are shown in fig. (5-13).
Figure 5-12. Training and testing superresolution results for the Grass_2 texture image considering two different regions of support (ROS). The solid lines correspond to training set results and the dashed lines are test set results. The curves related to the training data result from superresolving the Grass_2 128×128 image with the systems (features and LAMs) established to go from the Grass_2 128×128 image to the Grass_2 256×256 image. The curves related to testing result from superresolving the Grass_2 128×128 image with the systems established to go from the Grass_1 128×128 image to the Grass_1 256×256 image. Superresolved images corresponding to the two circled points ‘o’ are shown in fig. (5-14).
Figure 5-13. Visual comparison of the reconstruction results for the Asphalt_2 image. This image is part of a texture pair. The 128×128 image was reconstructed to 256×256. A zoomed section of the reconstructed results is displayed. The ‘training’ reconstruction utilized the 30 features and corresponding LAMs obtained in going from the Asphalt_2 128×128 image to the Asphalt_2 256×256 image with ROS of 5×5. The ‘testing’ reconstruction utilized the 30 features and corresponding LAMs obtained in going from the Asphalt_1 128×128 image to the Asphalt_1 256×256 image with ROS of 5×5 to reconstruct the Asphalt_2 image.
Figure 5-14. Visual comparison of the reconstruction results for the Grass_2 image. This image is part of a texture pair. The 128×128 image was reconstructed to 256×256. A zoomed section of the reconstructed results is displayed. The ‘training’ reconstruction utilized the 30 features and corresponding LAMs obtained in going from the Grass_2 128×128 image to the Grass_2 256×256 image with ROS of 5×5. The ‘testing’ reconstruction utilized the 30 features and corresponding LAMs obtained in going from the Grass_1 128×128 image to the Grass_1 256×256 image with ROS of 5×5 to reconstruct the Grass_2 image.
The features particular to the texture images are more difficult to heuristically characterize as compared with the features of the images of fig. (4-1). The images of fig. (4-1) had many edge features associated with them. Those features were found to translate well to test images for their superresolution using 3×3 regions of support. The better reconstruction performance with a larger ROS for the texture images suggests a more regular or recurring local structure for these images as compared with the nontexture images. The presence of a recurring structure in textures is intuitively expected. As such, the need for increasingly specialized features (as the ROS gets smaller) is not as necessary for "good" reconstruction performance in texture images.

5.1.1.3 Additional results and comments

Thus far we have presented reconstruction results for increasing the sample density by 4, that is, \( G_1 = G_2 = 2 \). Fig. (5-15) illustrates results when reconstructing by a factor of 9 \( (G_1 = G_2 = 3) \). Here, a Lena image which was 170×170 in size was reconstructed to a size of 510×510. The low resolution image was obtained via the decimation model of Chapter 4 from a 510×510 Lena image. The testing image which is pictured results from reconstructing the Lena 170×170 image with the features and LAMs used to reconstruct the Peppers 510×510 image from the Peppers 170×170 image. We can see in this figure that the LAM based superresolution procedure results in a more faithfully reconstructed image as compared to the other techniques. In fig. (5-16), we see the results of reconstructing the Lena 128×128 image by a factor of 4·4 = 16 utilizing two successive \( G_1 = G_2 = 2 \) reconstruction stages. For the test image, the 30 features and LAMs used were the ones obtained in training to reconstruct the Peppers 256×256 image from the Peppers 128×128 image with an ROS of 3×3. To perform the reconstruction, we first
reconstruct the Lena 128×128 image by a factor of 2 through each image axis. The resulting image is then further reconstructed by a factor of 2 on each axis. We used the same features and LAMs on both reconstruction stages. We can notice that the LAM reconstructed image is crisper than the other images. This further helps to reinforce our observations of the interdependence of image neighborhoods across scales which can be exploited for superresolution.

Fig. (5-17) illustrates a case of what can happen when inappropriate sets of bases are used for the image reconstruction. In this figure, the Lena 128×128 image is being reconstructed to 256×256. The original image is given in fig. (5-17a). The image reconstructed from the 30 features and LAMs obtained in training to reconstruct the Peppers 256×256 image from the Peppers 128×128 image is given in fig. (5-17b) and the image reconstructed from the 30 features and LAMs obtained in training to reconstruct the Left Pentagon 256×256 image from the Left Pentagon 128×128 image is given in fig. (5-17c). It has been shown (in figs. (5-3), (5-6) and (5-15)) that the systems of the Peppers is appropriate for the reconstruction of Lena. However, the systems of the Left Pentagon are not as appropriate for the reconstruction of Lena. This is particularly seen by the reconstruction performance about the right portion of the forehead and hat in fig. (5-17c). Incorporating the correct a priori information into the reconstruction process can be beneficial to superresolution but introducing the wrong information can have the opposite effect. In fig. (5-17d) we compensate for the lack of proper bases by incorporating the appropriate bases obtained from the Peppers image used in reconstructing (5-17b). The appropriate bases were simply “appended” to the inappropriate set from the Left Pentagon that was used in this example.
Figure 5-15. Example of reconstruction by a factor of 9 (3×3). A zoomed section of the reconstructed results is displayed. The low resolution Lena image was 170×170. It was reconstructed to 510×510 using several kernels. The ‘test’ image utilized the 30 features and LAMs trained to reconstruct the Peppers 510×510 image from the Peppers 170×170 image with an ROS of 5×5.
Figure 5-16. Example of reconstruction of Lena 128×128 image by a factor of 16 (a factor of 4 on each image axis). The reconstruction was accomplished using two successive stages of reconstruction, each by a factor of 4. The same features and LAMs were used in each stage. The test image was reconstructed using the features and LAMs used in reconstructing the Peppers 256×256 image from the Peppers 128×128 image with a ROS of 3×3.
Figure 5-17. Compensating for the effects of reconstruction with “inappropriate” bases. The results displayed show a portion of the Lena 256×256 image reconstructed from the Lena 128×128 image. (a) Original (b) Reconstructed with the 30 features and LAMs used in reconstructing the Peppers 256×256 image from the Peppers 128×128 image; this was shown to yield a good reconstruction (c) Reconstructed with the 30 features and LAMs used in reconstructing the Left Pentagon 256×256 image from the Left Pentagon 128×128 image. The inappropriate reconstruction is most noticeable in the right portion of the forehead and on portions of the hat (d) Reconstructed with 60 features and LAMs: 30 from the Left Pentagon image used in (c) and 30 from the Peppers image used in (b). We did not have to retrain our system in establishing an appropriate set of bases. We simply “append” the appropriate features and LAMs of the Peppers image to the existing set from the Left Pentagon image to reconstruct an adequate image.
In this manner, we did not have to retrain a system from scratch in order produce an appropriately reconstructed image. Available bases for reconstruction can simply be incorporated into an existing system to produce adequately reconstructed images. This is possible because of the hard partitioning scheme our procedure is implementing.

5.1.2 Nonlinear Associative Memories

In this section we examine the performance of our superresolution methodology with NLAMs rather than with LAMs. Now, instead of using an affine mapping of our image neighborhoods to reconstruct a high resolution neighborhood, a nonlinear mapping by way of an MLP is used. The results reported in this section are for the Lena and Peppers images. They will be directly compared against the results of figs. (5-3) and (5-4). Results using these two images are sufficient to convey how our superresolution methodology typically performed using NLAMs on other images not reported here.

In order to provide a “fair” comparison between the performance of LAMs and NLAMs, the NLAM topologies were selected such that their number of free parameters were approximately equal to that of the LAMs they were being compared against. The NLAM topologies consisted of an MLP with a single hidden layer. It has been shown that this topology is a universal function approximator [Hornik et al., 1989]. The input and output layers of this topology are fixed in size for a specified ROS. It was shown in Chapter 4 that for superresolving an image by a factor of $G_1 \times G_2$ using an ROS of $H_1 \times H_2$, each $H_1 \times H_2$ length vector representing a low resolution neighborhood was mapped to a $(2G_1 - 1)(2G_2 - 1)$ length vector representing a high resolution neighborhood. Therefore, when superresolving by a factor of 2 through each image axis ($G = G_1 = G_2 = 2$) and utilizing an $H = H_1 = H_2 = 3$ ROS, each LAM has
\[ H^2(2G-1)^2 = 9 \cdot 9 = 81 \text{ weight and } (2G-1)^2 = 9 \text{ bias parameters for a total of } 81 + 9 = 90 \text{ free parameters.}\] Performing a similar calculation for \( G = G_1 = G_2 = 2 \) and \( H = H_1 = H_2 = 5 \) ROS, each LAM has 234 free parameters. The NLAM topology tested here for \( G = 2 \) and \( H = 3 \) had 5 units in the single hidden layer. This yields \( H^2 \cdot 5 = 9 \cdot 5 = 45 \text{ weight parameters from the input layer to the hidden layer and } 5 \cdot (2G-1)^2 = 5 \cdot 9 = 45 \text{ weight parameters from the hidden layer to the output layer.} \] There are also \( 5 + (2G-1)^2 = 5 + 9 = 14 \) bias parameters from the hidden and output layers. This yields a total of \( 45 + 45 + 14 = 104 \) free parameters for this NLAM topology. The NLAM for the \( G = 2, H = 5 \) case utilized 7 units in its single hidden layer for a total of 254 free parameters. Noteworthy also is that 10% of the exemplars used in training each NLAM were randomly selected to establish a cross-validation set for the training.

The results of the superresolution performance using NLAMs are illustrated in figs. (5-18) and (5-19). In fig. (5-18), the Lena 128×128 image is superresolved by a factor of 2 in each image axis. Results using two regions of support are reported: 3×3 and 5×5. The training set results are given by the solid lines. Here the Lena 128×128 image was reconstructed using the features and NLAMs obtained from training to reconstruct the Lena 128×128 image to the Lena 256×256 image. The dashed lines illustrate test set performance. Here, the Lena 128×128 image was reconstructed using the features and NLAMs obtained in training to reconstruct the Peppers 128×128 image to the Peppers 256×256 image. The plots in this figure show very similar performance to those of fig. (5-3) where LAMs were used to reconstruct the Lena image.

In fig. (5-19), the Peppers 128×128 image was superresolved by a factor of 2 in each image axis. We report results using the same two regions of support as before. Here
the Peppers 128×128 image was reconstructed using the features and NLAMs obtained from training to reconstruct the Peppers 128×128 image to the Peppers 256×256 image. For the test set, the Peppers 128×128 image was reconstructed using the features and NLAMs obtained in training to reconstruct the Lena 128×128 image to the Lena 256×256 image. The plots in this figure also show very similar performance to those of fig. (5-4) where LAMs were used to reconstruct the Peppers image.

The similarity between the superresolution performance plots using LAMs and NLAMs indicates that there is an approximate affine relationship between samples of our low resolution neighborhoods’ structure and the high resolution neighborhood structure we wish to create. As such, the use of nonlinear associative structures (NLAMs) offers no significant advantage to the superresolution methodology presented in this work at least for the reconstruction by a factor of 4. I expect that higher reconstruction ratios will behave differently.

This means that the approximation to the complex high resolution neighborhood structure space by piecewise linear hyperplanes, one corresponding to each partition (Voronoi cell) of the input space, is sufficiently good. This comparable performance between the use of LAMs and NLAMs was generally observed in the images we tested. The comparable performance in reconstruction can be seen visually by the images illustrated in fig. (5-20). They correspond to the circled points in fig. (5-3) for the LAM case and fig. (5-18) for the NLAM case. Here the 30 features and LAMs (NLAMs respectively) obtained in training to reconstruct the Peppers 256×256 image from a size of 128×128 with an ROS of 3×3 were used for the reconstruction.
Figure 5-18. Training and testing superresolution results using NLAMs for the Lena image considering two different regions of support (ROS). The solid lines correspond to training set results and the dashed lines are test set results. The curves related to the training data result from superresolving the Lena 128×128 image with the systems (features and NLAMs) established to go from the Lena 128×128 image to the Lena 256×256 image. The curves related to testing result from superresolving the Lena 128×128 image with the systems established to go from the Peppers 128×128 image to the Peppers 256×256 image. Similar performance on training and testing set data is observed when using NLAMs or LAMs in our superresolution methodology. The superresolved image corresponding to the circled point ‘o’ is shown in fig. (5-20).
Figure 5-19. Training and testing superresolution results using NLAMs for the Peppers image considering two different regions of support (ROS). The solid lines correspond to training set results and the dashed lines are test set results. The curves related to the training data result from superresolving the Peppers 128×128 image with the systems (features and NLAMs) established to go from the Peppers 128×128 image to the Peppers 256×256 image. The curves related to testing result from superresolving the Peppers 128×128 image with the systems established to go from the Lena 128×128 image to the Lena 256×256 image. Similar performance on training and testing set data is observed when using NLAMs or LAMs in our superresolution methodology.
Figure 5-20. Comparing the reconstruction of Lena using LAMs and NLAMs. The overall PSNR performance of the LAM and NLAM based results were very similar. This was illustrated in figs. (5-3), (5-4), (5-18) and (5-19). The similarity in performance is also seen in the images reconstructed. Here the Lena 128×128 image is reconstructed by a factor of two in each image axis. The 30 features and LAMs (NLAMs respectively) used in the reconstruction were those obtained from training to reconstruct the Peppers 256×256 image from the Peppers 128×128 image with an ROS of 3×3.

5.1.3 Mixture of Experts

In this section we examine the performance of superresolution with the MOE architecture which was discussed in the last chapter. Recall that the input space is soft partitioned in the MOE. This is in contrast to the hard partitioning being performed by our superresolution architecture. The results reported here are for the Lena and Peppers images using two regions of support: 3×3 and 5×5.

In order to provide a “fair” comparison of the results of the MOE with our superresolution methodology, we used a hierarchical gating network composed of linear gating subnetworks. The hierarchical network was a tree-like structure which branched into two at each successive hierarchical level. This provides the most flexible soft partitioning of the input space with linear gating networks operating in a hierarchical
fashion, that is, at the highest level, the input space could be partitioned in two and each of these partitions could be partitioned in two and so on. Due to this partitioning scheme, we present results with 2, 4, 8 and 16 experts. Each of the experts was an affine transformation—exactly as with the LAMs used in our methodology.

Fig. (5-21) illustrates the results for the reconstruction of the Lena 128×128 image with the hierarchical MOE architecture. The training set results consisted of superresolving the Lena 128×128 image by a factor of 2 in each axis with the MOE parameters obtained in training to reconstruct the Lena 256×256 image from the Lena 128×128 image with regions of support 3×3 and 5×5. These results are given by the solid lines in the figure. The test set results consisted of superresolving the Lena 128×128 image by a factor of 2 using the MOE parameters obtained in training to reconstruct the Peppers 256×256 image from the Peppers 128×128 image using the same two regions of support. The training and test image reconstruction results are those denoted by the circled points on the plot for 2, 4, 8 and 16 experts. The lines connecting these points just serve to visually distinguish training from test set results for each ROS considered.

Fig. (5-22) illustrates the results for the reconstruction of the Peppers 128×128 image with the hierarchical MOE architecture. The training set results consisted of superresolving the Peppers 128×128 image by a factor of 2 in each axis with the MOE parameters obtained in training to reconstruct the Peppers 256×256 image from the Peppers 128×128 image with regions of support 3×3 and 5×5. The test set results consisted of superresolving the Peppers 128×128 image by a factor of 2 using the MOE parameters obtained in training to reconstruct the Lena 256×256 image from the Lena 128×128 image using the same two regions of support.
Figure 5-21. Training and testing superresolution results using the hierarchical MOE on the Lena image considering two different regions of support (ROS). Only architectures with 2, 4, 8 and 16 experts were tested. The PSNR of these architectures are denoted by the circled points. The solid lines correspond to training set results and the dashed lines are test set results. The curves related to the training data result from superresolving the Lena 128×128 image with the parameters established to go from the Lena 128×128 image to the Lena 256×256 image. The curves related to testing result from superresolving the Lena 128×128 image with the parameters established to go from the Peppers 128×128 image to the Peppers 256×256 image.
Figure 5-22. Training and testing superresolution results using the hierarchical MOE on the Peppers image considering two different regions of support (ROS). Only architectures with 2, 4, 8 and 16 experts were tested. The PSNR of these architectures are denoted by the circled points. The solid lines correspond to training set results and the dashed lines are test set results. The curves related to the training data result from superresolving the Peppers $128\times128$ image with the parameters established to go from the Peppers $128\times128$ image to the Peppers $256\times256$ image. The curves related to testing result from superresolving the Peppers $128\times128$ image with the parameters established to go from the Lena $128\times128$ image to the Lena $256\times256$ image.
Here, the training and test image reconstruction results are those denoted by the circled points on the plot for 2, 4, 8 and 16 experts. The lines connecting these points also serve to visually distinguish training from test set results for each ROS considered.

The MOE performed better than the common approaches with respect to PSNR. The results of figs. (5-21) and (5-22) do not compare similarly to those from the Lena and Peppers superresolution results when using LAMs and NLAMs. A comparison with our results for the few cases that could be tested with the MOE can only be offered. It is seen in figs. (5-21) and (5-22) that training set performance was better for the 5x5 ROS than with the 3x3 ROS. This is consistent with the performance of our superresolution methodology. However, the increasing PSNR trend of the training set in figs. (5-3) and (5-4) was not observed here. The reason for this is not clear but a few explanations can be conjectured. First, this might be caused by the partitioning afforded by the hierarchical gating network which is not the same as the Voronoi cell-type partitioning of Kohonen’s SOFM. Second, the issue of soft partitioning vs. hard partitioning might be more sensitive than currently thought. Third, training issues regarding the MOE are very important to the performance of the trained network. Let us now provide some observations regarding each of these issues.

When using linear gating subnetworks, the partitions afforded by Kohonen’s SOFM cannot typically be matched by the hierarchical network. As such, direct comparisons of the two partitioning schemes cannot naturally be performed. A direct comparison can be performed when only 2 experts or 2 LAMs are used. Here, the partition is just a single hyperplane through the input space. In the tests conducted, the LAM based method was trained using 2 LAMs and the hyperplane resulting from the two
established Kohonen features was noted. The linear gating network was then initialized to this input space partition and the MOE architecture was trained. The MOE's resulting PSNR was never noted being larger than with the LAM based approach. Again, this might be due to soft vs. hard partitioning considerations and/or to the issue of relative training speed between the gating network and the expert networks.

As for the other curves in figs. (5-21) and (5-22), the training set performance was better than the test set performance (as expected). However the test set performance was better with a 5×5 ROS that with the 3×3 ROS. This is in contrast to the performance of the LAM and NLAM based results shown previously. The reason for this is not understood. A closer scrutiny into the previously discussed issues regarding the MOE methodology might impart the answers.

### 5.2 Feature Extraction

The feature extraction stage is particular to the superresolution methodology presented herein. It does not have any bearing on results obtained using the reconstructors with universal bases. Earlier, we stated that our approach was forming clusters based on the most correlated local structural information contained in images. Examples of the formed clusters as well as the features extracted from some of the images already presented will be shown in the next two sections.

#### 5.2.1 Image Features

The features which result from the SOFM approach have been interpreted as corresponding to the mean of the statistical processes which generated the local structure present in the images studied. The standard images examined are usually very edgy in nature. The features of fig. (5-23) reflect this fact. Here we see the 30 features extracted
from the Peppers 128×128 image of fig. (4-1) using an ROS of 3×3. Notice how “regular” and edgy the features corresponding to the Peppers image are. These features were used in obtaining the LAM test results of figs. (5-6), (5-16), (5-17b), (5-17d), (5-20) and the training results of fig. (5-7). The number below each feature signifies the maximum gray level difference between the largest and smallest value present in each feature. Therefore, the feature with the ‘1’ below it can be considered a constant feature, that is, one that contains practically no structure. This feature corresponds to those image portions that are very smooth. The features have not been scaled here; instead, the constant feature (which is the zero vector) is represented by gray (128 in an 8 bit scale). Positive feature values become lighter and negative feature values are represented by a proportionately darker shade of gray.

Figure 5-23. Features extracted from the Peppers 128×128 image. They were used in superresolving the Lena 128×128 image to a size of 256×256; these results are given in fig. (5-6). Notice the largely edgy nature of these features which is in contrast to the features from texture images. The features have not been scaled here; instead the constant feature is represented by gray (128 in an 8 bit scale). The number below each feature represents the maximum 8 bit gray level difference between the largest and smallest value of that feature.
The features corresponding to the textured images are more difficult to heuristically pinpoint since edges are not the dominating image characteristic. This is evidenced in fig. (5-24). These 30 features were extracted by considering an ROS of $5 \times 5$ from the Asphalt_l $128 \times 128$ image of fig. (5-1). They were used in superresolving the LAM test image of fig. (5-13). Again, the number below each feature signifies the maximum gray level difference between the largest and smallest value present in each feature. These results however have been scaled to better show the features characteristic of the Asphalt_l texture image. Here the smallest feature value takes on the shade of black while the largest feature value is represented by white. Notice that there are no edge features and no constant features either. It would be difficult to assign these features a priori in a heuristic fashion. However, they are readily obtained using the SOFM.

![Figure 5-24. Texture features extracted from the Asphalt_l $128 \times 128$ image. They were used in superresolving the Asphalt_2 $128 \times 128$ image to a size of $256 \times 256$; these results are given in fig. (5-13). Notice the nonedgy nature of these features as we might expect with textures. The features have been scaled between black and white for visual enhancement. The number below each feature represents the maximum 8-bit gray level difference between the largest and smallest value of that feature.](image-url)
5.2.2 Correlated Image Structure

Here, we illustrate the correlated image structure that we mentioned is exploited by our superresolution process. The correlated neighborhoods are those that belong to a particular cluster. Fig. (5-25) serves to illustrate the clusters that were formed for the Lena, Peppers and left Pentagon images of fig. (4-1). Each of these images is 128×128 in size. The clusters were formed using each 3×3 neighborhood contained within these images. Thirty clusters were formed per image and each is represented by a unique gray level. We have evenly distributed these 30 gray levels between shades of black and white (0 and 255). Each cluster was randomly assigned one of these 30 shades of gray.

![Figure 5-25](image_url)

Figure 5-25. Illustration of formed clusters on the training images of Lena, Peppers and the left Pentagon. Neighborhoods that correspond to the same cluster are represented by the same shade of gray. There were 30 clusters formed per image and each image was 128×128 in size. These figures help to visually distinguish the most correlated local information contained in these images – which we use to establish our family of kernels. (a) Lena, (b) Peppers, (c) Left Pentagon.

The feature extraction has, in essence, performed a class labeling of image neighborhoods. This labeling of correlated information is evident from fig. (5-25). Regardless of the shade of gray assigned to each cluster, the visually perceived “information content” of these images is preserved.
Fig. (5-26) shows the 30 clusters formed for the texture images of Asphalt_1 and Grass_1 in fig. (5-1). They were of size 128×128. Each of the 30 clusters is also represented by unique shade of gray. It is much more difficult to visualize the correlated nature of information in these images. In fact, it seems as if there is less interblock correlation between the neighborhoods of these images relative to those of the standard images. Nonetheless, the formed clusters were shown to be useful to the superresolution problem.

![Image clusters](image)

Figure 5-26 Image clusters formed for the Asphalt_1 and Grass_1 128×128 texture images of fig. (5-1). Neighborhoods that correspond to the same cluster are represented by the same shade of gray. There were 30 clusters formed per image. (a) Asphalt_1 clusters, (b) Grass_1 clusters.

At present only the interblock correlations established by the SOFM are being exploited. Let us now note that “higher order” relations between neighborhood clusters might be incorporated into the superresolution procedure. With this statement we are alluding to the use of information relating the topological ordering of image neighborhoods obtained with the SOFM. The use of this additional information might aid in enhancing the superresolution process.
CHAPTER 6
SUPERRESOLUTION OF SYNTHETIC APERTURE RADAR IMAGES

The superresolution of SAR images, like with optical images, requires the establishing of an appropriate set of bases by which the collected radar samples can be weighted for the reconstruction of new image points. The bases are established from the models used to describe the energy backscatter from the objects in the radar illuminated scene and the basis function used in the creation of an image point is dependent on how well the appropriate model resides in the SAR phase history. Presently, the imaging of SAR data is almost exclusively based on the trihedral reflector (point scatterer) model. Many algorithms have been developed to deal with imaging based on this model – this was discussed in Chapter 3. However, not every object in a radar illuminated scene can be approximately modeled as a trihedral reflector. Because of this, an ideally formed SAR image would include models that account for the electromagnetic reflection of other objects in the imaged scene. The methodology presented in this chapter allows for this type of imaging. Hence, imaging based on the point scatterer model is a special case of the imaging paradigm presented here. Here, we are concerned with a methodology that can incorporate various models into a SAR superresolution procedure rather than focusing on approaches to imaging with any one particular model.

There are two main issues regarding the superresolution of SAR images with multiple models:
• selecting the appropriate model by which to form an image point

• determining the optimal projection of the phase history samples once the model to image with is known

These issues will be discussed in this chapter. We will first discuss the use of the superresolution methodology for the automatic target detection (ATD) problem. This is followed by a description of the architecture for implementing the methodology. Also included are details regarding the superresolution process for SAR imagery. Finally, experimental results are presented using MSTAR slicey as well as target data [Wright Laboratories, 1997]. These results are analyzed and compared with some common methods used in the imaging of SAR data and the performance of these methods for the ATD problem is assessed.

6.1 Imaging with Multiple Models

Multiple objects of varying composition typically comprise a radar illuminated scene. Examples of such objects are armored vehicles (and their associated component parts), trees, open field, buildings, etc. If an accurate model relating each of these objects to their phase history representation was available, then the imaging process would ideally be based on the model best describing the local neighborhood to be imaged. The need for more accurate imaging of SAR data is motivated by, among other things, the ATD/R problem. In particular, we are concerned herein with the automatic target detection (ATD) problem.

The ability to “preserve” clutter without destroying or distorting it is known to be important to the ATD and semi-automatic target detection (SATD) problem. ATD and SATD performance based on a clutter preserving algorithm [Benitz, 1997] has proven
beneficial to this problem [Novak, 1997]. This algorithm, however, does not directly account for the clutter, that is, it images based solely on the trihedral reflector model. As such, the introduction of models other than the point scatterer model into the imaging process would allow for better control of the imaging of clutter. This could prove beneficial to the ATD problem and will be addressed by the superresolution methodology of this chapter.

6.2 Architecture Specifics

The superresolution methodology for SAR images is similar in spirit to the optical image superresolution methodology. Both methods seek to use the most appropriate model (from the ones available) for the imaging of a particular point. There are two steps in the methodology:

- comparison of global phase history data with established image models
- superresolution based on the appropriate model

The first step chooses the most appropriate model while the second step estimates the image by weighting the phase history samples based on one of the available models for the point being imaged. Because of the nature in which SAR data is imaged, the basis functions employed in weighting the phase history samples are global in extent. This is because SAR images are viewed in the frequency domain — so the global weighting of our samples enables us to image with as much resolution as possible. These global basis functions are in sharp contrast to the local weighting encountered for optical images. A block diagram of the approach is pictured in fig. (6-1). As mentioned earlier, the first step is to compare the phase history data against available models for the point being imaged.
Figure 6-1. The superresolution architecture for SAR images.
When the appropriate model is determined, the basis for which to project the phase history samples is determined by the model estimation blocks in the figure. Each of the model estimators in fig. (6-1) can be interpreted as an expert specialized in a type of imagery. Note that the modules (model estimators) in fig. (6-1) are no longer given by local kernels as they were when superresolving optical images.

Each of these blocks has purposely been denoted as a model estimator due to the manner in which points are being imaged. If the imaging was solely based on the trihedral reflector model (as is typically done), then the imaging of points is the spectral estimation problem owing to the Dirac delta nature of the ideal trihedral reflector model in the frequency (image) domain.

In SAR imagery there is no training associated with establishing the optimal transformation of phase history samples to high resolution image points. Rather, this is based on the degree to which a model agrees with the phase history data. Note that the adaptive filtering spectral estimation schemes mentioned in the review of SAR superresolution imaging in Chapter 3 can be incorporated into our approach for the imaging of a point based on the trihedral reflector model.

6.3 Models Used

Two models have been implemented for the ATD results reported herein: the point scatterer model and a background clutter model. The point scatterer model assumes that the scene illuminated by the radar is composed of trihedral reflectors. All high resolution spectral estimators use this model. The point scatterer model phase history is given by

$$m_q = \text{vec}(M_q)$$ where the matrix $M_q = m_q[n_1, n_2] = \exp[j(\omega_{q_1} n_1 + \omega_{q_2} n_2)]$ for
\[ n_i = 0, \ldots, N_i - 1 \quad \text{and} \quad \omega_{q_i} = \frac{2\pi}{Q_i} q_i \quad \text{for} \quad q_i = 0, \ldots, Q_i - 1, \quad (i = 1, 2). \]

\( q = [q_1, q_2] \) here denotes the 2D frequency grid location at which points are imaged. \( N_i \) denotes the size of the matrix corresponding to the phase history samples and \( Q_i \) is the size of the superresolved image. In general, \( Q_i \geq N_i \) for \( i = 1, 2 \) for superresolution. By solely projecting the phase history vector onto the \( \mathbf{m}_q \) vectors, the equivalent of evaluating the DSFT on a discrete frequency grid à la eqn. (3-10) results.

Our clutter model seeks to account for the background interference present in SAR images. We wish to do this in a manner that does not eliminate its effects. This is because the background in a SAR image is used in the automatic detection of targets by computer [Kreithen et al., 1993]. The background clutter here is modeled as zero-mean complex Gaussian noise in the phase history. This corresponds well with the phase history of background chips from the MSTAR data set which we worked with. A 100 bin histogram of the real and imaginary parts of the phase history corresponding to background clutter from 90 different MSTAR target chips is illustrated in fig. (6-2). Note that the histograms resemble identically distributed Gaussian probability density functions.

A realization of the clutter model will be denoted by a vector \( \mathbf{g} \) where the samples are drawn from the Gaussian distribution described by \( \mathbf{g} \sim N(0, \sigma^2_r / 2) + jN(0, \sigma^2_i / 2) \). The variance of the real and imaginary portions of the complex Gaussian distribution are denoted by \( \sigma^2_r / 2 \) and \( \sigma^2_i / 2 \), respectively. When \( \sigma^2_r = \sigma^2_i = \sigma^2 \), this noise model has a power of \( \sigma^2 \).
Figure 6-2. Histogram of real and imaginary parts of phase history from background clutter from 90 MSTAR target chips. The zero-mean Gaussian model we use for background clutter is supported with this data. (a) Real part, (b) Imaginary part.
6.4 The Methodology

The superresolution approach utilizing multiple models as presented here is based on the MV method. This method was shown to produce the optimal set of basis functions such that the output power of the image was minimized subject to how well the image matched the assumed model. In the subsections that follow, we will describe how we determine the appropriate model to use during the superresolution process and describe how this process was implemented.

6.4.1 The Imaging Criteria

The superresolution approach utilizing multiple models as presented here is based on minimum variance (MV) experts. The estimate of an imaged point \( q \) is dependent on how well the phase history matches one of the available models. Specifically, for phase history \( x \), we determine the projection \( w \) which minimizes the output power of the estimate \( s_q \) based on its match to the model \( m_q \). This is written

\[
s_q = \min_w E[x^Hw^2] = \min_w w^HRw \quad \text{subject to} \quad w^Hm_q = 1 \quad (6-1)
\]

Note here that \( x \) is a vector corresponding to the \( N_1 \times N_2 \) matrix \( X \) of phase history samples, that is, \( x = \text{vec}(X) \) where \( \text{vec}(\cdot) \) is the operation that stacks the columns of a matrix one on top of the other to form a vector. This correspondence also holds for model vector \( m_q \) and \( q = [q_1, q_2] \) denotes the image location in the formed \( Q_1 \times Q_2 \) SAR image. \( R \) is the autocorrelation matrix of the phase history samples. The solution to the problem of eqn. (6-1) is well-known and is readily solved with the use of Lagrange multipliers [Stoica and Moses, 1997]. The model estimate at image point \( q \) is given by

\[
s_q = \frac{1}{m_q^HR^{-1}m_q} \quad (6-2)
\]
Notice that when our model \( m_i \) is a 2D complex exponential, then the model estimation problem is equivalent to spectral estimation.

### 6.4.2 Selecting Between Models

Each model estimator of fig. (6-1) corresponds to one of \( C \) models \( m_1(q), \ldots, m_C(q) \). Note that each model is dependent on the point \( q \) being imaged. The phase histories of these models must be used to determine which model, from the given ones, is most appropriate for the formation of an image point. Everything about the model is known except for its amplitude. Therefore we must estimate each model’s amplitude at the point being imaged. Having done this, a comparison of the models’ phase history (along with estimated amplitude) can be performed against the actual phase history data in order to select the best matched model.

Let us define the cost of mismatch between the given phase history samples \( x \) and the phase history samples of model vector \( m_{c(q)} \) as

\[
J_{c(q)}(a_{c(q)}) = (x - a_{c(q)}m_{c(q)})^H(x - a_{c(q)}m_{c(q)})
\]  

(6-3)

Notice that this cost is a function of the unknown model amplitude at the point being imaged \( a_{c(q)} \). The amplitude we are solving for must be a real number. If it is complex, then the estimated amplitude would contain the phase shift information for which that particular model best fits in the phase history – not necessarily at the point we desire to image. The amplitude for which model \( m_{c(q)} \) best matches the phase history is said to minimize cost \( J_{c(q)} \). This is easily obtained through the following equation:
\[
\frac{\partial J_{c(q)}}{\partial a_{c(q)}} = \frac{\partial}{\partial a_{c(q)}} \left( x^H x - a_{c(q)} x^H m_{c(q)} - a_{c(q)} m_{c(q)}^H x + a_{c(q)}^2 m_{c(q)}^H m_{c(q)} \right) \\
= -x^H m_{c(q)} - m_{c(q)}^H x + 2a_{c(q)} m_{c(q)}^H m_{c(q)} \\
= -2 \Re[x^H m_{c(q)}] + 2a_{c(q)} m_{c(q)}^H m_{c(q)}. \quad (6-4)
\]

Setting eqn. (6-4) to 0 and solving for \(a_{c(q)}\) gives the amplitude for the phase history of model \(m_{c(q)}\) which yields the smallest mismatch in terms of squared error. This amplitude is

\[
a_{c(q)} = \frac{\Re[x^H m_{c(q)}]}{m_{c(q)}^H m_{c(q)}} \quad (6-5)
\]

Substituting eqn. (6-5) into eqn. (6-3) yields the cost \(J_{c(q)}\) which is used in determining the best matched model to the phase history at the point being imaged. This is

\[
J_{c(q)} = (x - \frac{\Re[x^H m_{c(q)}]}{m_{c(q)}^H m_{c(q)}} m_{c(q)})^H (x - \frac{\Re[x^H m_{c(q)}]}{m_{c(q)}^H m_{c(q)}} m_{c(q)}) \quad (6-6)
\]

One way to introduce computational savings is to normalize all model vectors. In this way, \(m_{c(q)}^H m_{c(q)} = 1\) and eqn. (6-6) somewhat simplifies. Finally, the function which describes the model most appropriate for the imaging of point \(q\) on the image grid given the phase history \(x\) is

\[
\gamma_q(x) = \arg \min_c \left( J_{c(q)} \right), \ c = 1, 2, ..., C \quad (6-7)
\]

### 6.4.3 The Superresolution Process

The process of superresolving an image with the methodology presented here is very simple. The optimal estimator of eqn. (6-2) is the basis for our imaging. First, we determine, from eqn. (6-7), the appropriate model to use in forming each image point. Here we will denote the point scatterer model for frequency grid location \(q\) as \(m_q\) and a
realization of the complex Gaussian noise model as \( g \). Once the model is determined, the appropriate model is utilized in eqn. (6-2) to yield a superresolved point.

For a complex 2D phase history which is \( N_1 \times N_2 \) samples in size, we create an image which is \( Q_1 \times Q_2 \) samples in size. In this manner, we are imaging at more \( (Q_i > N_i) \) frequency locations than that necessary for the perfect reconstruction of the phase history. Each point \( \mathbf{q} = [q_1, q_2] \) in the superresolved SAR image corresponds to frequency coordinates \( (\omega_{q_1}, \omega_{q_2}) = \left( \frac{2\pi}{Q_1} q_1, \frac{2\pi}{Q_2} q_2 \right) \) where \( q_i = 0,1,\ldots,Q_i, (i = 1,2) \).

Let the point scatterer model \( \mathbf{m}_q \) correspond to model 1 and the noise vector \( \mathbf{g} \) correspond to model 2. The model estimate, or imaged point, at location \( \mathbf{q} \) will be denoted \( s_q \). It is given by

\[
 s_q = \frac{1}{\mathbf{v}^T \mathbf{R}^{-1} \mathbf{v}}
\]

where

\[
 \mathbf{v} = \begin{cases} 
 \mathbf{m}_q & \gamma_\mathbf{q} (\mathbf{x}) = 1 \\
 \mathbf{g} & \gamma_\mathbf{q} (\mathbf{x}) = 2
 \end{cases}
\]

The superresolved image is thus given by the matrix \( \mathbf{S} = [s_q] = s[q_1, q_2] \) which is \( Q_1 \times Q_2 \) samples is size. The reason we have chosen to project onto a randomly sampled noise vector \( \mathbf{g} \) is to maintain some variability in the model estimates of background clutter. This preserves some of the clutter structure in the image but still reduces the clutter variance which is useful for the ATD problem. In experimental tests it has been found useful, for each point to image, to generate a few noise realizations \( \mathbf{g} \) (usually 10) and then determine the average cost of these realizations against the phase history via eqn. (6-6). This is done to aid in determining which model to form the image point with. Other implementation
details to note are that the superresolved imagery are formed using a mosaicking of 50% overlapping subblocks. The autocorrelation matrix $R$ was estimated using a forward-backward averaging of subwindows within the overlapping subblocks' phase history as described by eqn. (3-16).

**6.5 Experimental Results**

The results reported in this section compare the performance of four spectral estimation methods. The complex phase histories of the data reported in this section will be processed with (1) the unwindowed (or rectangular windowed) Fourier transform, (2) the Taylor windowed (n=5, -35dB) Fourier transform, (3) the MV (Capon) method and (4) the proposed superresolution architecture; we will refer to this here as the dual model (DM) approach. We now note that SAR imaging via the Fourier transform is by far the most common processing method. Imaging in this manner has been included to serve as a comparative baseline to the other methods.

**6.5.1 MSTAR Slicy Data**

Prior to reporting on the superresolution results of MSTAR slicy data, the four methods mentioned were tested on a synthetically generated phase history. Fig. (6-3) illustrates the results of processing a synthetically generated phase history with the above mentioned estimators. The phase history is composed of three complex plane waves embedded in complex zero-mean Gaussian noise. The scatterers' signal power was 0 dB, -3 dB and -6 dB respectively. The complex noise power was 0 dB. The phase history was $25 \times 25$ samples in size and the images pictured in the figure were imaged to $50 \times 50$ samples. The three scatterers can easily be distinguished in this figure. The unwindowed FT retains the sidelobes corresponding to the scatterer with the most power.
The side lobes of the other scatterers are well below the 0 dB noise floor. The Taylor windowed FT has suppressed the sidelobes of all scatterers at the expense of a wider main lobe width. The MV and DM approaches have more accurately detected the scatterers relative to the FT approaches. The background (complex Gaussian noise) appears very different for each of these methods. The Taylor windowed FT and the MV approaches have distorted the variance in the background considerably. The unwindowed FT provides a typical SAR image background while the DM approach has reduced the background variance while correctly detecting the present scatterers.

Figure 6-3. Images of a synthesized phase history consisting of three point scatters embedded in complex zero-mean Gaussian noise. The synthesized phase history size was 25×25 samples and the illustrated SAR images are 50×50 samples in size. (a) Unwindowed FT, (b) Taylor windowed FT, (c) MV method and (d) DM method.

Figs. (6-4) and (6-5) illustrate the performance of these superresolution techniques on MSTAR slicy data. This is real SAR data for which certain geometric structures are illuminated by radar. These structures, which typically illicit one or more large energy returns, are used for testing several aspects of the collected radar data. The phase history collected for each of these structures was 40×40 samples. They were imaged to 80×80 samples. In fig. (6-4), there are three visible prominent scattering returns.
6.5.2 MSTAR Target Data

Figs. (6-6) and (6-7) illustrate the performance of these superresolution techniques on MSTAR target data. The targets in these figures are T-72 tanks at different poses.
relative to the radar. The radar data collected was at a depression angle of 17 degrees. The phase history for each target collected was 100×100 samples. They were each imaged to 200×200 samples. The observations made regarding the ability to resolve bright point scatterers for the slicy data also hold for target data too.

In these figures, we can notice that the background with the dual model approach has much smaller variance than the background which results from the other imaging methods. This is important when considering issues such as target detection in SAR imagery.

With the DM approach, every point to be imaged is either “classified” as a bright point scatterer or not a bright point scatterer. The nonpoint scatterer class in this method is modeled as background clutter (complex Gaussian noise). As such, points on targets which do not resemble the bright point scatterer model are imaged as background clutter and hence lead to the “spotty” target produced by the DM approach. From the point of view of visually pleasing images, this does not result in a natural representation of the imaged target. Obviously, the structure of a tank is much more complex than to be adequately represented by two models, let alone by a single point scattering model. The proper inclusion of more models will solve this problem.

6.5.3 Detection Performance

From the point of view of CFAR target detection, the DM approach outperforms the other approaches. To detect a target, a simple test comparing the imaged points on target to simple statistics of the region around the targets’ perimeter is performed. If the test, known as the CFAR statistic, is greater than a preset threshold, then the image point about which the test is being conducted is said to belong to a target.
Figure 6-6. Images of MSTAR target data (hb03931.017). The phase history size was 100×100 samples and the illustrated SAR images are 200×200 samples in size. (a) Unwindowed FT, (b) Taylor windowed FT, (c) MV method and (d) DM method.
Figure 6-7. Images of MSTAR target data (hb03940.017). The phase history size was 100×100 samples and the illustrated SAR images are 200×200 samples in size. (a) Unwindowed FT, (b) Taylor windowed FT, (c) MV method and (d) DM method.
The CFAR output for pixel \( x \) under test is defined as \( (x-m)/\sigma \) where \( m \) is the sample mean and \( \sigma \) is the standard deviation of pixels in the perimeter region surrounding the target. Target detection performance is typically evaluated by analysis of a receiver-operating-characteristic (ROC) curve. Since ground truthed data containing targets embedded in clutter is not available we must perform an analysis based on the data that was available: clutter only data and target only data. Thus we can establish the average gain in performance of detection in targets vs. clutter for each superresolution method and then compare these with each other.

Fig. (6-8) illustrates the average detection performance on targets vs. clutter for the four methods we have been discussing. There were 90 target and 90 clutter chips that were tested. The 90 targets were T-72 tanks at randomly selected pose angles at 17 degrees of depression. The 90 clutter chips were selected randomly from nonoverlapping sections of a large natural clutter scene containing many trees and vegetation interspersed with sections of open ground along with their associated shadows (hb06178). All chips had a phase history of 100×100 samples. They were imaged to 200×200 samples. The targets in each chip were within the center 67×67 portion of the superresolved image. The guardband was a square of 134 pixels on a side with a 1 pixel width. For each target, we report the maximum CFAR output for pixels under test (these are the center 67×67 portion of all of the chips tested). This is because the CFAR tests that are run classify clusters of points as target or nontarget based on the maximum value in the cluster which exceeds the preset threshold.
Figure 6-8. Illustration of the detection performance resulting from the methods tested. The DM approach has increased the ability to detect targets well over the other methods. It also produces higher returns on clutter data but the relative increase between target and clutter returns is largest for the DM approach. Note that here the maximum CFAR output for each target and clutter chip considered has been sorted according to the DM results.
The detection performance for each target and clutter chip is reported in fig. (6-8). The results have been sorted with respect to the DM results for target as well as clutter chips. The performance of the superresolution methods remain consistent throughout these tests.

It is evident that, in both target and clutter, the DM approach consistently produces a higher CFAR output than the MV method and that this method produces a higher result relative to the FT approaches. There is little difference between performance with the unwindowed FT and the Taylor windowed FT. The increase in output from the DM approach comes primarily from the lower variance (standard deviation) which results from the superresolved background as well as the higher spectral peaks resulting from the MV method when a point was imaged as a point scatterer. The average performance of these methods is compared in tables 6-1 through 6-3. The numbers tabulated here come from analysis of the results shown in fig. (6-8).

Table 6-1 reports on the average performance of these algorithms relative to each other on the 90 targets. Each entry in this table, the average target gain (ATG), was computed by

\[
\text{ATG}(a, b) = \frac{1}{90} \sum_{n=1}^{90} \frac{\text{maximum CFAR output on target } n \text{ with approach } a}{\text{maximum CFAR output on target } n \text{ with approach } b}
\]

Table 6-2 reports on the average performance of the algorithms relative to each other on the 90 clutter chips. Each entry in this table, the average clutter gain (ACG), was computed by

\[
\text{ACG}(a, b) = \frac{1}{90} \sum_{n=1}^{90} \frac{\text{maximum CFAR output on clutter chip } n \text{ with approach } a}{\text{maximum CFAR output on clutter chip } n \text{ with approach } b}
\]
Table 6-3 reports on the average performance of the algorithms relative to each other based on how they performed on the 90 targets relative to the 90 clutter chips. Each entry in this table, the average target to clutter gain (ATCG), was computed by

$$ATCG(a, b) = \frac{ATG(a, b)}{ACG(a, b)}$$

Table 6-1. Average relative gain over 90 imaged MSTAR target chips using several methods.

<table>
<thead>
<tr>
<th>b \ a</th>
<th>UN</th>
<th>TAY</th>
<th>MV</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN</td>
<td>1</td>
<td>0.97</td>
<td>1.53</td>
<td>1.90</td>
</tr>
<tr>
<td>TAY</td>
<td>x</td>
<td>1</td>
<td>1.58</td>
<td>1.95</td>
</tr>
<tr>
<td>MV</td>
<td>x</td>
<td>x</td>
<td>1</td>
<td>1.24</td>
</tr>
<tr>
<td>DM</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6-2. Average relative gain over 90 imaged clutter chips using several methods.

<table>
<thead>
<tr>
<th>b \ a</th>
<th>UN</th>
<th>TAY</th>
<th>MV</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN</td>
<td>1</td>
<td>0.99</td>
<td>1.42</td>
<td>1.59</td>
</tr>
<tr>
<td>TAY</td>
<td>x</td>
<td>1</td>
<td>1.42</td>
<td>1.59</td>
</tr>
<tr>
<td>MV</td>
<td>x</td>
<td>x</td>
<td>1</td>
<td>1.12</td>
</tr>
<tr>
<td>DM</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6-3. Average relative target to clutter gain using several methods.

<table>
<thead>
<tr>
<th>b \ a</th>
<th>UN</th>
<th>TAY</th>
<th>MV</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN</td>
<td>1</td>
<td>0.98</td>
<td>1.08</td>
<td>1.20</td>
</tr>
<tr>
<td>TAY</td>
<td>x</td>
<td>1</td>
<td>1.11</td>
<td>1.23</td>
</tr>
<tr>
<td>MV</td>
<td>x</td>
<td>x</td>
<td>1</td>
<td>1.11</td>
</tr>
<tr>
<td>DM</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>1</td>
</tr>
</tbody>
</table>

Let us now note that the x’s in the tables are values that are not significant since the relative performance of imaging method a to b is nearly reciprocal to that of method b to a. We can see from table 6-3 that the DM method performed equal or better in ATCG
when compared with the other methods. This serves as an indicator as to how these methods will perform in an ROC test.
CHAPTER 7

CONCLUSIONS AND FUTURE WORK

Here we present a recapitulation of the research performed and the contributions made towards the solution of the problems addressed. The thrust of this work was in establishing a unified methodology for the superresolution of optical and SAR images. We saw that the superresolution problem was essentially the problem of perfect reconstruction defined in the spatial domain for optical images and in the frequency domain for SAR images. The superresolution problem had not been framed in this fashion previously. The theoretically established sets of bases for perfect reconstruction in the optical and SAR domains was given by the sinc and Fourier bases, respectively. These bases were established in terms of the sets of linear projections that spanned the signal space. Perfect reconstruction was only achievable when an infinite amount of collected data was specified. In practice, we are limited to working with a finite amount of data so the reconstruction (or resolution) capabilities of utilizing these established sets of basis functions was compromised.

To overcome these limitations, it was shown that a priori information relating the collected samples to the signal to reconstruct could be used to establish a set of bases for which a better image reconstruction could be achieved. This reconstruction procedure was embodied by a modular architecture where each module was responsible for specifying the basis function for which the collected set of samples could be projected in
order to form an image. We presented several sets of accepted bases used in the reconstruction of optical and SAR images in Chapter 3. The results of reconstruction via the superresolution methodology presented here were compared to the accepted reconstruction methods and were shown to perform better than the standard approaches. This was done in chapter 5 for optical images and in chapter 6 for SAR images.

Establishing the set of bases was key to the success of the superresolution procedure. The bases had to be "broad" enough to be useful on a class of images of interest. For optical images, we were able to establish a set of bases in an unsupervised manner by exploiting the correlated structure of local information in images as well as the interdependencies exhibited by these neighborhoods across scales. In this manner, the sets of bases would come from optimally associating the correlated homologous neighborhoods of a pair of counterpart high and low resolution images. Also, we saw that the most effective basis functions had to be local in space. Typically, for very structured images, test set reconstruction performance was observed to be best for highly localized basis functions in space. For optical images with more "regularity" such as textures, the best performing set of bases were not as locally concentrated in space.

In SAR images, we saw that the basis functions had to be global in extent in the phase history domain. This was necessary to afford maximal resolution in the SAR image which is viewed in the frequency domain. The main contributions in this regard were that of allowing for multiple models in the construction of a SAR image via the architecture for superresolution of this work. Two models were used in the imaging of SAR phase history data: the point scatterer model and a background clutter model which was complex zero-mean Gaussian noise. The use of imaging with these models in a nonparametric fashion
led to ease of implementation and increased automatic target detection performance over the currently established methods based on direct use of the Fourier transform.

The importance of this research can be summarized as follows:

1. The approach herein unified the superresolution problem for two image domains which are very different in nature. Hence, further study into the advances of one of these domains could prove applicable to imaging in the other domain.

2. The approach exploited the correlated structure of optical images in a new fashion. Only very recently has the correlated information of images been suggested for image compression. Here, this information as well as the interdependencies among homologous neighborhoods across scales was utilized for image superresolution.

3. The approach allowed for the imaging of SAR data based on multiple models. Until now, no nonparametric SAR imaging approach that the author is aware of has allowed for this. By establishing the proper projections which decrease clutter variance, automatic target detection performance is enhanced.

Issues that need further addressing constitute the future work suggested here. The issues suggested for further analysis are:

1. Develop a stronger mathematical link to sampling theory. Among the important issues here are the ability to show that the established bases default to the sinc bases when all of the conditions regarding the perfect reconstruction of the Shannon sampling theory are satisfied. This also includes the issue of establishing bounds for the accuracy of the reconstruction given only partial knowledge of the problem.
2. Establish a natural set of bases for images. This issue is quite broad but is a direct consequence of the work introduced in this study. Here we are speaking towards the manner in which the established set of bases are developed in an autoassociative setting. Methods such as PCA and independent component analysis (ICA) utilize an autoassociative approach to the establishing of bases for the problem of signal representation. Here the autoassociation occurs with image samples across scales. As such, the resolution issue dealt with here is the reconstruction problem rather than the representation problem usually addressed by PCA and ICA. The implications of such an established set of bases include increased compression gains and improved quality images in digital video systems such as video conferencing and HDTV.

3. Include “higher order” local interdependent information into the optical superresolution process. This entails further exploring relationships among the local neighborhoods of images that have been used for establishing our reconstruction bases. An example of this is to consider the effect of the topological ordering obtained from the SOFM approach that is being used to partition the input space. Relationships among the resulting ordered feature vectors could serve to more appropriately constrain the input space partitioning and/or the obtained reconstructed neighborhoods.

4. Incorporate other models into the SAR superresolution procedure. The results of this work with the two models utilized indicate the potential of the approach for tackling more difficult problems such as recognition and classification if multiple models are properly incorporated. The main issues here are the establishing of appropriate models
representative of the objects in the imaged environment and their proper inclusion in the imaging methodology.

5. Test further the SAR superresolution methodology. In particular, the approach should be tested in an end-to-end detection system with ground truthed data in order to properly quantify this method’s performance. Other related issues are testing and developing other model estimation procedures for the construction of an image point. As an example, it was pointed out that several spectral estimation methods have been developed for imaging based on the point scatterer model. These tests should be performed in the aforementioned end-to-end system.

6. Examine the potential applications of this methodology further. Throughout this study, applications of this methodology have been suggested but only the direct reconstruction performance relating to optical images and the automatic target detection performance of SAR images has been discussed. Foreseeable applications in the superresolution of optical images are in the postprocessing of decompressed images, better quality and higher resolution images in video conferencing and HDTV systems, restoration of old movies with low resolution or blurred appearance and increased graphics resolution for printers and similar display devices. Applications for SAR images include the increased ability to discriminate between target and nontarget structures and also classify targets and clutter. It also includes the increased visual appearance of SAR images to more closely resemble their optical image representation. This is useful in the SATD/R problem where human operators help monitor (in real time) the automatic discrimination and classification decisions made by a computer.
REFERENCES


Webster’s New Universal Unabridged Dictionary (1989), Barnes and Noble Bookstores, Dilithium Press.


BIOGRAPHICAL SKETCH

Frank Candocia was born on June 14, 1967. He obtained the Bachelor and Master of Science both in electrical engineering from Florida International University in Miami in 1990 and 1993, respectively. His master’s thesis was on passive depth extraction from two-dimensional binocular stereo images. He began his doctoral studies in the Department of Electrical and Computer Engineering in 1994 at the University of Florida, Gainesville, where he was the recipient of a UF Graduate Minority fellowship and a Motorola fellowship. His research work was supervised by Dr. Jose C. Principe and focused on the superresolution of images. During the summer of 1997 he went to MIT Lincoln Laboratory to continue his research and collaborate with Dr. Leslie Novak. His main areas of research interest are in image processing, artificial neural systems, digital signal processing and computer vision.
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jose C. Principe, Chairman
Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John M. M. Anderson
Associate Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

A. Antonio Arroyo
Associate Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John G. Harris
Assistant Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Janice Cox Honeyman
Associate Professor of Computer and Information Science and Engineering
This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 1998

__________________________
Winfred M. Phillips
Dean, College of Engineering

__________________________
Karen A. Holbrook
Dean, Graduate School