Title
Modelling and MMSE reconstruction solutions for image and video super-resolution

Permalink
https://escholarship.org/uc/item/67n4789v

Author
Prendergast, Ryan Strong

Publication Date
2008-01-01

Peer reviewed|Thesis/dissertation
Modelling and MMSE Reconstruction Solutions for Image and Video Super-Resolution

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Electrical Engineering

by

Ryan Strong Prendergast

Committee in charge:

Professor Truong Q. Nguyen, Chair
Professor Pamela C. Cosman
Professor William S. Hodgkiss
Professor Dimitris N. Politis
Professor Nuno Vasconcelos

2008
Copyright
Ryan Strong Prendergast, 2008
All rights reserved.
The dissertation of Ryan Strong Prendergast is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2008
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature Page</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>xi</td>
</tr>
<tr>
<td>Vita and Publications</td>
<td>xiii</td>
</tr>
<tr>
<td>Abstract</td>
<td>xiii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Super-Resolution Application Interest</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Comments on Style, Notation, and Nomenclature in this Dissertation</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Organization and Intent</td>
<td>5</td>
</tr>
<tr>
<td>2 Modelling the Image Degradation Process</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>7</td>
</tr>
<tr>
<td>2.2 Degradation Modelling</td>
<td>8</td>
</tr>
<tr>
<td>2.3 Standard Models</td>
<td>10</td>
</tr>
<tr>
<td>2.3.1 Distortion</td>
<td>10</td>
</tr>
<tr>
<td>2.3.2 Noise</td>
<td>15</td>
</tr>
<tr>
<td>2.3.3 Sampling</td>
<td>18</td>
</tr>
<tr>
<td>2.4 Classic Image Restoration</td>
<td>28</td>
</tr>
<tr>
<td>2.4.1 LMMSE (Wiener) Restoration</td>
<td>28</td>
</tr>
<tr>
<td>2.4.2 Regularization</td>
<td>31</td>
</tr>
<tr>
<td>2.4.3 Richardson-Lucy</td>
<td>32</td>
</tr>
<tr>
<td>2.5 Registration</td>
<td>33</td>
</tr>
<tr>
<td>2.5.1 Global Registration</td>
<td>35</td>
</tr>
<tr>
<td>2.5.2 Local Registration</td>
<td>37</td>
</tr>
<tr>
<td>3 Content Modelling</td>
<td>40</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>40</td>
</tr>
<tr>
<td>3.2 Inherent Models</td>
<td>41</td>
</tr>
<tr>
<td>3.3 Specified Models</td>
<td>43</td>
</tr>
<tr>
<td>3.3.1 Global Correlation and PSD Modelling</td>
<td>46</td>
</tr>
<tr>
<td>3.3.2 Local Correlation Modelling</td>
<td>80</td>
</tr>
<tr>
<td>3.4 Acknowledgement</td>
<td>88</td>
</tr>
</tbody>
</table>
4 Still Image Super-Resolution Reconstruction
4.1 Introduction
4.2 Problem History
4.3 Fourier Based Reconstruction
   4.3.1 MMSE Solution
   4.3.2 Fundamental Limitations of Super-Resolution in the Frequency Domain
4.4 Alternative Approaches
   4.4.1 Iterative Interpolation and Deblurring via Regularization
   4.4.2 Direct Spatial Domain Interpolation
4.5 Example Reconstruction Results
   4.5.1 First Test
   4.5.2 Second Test
   4.5.3 Third Test
4.6 Acknowledgement

5 Video Sequence Super-Resolution Reconstruction
5.1 Introduction
5.2 Prior Approaches
   5.2.1 State-Space Implementation
   5.2.2 Adaptive Wiener Filter Implementation
5.3 Proposed Video Super-Resolution Enhancement Algorithm
   5.3.1 Design Goals
   5.3.2 Principal Algorithm
   5.3.3 Computational Analysis
5.4 Simulation Results
5.5 Acknowledgement

6 Summary of Contributions
6.1 Contributions
6.2 Importance of the Contributions and Extensions

A MMSE Solution to Generalized Undersampling Problem
A.1 Problem Setup
A.2 Synthesis Filter Optimization
A.3 Extensions and Additional Considerations
   A.3.1 Optimality of Decimated Output Sequences
   A.3.2 Additive Noise and Interference
   A.3.3 Numerical Considerations
A.4 Acknowledgement

Bibliography
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Simple linear image degradation system.</td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>Purely discrete sample rate conversion system.</td>
<td>23</td>
</tr>
<tr>
<td>2.3</td>
<td>Bayer color filter mosaicing pattern, representing the color of light admitted through the filter to individual CCD elements. Periodic repetition of this pattern leads to regular sampling in all three color spaces, where the green “luminance” color space receives twice the number of samples as either of the red or blue “chrominance” color spaces.</td>
<td>27</td>
</tr>
<tr>
<td>2.4</td>
<td>LMMSE degradation and restoration block diagram. Signals and filters are labeled for the spatial domain, equivalent frequency domain version is applicable.</td>
<td>29</td>
</tr>
<tr>
<td>2.5</td>
<td>Example effects of registration error.</td>
<td>34</td>
</tr>
<tr>
<td>3.1</td>
<td>Example Gibbs phenomena.</td>
<td>43</td>
</tr>
<tr>
<td>3.2</td>
<td>Circular shift equivalent images.</td>
<td>50</td>
</tr>
<tr>
<td>3.3</td>
<td>Example axis artifacts. Two windowed sinusoid functions of different frequencies (a,d) dictate the horizontal amplitudes in images (b,e) leading to very different periodogram estimated spectra (c,f).</td>
<td>51</td>
</tr>
<tr>
<td>3.4</td>
<td>Example axis artifact reduction. Image from Fig. 3.3(e) is modified via Hamming window (a) and boundary blurring (c) to produce respective periodogram estimates in (b) and (d).</td>
<td>54</td>
</tr>
<tr>
<td>3.5</td>
<td>Example images for discussion. Pentagon image obtained from USC database [69].</td>
<td>58</td>
</tr>
<tr>
<td>3.6</td>
<td>Spectra of images in Fig. 3.5.</td>
<td>58</td>
</tr>
<tr>
<td>3.7</td>
<td>Plot of the radial measurement, given by the weighted angular average of (3.9) for the images of Fig. 3.5.</td>
<td>61</td>
</tr>
<tr>
<td>3.8</td>
<td>Plot of the angular measurement, given by the weighted radial average of (3.10) for the images of Fig. 3.5.</td>
<td>62</td>
</tr>
<tr>
<td>3.9</td>
<td>Portion of Lena image with additive noise ($\sigma^2 = 100$) (a), and corresponding original portion (b). Considering the radial averages of Fig. 3.7, this illustrates how relatively high noise levels can affect perceptual quality of an image, while having minimal effect on the lower frequencies of the image’s spectrum.</td>
<td>64</td>
</tr>
<tr>
<td>3.10</td>
<td>Cameraman image (a) and spectrum (b) which demonstrates aliasing of spectra ray features with insufficient anti-aliasing filtering.</td>
<td>65</td>
</tr>
<tr>
<td>3.11</td>
<td>Plot of function (3.20) for varying $\alpha$. Detail box in upper-right depicts function for small values of $\Omega_r$.</td>
<td>69</td>
</tr>
<tr>
<td>3.12</td>
<td>Test image used to evaluate the spectral models.</td>
<td>76</td>
</tr>
</tbody>
</table>
Figure 3.13: PSNR (a) and QI (b) for the MMSE recovery of the image shown in Fig. 3.12 using different PSD models for reconstruction filter design.

Figure 3.14: Weighted angular averages for the image in Fig. 3.12 (thin black curve) and its smooth (thick blue curve) and combined spectral ray (dashed red curve) models. (a) Original image, 10 zones. (b) Degraded image, 10 zones. (c) Original image, 80 zones. (d) Degraded image, 80 zones.

Figure 4.1: Filter bank model for super-resolution. A single HR image is degraded via the analysis bank to produce C LR images, which are processed and combined in the synthesis bank to reconstruct the super-resolved output.

Figure 4.2: Test image used to demonstrate improvement over prior bandlimited frequency domain reconstructions.

Figure 4.3: Magnified portions of test scene (a), one of four LR frames at 1/16th resolution (b), frequency-domain reconstruction under bandlimited assumption (c), the proposed LMMSE approach assuming full knowledge of the HR spectrum (d).

Figure 4.4: Super-resolution comparisons for several methods from a set of four degraded LR images with known registration but unknown blurring and noise. Depicted images are: (a) original boat image, (b) one of four degraded LR images (resolution increase using bicubic interpolation for display), (c) MMSE reconstruction using original image PSD, (d) MMSE reconstruction using parametric PSD model estimated from the LR image set, (e) Delaunay triangulation based cubic interpolation using the MATLAB `griddata()` function, and (f) iterative interpolation/deblurring using $L_2$ regularization.

Figure 4.5: Comparison of spectra used in the second super-resolution test. Spectra are put through a log10 scaling to produce visually discernable details for the figure. The PSD of the original image is shown in (a), followed by the model obtained from the LR data in (b), and the PSD of one of the undersampled LR images in (c).

Figure 4.6: PSNR results for several methods of super-resolution with varying amounts of AWGN. Gaussian PSF with (a) $\sigma = 0.6$ and (b) $\sigma = 1.0$.

Figure 4.7: Quality index results for several methods of super-resolution with varying amounts of AWGN. Gaussian PSF with (a) $\sigma = 0.6$ and (b) $\sigma = 1.0$. 

vii
Figure 4.8: Comparison of several images: (a) camera-obtained HR image, (b) $2 \times 2$ bicubic interpolation of individual LR image, (c) proposed MMSE reconstruction from 2 LR images, (d) proposed MMSE reconstruction followed by non-linear post-processing for artifact removal, (e) $L_1$ regularization reconstruction from 2 LR images, and (f) $L_1$ regularization reconstruction from 13 LR images.

Figure 5.1: Illustrative case of core, source information, and correlation blocks superimposed over known pixels of LR frame. The HR pixels to be interpolated correspond here to a $2.5 \times 2.5$ scaling enhancement.

Figure 5.2: Frame super-resolution from eleven consecutive LR frames, three of which are shown in (a): frame 10 in the upper-left, frame 20 in the upper-right, and the base LR frame 15 in the bottom. The original HR frame 15 is shown in (b). The proposed method is applied to the base frame alone to produce (c) and with the ten auxiliary frames to produce (d). Comparison with the MATLAB Delaunay triangulation-based cubic interpolation is shown using the base frame alone for (e) and data from the ten auxiliary frames to obtain (f).

Figure 5.3: Reconstruction PSNR as a function of frame index for the mobile sequence.

Figure 5.4: Reconstruction PSNR as a function of average sampling density, simulating the performance in the case of ideal registration.

Figure 5.5: Reconstruction results for simulated ideal registration. LR base frame (a) and desired HR equivalent (b). The proposed reconstruction at 0.16 samples/HR pixel (c) and 1.3267 samples/HR pixel (d). MATLAB cubic reconstruction with deblurring at 0.16 samples/HR pixel (e) and 1.3267 samples/HR pixel (f). Results (c) and (e) use only the uniformly spaced base frame pixels from (a).

Figure 5.6: Original Athens image used for reconstruction of purely translational motion.

Figure 5.7: Reconstructions of the Athens image (Fig. 5.6) determined through: (a) Hardie reconstruction using three LR frames, (b) Hardie reconstruction using a single LR frame, (c) proposed reconstruction using three LR frames, and (d) proposed reconstruction using a single LR frame.

Figure 5.8: Three LR images simulating zoom of the original elephant image. A simulated zoom factor of 1.2 is present between (b) and (a), and between (c) and (b).
Figure 5.9: Reconstructed images attempting to recover original elephant image: (a) proposed reconstruction using a single LR frame—Fig. 5.8(a), (b) proposed reconstruction using all three LR frames from Fig. 5.8, (c) Hardie reconstruction using a single LR frame, and (d) original HR image.  

Figure 5.10: Pair of distorted LR images used for the optical flow simulation. 

Figure 5.11: Original Einstein image (a), along with reconstructions: (b) Hardie reconstruction using a single LR frame, (c) proposed reconstruction using a single LR frame, and (d) proposed reconstruction using both LR frames from Fig. 5.10. 

Figure 5.12: Einstein reconstructions with varying observation window sizes. Single frame reconstructions are shown in (a)-(d) with respective $s_S = 3, 5, 7, 9$ and two-frame reconstructions are shown in (e)-(h) with respective $s_S = 3, 5, 7, 9$. 

Figure A.1: General multirate filter bank structure. 

Figure A.2: Equivalent reconstruction models used to determine the MSE of a decimated output sequence. 

Figure A.3: Dual analysis bank model used to consider the presence of an interfering source. The system can be generalized to consist of $K + 1$ merging analysis banks to model $x[n]$ in the presence of $K$ distinct interfering sources.
LIST OF TABLES

Table 5.1: Required computation time for correlation estimate, varying $s_{AC}$. The combined remaining calculations require approximately 15 additional seconds. 133
Table 5.2: Required computation times (in seconds) for different components of the restoration process with varying resolution enhancement factor. 133
Table 5.3: Computation times (in seconds) and number of block iterations for processing of 3240 LR pixels, varying core block size. 134
Table 5.4: Computation times (in seconds) for several different sets of block sizes, $m = 2.5$ for all cases. 135
Table 5.5: Computation times (in seconds) for reconstructions with and without auxiliary frame data. 135
Table 5.6: Summary of reconstruction PSNR measurements in the case of ideal registration. 147
Table 5.7: Reconstruction PSNR as size of observation window is increased. 149
ACKNOWLEDGEMENTS

I wish to acknowledge and thank all the individuals who helped bring this dissertation into being. Most importantly, I thank my advisor, Professor Truong Nguyen, who provided academic and financial support during my time at UCSD. I offer my thanks to my fellow students of the UCSD Video Processing Group, with whom I’ve shared many important discussions on research over the years, and the members of my dissertation committee for their suggestions and time. I would also like to express my appreciation to the faculty of UCSD and UC Davis who provided education and inspiration, as well as the many staff members who helped me navigate through the sometimes confusing university system to this point of completion. Finally, I wish to thank everyone who has been a part of my life outside the academic world: my friends, parents, and Nora. Your patience, advice, encouragement, and love provide motivation for everything I do.


Portions of the text of Chapter 4 are adapted from material that has been published as: R. S. Prendergast and T. Q. Nguyen, “Improving frequency domain super-resolution via undersampling model,” in Proc. IEEE Int. Conf. on Image Processing, 2005, and material submitted for publication as: R. S. Prendergast and T. Q. Nguyen, “Spectral modelling and Fourier domain recovery of high-resolution images from jointly undersampled image sets,” under review with IEEE Trans. Image Proc., 2008. The dissertation author was the primary researcher for these
publications, and the co-author directed and supervised the research which forms the basis for this chapter.

Portions of the text of Chapter 5 are adapted from material that been submitted for publication as: R. S. Prendergast and T. Q. Nguyen, “Digital video super-resolution,” under review with *IEEE Trans. Image Proc.*, 2008. The dissertation author was the primary researcher for this publication, and the co-author directed and supervised the research which forms the basis for this chapter.

The text of Appendix A, in full, is adapted from material that has been published as: R. S. Prendergast and T. Q. Nguyen, “Minimum mean-squared error reconstruction for generalized undersampling of cyclostationary processes,” *IEEE Trans. Sig. Proc.*, 2006. The dissertation author was the primary researcher for this publication, and the co-author directed and supervised the research which forms the basis for this chapter.
<table>
<thead>
<tr>
<th>Year</th>
<th>Position</th>
<th>Institution</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>Bachelor of Science with Highest Honors</td>
<td>University of California, Davis</td>
<td>Davis</td>
</tr>
<tr>
<td>2001-2003</td>
<td>Graduate Teaching Assistant</td>
<td>University of California, Davis</td>
<td>Davis</td>
</tr>
<tr>
<td>2002-2003</td>
<td>Graduate Research Assistant</td>
<td>University of California, Davis</td>
<td>Davis</td>
</tr>
<tr>
<td>2003</td>
<td>Master of Science</td>
<td>University of California, Davis</td>
<td>Davis</td>
</tr>
<tr>
<td>2003-2004</td>
<td>Graduate Teaching Assistant</td>
<td>University of California, San Diego</td>
<td>San Diego</td>
</tr>
<tr>
<td>2004-2008</td>
<td>Graduate Research Assistant</td>
<td>University of California, San Diego</td>
<td>San Diego</td>
</tr>
<tr>
<td>2008</td>
<td>Doctor of Philosophy</td>
<td>University of California, San Diego</td>
<td>San Diego</td>
</tr>
</tbody>
</table>
ABSTRACT OF THE DISSERTATION

Modelling and MMSE Reconstruction Solutions for Image and Video Super-Resolution

by

Ryan Strong Prendergast
Doctor of Philosophy in Electrical Engineering

University of California San Diego, 2008

Professor Truong Q. Nguyen, Chair

Super-resolution considers the problem of increasing the spatial resolution of an image or video from one or more observation images or frames. In both cases, the problem seeks to determine a representation of the content at a higher spatial resolution than was originally possible to acquire, store, or transmit. For the case of images, the problem has been considered in the literature for over two decades and a variety of techniques exist. Far fewer results exist for the case of digital video, which is becoming a problem of increasing importance as presence of digital video becomes more prevalent.

The problem is considered through two separate processes: modelling, which describes how multiple individual low-resolution observed images/frames are related to a single high-resolution equivalent, and reconstruction, which recovers the unknown high-resolution version from the observations and the results of the modelling process. As presented, the complete problem is driven by the proposed reconstruction solution, and novel aspects of the modelling problem are introduced based on the needs of the particular reconstruction solution.

For the case of images, a linear minimum mean-squared error (LMMSE) frequency domain solution is proposed using a filter bank model and a stationary stochastic signal assumption. The solution requires estimation of a high-resolution image’s spectral density from its low-resolution observations. Novel parametric
spectral models for images are introduced and applied to the problem.

In the case of video sequences, the presence of temporal motion over multiple frames necessarily leads to more complex registration models, generally prohibiting the application of most standard still-image solutions. Previously, super-resolution methods intended for video have been limited to relatively simple motion models, e.g., global translational motion, based on a reconstruction requirement that the distortion and motion models commute. Relying on a reverse motion model, the proposed approach removes this limitation, consequently extending the result to cases of arbitrary motion models. With the required modelling in place, a LMMSE spatial domain reconstruction is used to determine the reconstructed sequence.
1 Introduction

The super-resolution problem seeks to restore a single high-resolution (HR) representation of a scene from some number of known low-resolution (LR) representations. This inverse problem is generally ill-posed as the information present in the set of observed LR images is typically either insufficient to determine a unique HR representation (underdetermined), or sufficient but self-contradictory, prohibiting a solution from existing (overdetermined). The subject of inverse problems is well studied ([1, 2] are two of many different international journals on the topic), and many of the standard solution techniques for ill-posed problems are often applied to super-resolution scenarios. Investigation of the super-resolution problem has existed for some time [3], and these years have seen the development of a wide variety of techniques [4]. More recent results and ongoing research in the field are providing increasingly feasible solutions, improved reconstruction quality, and extending the more limited results of the past to provide real-time solutions for a variety of video and imaging applications.

The research examined in this dissertation considers the super-resolution problem in two phases: modelling and reconstruction. Although typically considered separately, these processes need to be designed together to produce a high-quality result. The purpose of the modelling phase is to determine all the required information describing the problem scenario. While the specific information required varies depending on the form of the reconstruction solution, typically some representation for the degradation process (a model relating the observed LR images to the desired HR image) and some description of the characteristics of the desired HR image are necessary. Oftentimes, much of this information is assumed.
known a priori, either based on known characteristics of the problem or simply assumed using a convenient mathematical description. In the most extreme case, the problem can be so limited that only a set of LR observations are known, requiring the degradation and content models to be estimated directly from the observations. More commonly, the problem scenario allows much of the degradation model to be assumed as known, although typically at least the positional relationship between individual LR images—known as the image registration problem—must be estimated directly from the observations. As for image characteristics, many reconstruction approaches avoid the problem by implicitly assuming some form of a content model, but reconstruction techniques relying on a specified model typically require additional estimation for its description. In any case, determining the complete model required by the reconstruction process relies heavily on assumptions, and the quality of the reconstruction is largely dependent on the validity of these assumptions.

The reconstruction phase simply takes the observed set of LR images and modelling information to produce the desired HR output. The quality of reconstruction depends on three components: the specific reconstruction algorithm used, the quality of the degradation and content models, and the quantity and quality of the LR observations. This last component is directly uncontrollable by the super-resolution system, which should always be expected to produce a better reconstruction when the amount of source information is increased (assuming accuracy of the observational model). The relative performances of certain reconstruction algorithms may vary depending on the size of observations (e.g., some algorithms may perform better than others in underdetermined scenarios but worse in overdetermined scenarios), but this is generally a secondary consideration since most algorithms are intended to be applicable to a general variety of scenarios. Similarly, although the reconstruction algorithm will dictate what form of information must be supplied by the modelling portion, the reconstruction approach generally must take the provided information as given (although the validity of specific models can certainly be considered within the reconstruction). Otherwise, the reconstruction phase is simply an image-oriented solution to the general ill-posed
1.1 Super-Resolution Application Interest

Image and video super-resolution is of interest to a wide variety of applications, both as an end goal and as a step in a larger analysis process. Potential applications can be found in almost any image or video display scenario due to the intrinsic value of increasing spatial resolution for human observation.

In the commercial sector the strongest interest in super-resolution lies in video. This is a relatively recent interest that is mainly drawn from the increasing prevalence of digital video systems, and the improvement in display technology. Video super-resolution provides a method for enhancing lower-resolution sequences for use with higher-resolution displays, e.g., standard-resolution video on HDTV. In cases where the desired HR sequence is unavailable, a super-resolved HR estimate produced from the available LR sequence provides the end user some measure of the high performance the display is capable of. As new, higher-quality displays become available, there will always be interest in viewing the previous generations' media. In some cases the sequence might only be able to be supplied at a low resolution, e.g., due to transmission or storage capacity constraints. Video super-resolution provides a means to maximize the quality based on the end-user’s resolution capabilities, rather than the signal transmission capabilities. This is of direct interest for cable and satellite television providers, streaming internet video, as well as user-to-user applications such as webcam and wireless video.

Outside of consumer-oriented commercial applications, super-resolution is of interest in security, surveillance, military, and scientific observational applications. Many of these applications differ from consumer applications in that real-time video processing is not required. In general, there is a trade-off between enhancement quality and required computation. For consumer applications requiring real-time processing, there are necessarily increased restrictions on computation time. In off-line processing scenarios, the highest quality output is desired, allowing more costly processing. Some specific still-image applications include astronomi-
cal observations, satellite-based terrestrial observations, or detail extraction from close-circuit television signals (e.g., crime scene analysis). Similar video equivalents for these applications exist as well.

1.2 Comments on Style, Notation, and Nomenclature in this Dissertation

Super-resolution is a widely studied problem, being examined from the standpoints of a number of different fields, each with its own ways of viewing and describing the same things. This section briefly goes over some preliminary discussion in an attempt to put readers on the right track of thought. Although, for the most part, the discussion in the later chapters should be easily accessible on its own.

The super-resolution problem is approached in this dissertation from the vantage of a traditional digital signal processing background. The notation from [5] is used, where time- or spatial-domain functions are referred to using $x(t)$ for the continuous case and $x[n]$ for the discrete case. The Fourier domain relies on a capitalized notation, with $X(\Omega)$ being the continuous Fourier transform and $X(e^{j\omega})$ indicating the discrete-time/space Fourier transform (normalized in frequency to be $2\pi$-periodic). The discrete-Fourier transform makes use of bracketed notation $X[k]$. The second-order statistics are expressed using $R_{xx}$ for correlation domain functions and $S_{xx}$ for the spectral functions, with subscripts specifying the labels of signals involved in the expectation (in this case $x$ with itself). As per signal labeling convention, the parenthesis and straight-bracket notation for respective continuous and discrete representations apply to these statistical functions.

As the content of the dissertation covers both still-image and video super-resolution, the term “image” is often used in general discussion to refer to both still-images and sequence frames. Cases where the term is applicable only to images and not frames (or vice-versa) are either explicitly noted or should be clear from the context. Although the video problem can be examined as a series of individual still-image problems, there are generally more complications which need to be
accounted for in the reconstruction (see Chapter 5).

The term “super-resolution” itself has different meanings in various uses. In this dissertation, the term is intended to specify resolution enhancement from multiple observational images. This is intended to differentiate the problem from that of the simpler single-image case, referred to as resolution enhancement. The multiple-image case requires some representation for the registration between images, which is one of the most significant complications of the super-resolution problem. Scenarios without resolution enhancement (e.g., noise and distortion removal) are referred to as restoration problems, regardless of the number of observation images. In this dissertation, restoration can be viewed as a subset of enhancement, lacking the complication of a required content model beyond the sampling rate of the individual observations. There are also scenarios which muddy the distinction between these categories.

1.3 Organization and Intent

This dissertation provides novel results in content modelling, still-image super-resolution, and video super-resolution. Each of these topics is the main subject of a chapter, with an additional chapter related to background and degradation system modelling.

Chapter 2 provides an introduction to the super-resolution scenario by detailing models for the degradation system. Individual LR images are represented as produced from the HR equivalent using a series blurring-sampling-noise system. Each of these degradation components is discussed in detail, providing a foundation for the contributions presented in later chapters. A discussion of image registration is also presented.

Content modelling is the subject of the Chapter 3. Prior and novel models are discussed. Separate examination is provided for global and local modelling approaches. Primary interest is on models used by the reconstruction algorithms in the subsequent chapters, with a focus on a novel global spectral model for still-image applications, and a local correlation domain equivalent for video.
Still-image super-resolution enhancement is provided in Chapter 4. A novel minimum mean-squared error (MMSE) solution approach based on an undersampled filter bank interpretation is provided. Alternative approaches are discussed and comparative simulation results are provided.

Video super-resolution is examined in Chapter 5. A discussion of the complications of the video scenario over the image scenario is provided. The proposed MMSE based approach is presented along with some alternative approaches which were intended for video applications. Reconstruction results are presented.

Chapter 6 provides a review of the contributions.
2 Modelling the Image Degradation Process

2.1 Introduction

This chapter discusses modelling of the image degradation, or image acquisition, process. The quality of a super-resolution reconstruction is largely dependent on the accuracy of this model describing how a set of LR images is obtained. The later chapters discuss the topics of content modelling, which often requires an accurate degradation model, and reconstruction, which requires accurate models for both degradation and content. There is consequently a high degree of importance placed on this chapter’s models. That being said, there are no new contributions presented in this chapter. The main purpose is to provide details on the types of models that are used and the characteristics of some specific models. Establishing an explicit model is necessary for determining what the super-resolution problem is, and how to solve it.

The second purpose of the chapter is to introduce some of the background necessary for the later chapters. As much of the novel content modelling and reconstruction results rely heavily on Fourier analysis, this chapter is intended to serve as a short review of the subject and introduce much of the notation used later on. As an alternative to simply listing the notation and background, the necessary subjects are mentioned as they come up naturally in the discussion. Organization is as follows. First, the standard image degradation model is presented in Section 2.2. A discussion of specific models for distortion, noise, and sampling are presented
in Section 2.3. With these models in place, a review of classic image restoration is provided in Section 2.4. The final component of the complete image acquisition model, registration (or motion modelling) is discussed in Section 2.5.

2.2 Degradation Modelling

A degradation model is used to describe how individual known LR images were obtained from their unknown HR equivalents. A block diagram representation of the most basic degradation system is depicted in Fig. 2.1. The three blocks in this system represent the commonly considered types of degradation:

**Distortion**, in which a transfer function acts on the initial scene. In the image acquisition process this is best considered using a continuous-space function that describes the misdirection of light from the desired 2D projection. Some of the commonly considered sources of acquisition distortion include motion blur, atmospheric blur, and lens/focus induced distortion ([6], Chapter 3.5). These are typically modelled using a linear point spread function (PSF). Distortion can also be induced digitally in HR-to-LR reductions, e.g., as an anti-aliasing process or as part of a coding transform. The digital model uses a discrete PSF.

**Sampling**, or resolution reduction, is responsible for producing the LR version of the (distorted) image. In a digital scenario this can be a simple decimation operation for integer scale reduction or combined sub-sampling interpolation/decimation for a more generalized rational scale reduction. For the continuous case, sampling is done on the camera through a charge coupled device (CCD) or CMOS sensor, which detects the intensity of the light focused through a lens to a specific position and produces a voltage signal corresponding to intensity. The sampling model can also be used to include registration information, giving the relative locations of individual pixels between images in the LR set (in some cases registration can also be a part of the distortion model). Individual images are typically only sampled uniformly,
but combined multiple images can introduce nonuniform sampling. Certain acquisition approaches or applications can also introduce alternative forms of sampling (e.g., mosaicing, in which each color space is sampled differently and requires special post-acquisition processing to produce the final image, or digitally induced interlacing [7]). Localized motion in video scenarios often introduces additional complications by requiring localized motion/sampling models.

**Additive noise**, which is the most common model for noise introduction. Alternatives like multiplicative noise exist in certain applications. Noise sources include: coding, quantization, CCD electrical error ([6], Chapter 4.5). Noise models can also be used to compensate for other model inaccuracies, such as registration error. Typically, statistical models are used requiring some information on the magnitude of noise and any correlation with the content or autocorrelation.

![Figure 2.1: Simple linear image degradation system.](image)

This thesis will assume the same operational order depicted in Fig. 2.1. In practice there can be instances of noise occurring prior to the sampling and distortion occurring subsequent to the sampling. However, in these cases it is always possible (assuming linear distortions and statistically independent additive noise) to determine an equivalent model in the form of Fig. 2.1. Most super-resolution reconstruction algorithms can still be applied to alternative degradation system models with only minor to moderate modification.

There is some difference in the modelling system depending on whether the input source is considered to be spatially continuous or discrete. While it is assumed that the true image is a continuous 2D projection of some real-world scene, it is often convenient to use a discrete representation for a couple reasons.
First, the infinite support of continuous models necessitates parametric modelling, and second, discrete models are more easily worked with by a computer solution. The choice between a continuous or discrete representation primarily affects only the distortion and sampling models since, subsequent to sampling, the LR version is always discrete. Although a continuous equivalent for the noise statistical model can be determined, there is no loss associated with using the discrete form. The form of the distortion and sampling models generally depends on the form of the image content model, which will be discussed in the next section. If a discrete model is used, it is generally assumed to provide an unadulterated representation (i.e., a Nyquist-rate sampling of a bandlimited continuous function [8–10]). In practice this is not possible, but the error resulting from a discrete version of the model can typically be kept sufficiently small, subject to image content. In scenarios using a digital original (e.g., resolution reduction of a digital image for lower bit-rate), the discrete model can be fully accurate, and is often known a priori, avoiding the computational and accuracy penalties associated with estimation.

The next section discusses specific degradation models.

2.3 Standard Models

2.3.1 Distortion

Almost always a linear model for distortion is assumed, where the pixels of the distorted output image are simply the weighted sum of the pixels of the input image. Only the discrete models will be presented here, and the continuous equivalents can be easily determined if needed. It simplifies things greatly if the distortion can be assumed shift-invariant, providing a linear shift-invariant (LSI) system, which has the notable advantages of being relatively simply represented and easily worked with using classic signal processing techniques (e.g., Fourier domain analysis). An introduction to traditional discrete-space 2D signal processing is found in [11].

LSI distortion is applied to an image through the convolution equation
which provides the output \( y[n_1, n_2] \) image as a function of the input \( x[n_1, n_2] \) and transfer function \( h[n_1, n_2] \). This is represented in the spatial domain as

\[
y[n_1, n_2] = x[n_1, n_2] * h[n_1, n_2]
\]

\[
= \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} x[m_1, m_2] h[n_1 - m_1, n_2 - m_2]
\]

\[
= \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} h[m_1, m_2] x[n_1 - m_1, n_2 - m_2]
\] (2.1)

In practice, the transfer function is typically finite impulse response (FIR) with a non-zero component significantly smaller than the total size of the image.

The 2D discrete-space Fourier transform is defined as

\[
X(\omega_1, \omega_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} x[n_1, n_2] e^{-j\omega_1 n_1} e^{-j\omega_2 n_2},
\] (2.2)

and its inverse is given by

\[
x[n_1, n_2] = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} X(\omega_1, \omega_2) e^{j\omega_1 n_1} e^{j\omega_2 n_2} d\omega_1 d\omega_2.
\] (2.3)

For analysis of LSI distortion, the most useful property of the Fourier transform is the equivalence

\[
x[n_1, n_2] * h[n_1, n_2] \longleftrightarrow X(\omega_1, \omega_2) H(\omega_1, \omega_2)
\] (2.4)

which, applied to 2.1, allows Fourier domain representation of the distorted image as a product of the input image and transfer function. The utility of Fourier domain processing extends from analysis tools (such as spectral modelling, see Section 3.3.1) to methods of reconstruction implementation.

For digital implementation, it is more convenient to use the 2D discrete Fourier transform (DFT) as an alternative to the discrete-space Fourier transform. This representation assumes periodic/impulsive duality for both the spatial and the frequency domain [11]. Due to the finite length of both digital images and distortion response models, this representation can be practical without further modification. (In contrast, there can exist some practical implementation problems in the case of non-finite time-domain signals and systems [5].) The 2D DFT
representation for a discrete \((N_1, N_2)\) signal is given by
\[
X[k_1, k_2] = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x[n_1, n_2] e^{-j(2\pi/N_1)n_1k_1} e^{-j(2\pi/N_2)n_2k_2}. \tag{2.5}
\]
As the DFT is \((N_1, N_2)\) periodic as well, it need only be computed on \(0 \leq k_1 \leq N_1\) and \(0 \leq k_2 \leq N_2\). The inverse DFT is given by
\[
n[n_1, n_2] = \frac{1}{N_1N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} X[k_1, k_2] e^{j(2\pi/N_1)n_1k_1} e^{j(2\pi/N_2)n_2k_2}. \tag{2.6}
\]
For a finite length signal, the representation (2.5) contains all information present in (2.2), meaning the continuous representation \(X(\omega_1, \omega_2)\) is redundant and need only be represented by its samples, uniformly\(^1\) spaced on the points \(\omega_1 = 2\pi k_1/N_1\) and \(\omega_1 = 2\pi k_2/N_2\). Portions of the continuous function \(X(\omega_1, \omega_2)\) existing between these sample points can be determined directly from the samples by convolving the sampled function with the discrete-space Fourier transform of
\[
\text{rect}(N_1, N_2) = \begin{cases} 
1 & \text{for } 0 \leq k_1 \leq N_1, \ 0 \leq k_2 \leq N_2 \\
0 & \text{elsewhere,} 
\end{cases} \tag{2.7}
\]
i.e., a Fourier-domain circular convolution [5] with a 2D sinc function. By far the most attractive feature of the DFT representation is that it can be computed very efficiently using fast Fourier transform (FFT) processing techniques [5,11].

**Specific Distortion Models**

This dissertation only considers linear distortion models. The most prevalent model used is a Gaussian PSF which is given in the continuous form by
\[
h(t_1, t_2) = \frac{1}{2\pi\sigma^2} e^{-\frac{(t_1^2 + t_2^2)}{2\sigma^2}}. \tag{2.8}
\]
The function is circularly symmetric, or isotropic, meaning its value varies only as a function of radius \((r = \sqrt{t_1^2 + t_2^2})\), and not as a function of angle. Interestingly, the function is also separable, meaning it can be written as the product of two distinct 1D functions \(h(t_1, t_2) = h_1(t_1)h_2(t_2)\), which greatly simplifies finding its

\(^1\)A nonuniform representation can also be considered [12]
continuous Fourier transform since the Fourier transform of a multi-dimensional separable function is the product of the 1D Fourier transforms of its component functions. Using the 1D Fourier transform pair [13]

\[ h(t) = e^{-\frac{t^2}{2\sigma^2}} \iff H(\Omega) = \sqrt{2\pi}\sigma^2 e^{-\frac{\Omega^2\sigma^2}{2}} \]  

(2.9)

it is easily shown that the 2D continuous Fourier transform of the Gaussian function (2.8) is

\[ H(\Omega_1, \Omega_2) = e^{-\frac{(\Omega_1^2 + \Omega_2^2)\sigma^2}{2}}, \]

(2.10)

which is also a Gaussian function. Note that a more generalized form of the 2D Gaussian function is the multivariate normal function which can be found as a rotated and scaled version of (2.8), with corresponding rotations and scalings carrying through to its Fourier transform [11, 14]. For discrete-space models, a Gaussian-like distortion can be reasonably approximated by sampling a scaled, finite portion of (2.8), providing an FIR PSF model. The samples of the model lost to truncation are typically of insufficient magnitude to noticeably affect the outcome. A discrete-space Fourier transform of this PSF can be determined numerically. Note that while the spatial function is determined from the truncated samples of a continuous Gaussian function, the discrete space Fourier transform of this approximation is not also Gaussian, due to sampling-induced aliasing (see Section 2.3.3).

The Gaussian PSF serves as a reasonable approximation for much of practical distortion found, such as lens and atmospheric induced blur. One specific form of distortion which is not well modelled by a Gaussian PSF is motion blur. For the most basic example, consider a uniform motion blur resulting from the movement of the image acquisition device in a single direction, given by angle \(\phi\), at a constant velocity, \(v_d\). Practical image acquisition is not instantaneous, and the recorded image is (for this example) simply modelled as the uniformly weighted combination of all light acquired over this acquisition time, \(t_a\). As shown in Chapter 3.5 of [6], this results in a linear blur length \(L = v_d t_a\) along the angle \(\phi\), which
can be considered in a continuous form as
\[
h(t_1, t_2) = \begin{cases} 
1/L & \text{if } \sqrt{t_1^2 + t_2^2} \leq L/2, \quad -\tan(\phi) = t_1/t_2 \\
0 & \text{otherwise.}
\end{cases}
\] (2.11)

This simply spreads the image along a straight line segment of length \( L \). As this is essentially a 1D “box function”, the Fourier transform of (2.11) will contain some zeros, i.e., theoretically irrecoverable content. A discrete model for (2.11) will require some approximation. The estimation of more complicated motion blur models (non-straight line) is presented in [15].

A final complication is introduced by the fact that real distortion may not be shift invariant. For instance the distortion introduced by the lens of the acquisition device can be more significant for the edge portions of the image than for the center portions. Similarly, localized motion of objects within an image can lead to different levels and directions of motion blurring throughout the image. The two main consequences of shift-variant distortion are that a more complicated model is required and that the solution approach must be adjusted locally. The effect of both these consequences is to increase the required computational complexity of the modelling process and solution. Fourier solution approaches, which are efficient but based on an assumption of shift-invariance, cannot be used in this general case as the convolution equation (2.1) no longer applies. Instead, the distorted output is given by a function such as
\[
y[n_1, n_2] = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} x[m_1, m_2] h[n_1, n_2][m_1, m_2],
\] (2.12)
in which each output pixel is still a linear combination of input pixels. However, unlike in (2.1), a unique PSF is defined by the index of the output pixel. A direct spatial-domain solution is often required in this case. It is also often possible to regionally designate a PSF, through which a shift-variant function is defined as the piecewise composite of several shift-invariant functions, a fact which can be exploited for improved computational efficiency, e.g., by applying the Fourier transform to local regions.
2.3.2 Noise

In contrast to distortion, which corrupts information from individual pixels of the image with each other, noise degradation corrupts the pixels of the image with an error distinct and separate from the content (although it should be noted that it is possible to model non-zero correlation between the noise and image content). The most common form of noise is additive, as depicted in the system of Fig. 2.1. While alternative forms of noise models exist for certain applications, e.g., multiplicative noise, the remainder of this thesis will consider only the standard additive noise scenario. Additional noise models and scenarios are provided in Chapter 4.5 of [6].

Excluding distortion, the noisy digital image is given by

\[ y[n_1, n_2] = x[n_1, n_2] + w[n_1, n_2], \] (2.13)

where \( x[n_1, n_2] \) is the uncorrupted image and \( w[n_1, n_2] \) is the noise. Ideal recovery of \( x[n_1, n_2] \) is found simply through subtracting the noise term from \( y[n_1, n_2] \). However, this is practically impossible since the exact noise term is unknown in all practical cases. Practical denoising must instead rely on an inexact description of the characteristics of \( w[n_1, n_2] \). This is most commonly done using a wide-sense stationary (WSS) statistical model. The WSS model assumes the first and second-order expectations are invariant to spatial shifts, respectively expressed as

\[ E[w[n_1, n_2]] = E[w[n_1 + v_1, n_2 + v_2]] = \mu_w \forall v_1, v_2 \in \mathbb{Z}, \] (2.14)

and

\[ E[w[n_1, n_2] w^*[m_1, m_2]] = E[w[n_1 + k_1, n_2 + k_2] w^*[m_1 + k_1, m_2 + k_2]] \]
\[ = R_{ww}[n_1 - m_1, n_2 - m_2] \]
\[ = R_{ww}[\tau_1, \tau_2] \forall k_1, k_2, \tau_1, \tau_2 \in \mathbb{Z}. \] (2.15)

Continuous space versions of these models are generally not needed since the noise can be represented as applied strictly to the discrete image. The property of general, or strict-sense, stationarity extends the requirements (2.14) and (2.15) to
all higher-order expectations. Higher-order statistical models have utility in certain image processing applications, e.g., texture classification, but are not considered further in this thesis. Restricting the models to the WSS case provides the distinct implementation advantage of allowing for linear system representation, which is efficiently processed with standard Fourier domain analysis as discussed above. The implications of this on the classic restoration problem are shown in Section 2.4.

The discrete-space Fourier transform of (2.15), the autocorrelation function, provides the power spectral density (PSD) function, denoted as

\[ S_{ww}(e^{j\omega_1}e^{j\omega_2}) = \mathcal{F}\{ R_{ww}[\tau_1, \tau_2] \} . \]  

(2.16)

The most commonly considered case of image noise is zero-mean \((\mu_w = 0)\) and white, meaning

\[ R_{ww}[\tau_1, \tau_2] = \sigma^2 \delta[\tau_1, \tau_2] \]  

(2.17)
in which the Kronecker delta function \(\delta[\tau_1, \tau_2]\) implies a lack of noise correlation between different pixels of \(y[n_1, n_2]\). Applying (2.16) to (2.17), the PSD of white noise is shown to be spectrally flat, or of equal power at all frequencies. Colored noise, e.g., as can arise through filtering white noise, is not spectrally flat.

Although not commonly considered in practice, the cross-correlation between the signal and noise, denoted as \(R_{xw}[\tau_1, \tau_2] = E[x[n_1, n_2]w^*[m_1, m_2]]\) can also be defined in the same fashion as (2.15), along with the corresponding cross spectral density function. In most cases it is assumed that \(R_{xw}[\tau_1, \tau_2] = 0\), but if a non-zero cross-correlation model is known to exist, it can be used to improve the restoration process as will be discussed in further detail in Section 2.4.

In addition to correlation, noise models often include a probability density function (PDF), indicating the distribution of noise amplitude. The two typical PDF models are Gaussian and uniform. The Gaussian (or normal) distribution is defined by the PDF

\[ p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} , \]  

(2.18)

for which \(\mu\) indicates the mean value and \(\sigma^2\) indicates the variance. An important property of Gaussian noise is that the linear combination of Gaussian variables
yields a Gaussian variable [17], which guarantees the noise stays Gaussian through any linear distortion. Additionally, the central limit theorem [17] provides that the distribution of the sum of \( n \) independent random variables approaches the normal distribution as \( n \to \infty \), which promotes the use of the normal distribution for otherwise intricate noise sources (e.g., CCD electrical error). These results allow the Gaussian PDF to serve as the default distribution for additive error. The proper selection for \( \sigma^2 \) can be estimated, known a priori, or it can be simply assumed to equal some reasonable value that provides sufficient quality of results.

The one alternative exception to the prevalence of the Gaussian PDF is the uniform distribution, which is of specific interest to model the effects of quantization and lossy encoding. Under a standard uniform quantization model [14], a pixel’s value is rounded to the amplitude level in the set \( kQ, \forall k \in \mathbb{Z} \), where \( Q \) is the quantization spacing. This model is most commonly seen for the digitization of a signal, but the same approach can be used in encoding applications, e.g., as applied transformed image data as in JPEG, which uses varying \( Q \) values for different portions of data having been subjected to a discrete cosine transform (DCT) [14]. Subsequent to rounding, the value of an individual pixel is given by

\[
p_q = p + n_q,
\]

where \( p \) is the true amplitude, \( p_q \) is the quantized amplitude, and the error between the two is given by \( n_q \), which must be bounded on the interval \([−Q/2, Q/2]\) due to nearest-neighbor rounding. The exact distribution of the error variable is highly dependent on the image characteristics, but is generally assumed to be uniform over the interval. However, this is not necessarily the case and alternative distributions can be used if known. Under the uniform assumption, the mean is zero and the variance is found to be \( \sigma^2_u = Q^2/12 \) [5].

Note that in practice it is often only necessary to know the distribution variance \( \sigma^2 \) (the mean is almost always zero), although the required model detail is dependent on the particular restoration or enhancement algorithm. In such cases, knowledge of the actual PDF is useful only to help determine this value. This is seen for the case of zero-mean white noise (2.17), which only makes use of
the variance (noise power).

### 2.3.3 Sampling

The sampling model determines the relationship between the high- and low-resolution versions of the scene. Although it is typically more convenient in implementation to assume that the original HR image (the input signal of Fig. 2.1) is discrete, it is useful to define the relationship between the “true” continuous version of the scene and any discrete (HR or LR) representation. It will generally be assumed that the discrete representation is obtained through rectangular uniform sampling. Non-uniform sampling is typically not encountered except for the case where multiple images or frames are examined together in a super-resolution problem. In this case the locations of individual samples (pixels) are defined by the image registration model, which is to be discussed in Section 2.5.

A continuous 2D scene, \(x(t_1, t_2)\), which is sampled uniformly can be conveniently examined using its Fourier domain equivalent,

\[
X(\Omega_1, \Omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t_1, t_2) e^{-j\Omega_1 t_1} e^{-j\Omega_2 t_2} dt_1 dt_2. \tag{2.20}
\]

Uniform sampling gives \(x[n_1, n_2] = x(n_1 T_{s1}, n_2 T_{s2})\), the discrete spatial domain signal which provides the instantaneous value of the continuous scene at the sample points, with uniform sampling rates \(T_{s1}\) and \(T_{s2}\). A strictly continuous version of the sampling can be mathematically represented by multiplication with a 2D impulse field of Dirac delta functions, which is defined as

\[
\delta(t_1, t_2) = \sum_{n_1 = -\infty}^{\infty} \sum_{n_2 = -\infty}^{\infty} \delta(t_1 - n_1 T_{s1}, t_2 - n_2 T_{s2}). \tag{2.21}
\]

The sampled representation is then found to be

\[
x_s(t_1, t_2) = x(t_1, t_2) \delta(t_1, t_2) = \sum_{n_1 = -\infty}^{\infty} \sum_{n_2 = -\infty}^{\infty} x(n_1 T_{s1}, n_2 T_{s2}) \delta(t_1 - n_1 T_{s1}, t_2 - n_2 T_{s2}). \tag{2.22}
\]

From this description it is seen that \(x_s(t_1, t_2) = 0\) for non-integer \(n_1\) or \(n_2\), meaning all information of \(x(t_1, t_2)\) outside the sampling points is lost. Recovery of
this information (either perfect recovery or a non-ideal estimate) requires some knowledge of the characteristics of the original continuous signal.

A Fourier domain equivalent of (2.22) often provides a better tool for analysis. Spatial domain multiplication between $x(t_1, t_2)$ and $\bar{\delta}(t_1, t_2)$ is equivalent to a frequency domain convolution of $X(\Omega_1, \Omega_2)$ and

$$
\tilde{\Delta}(\Omega_1, \Omega_2) = \mathcal{F}\{\bar{\delta}(t_1, t_2)\}
$$

$$
= \frac{2\pi}{T_{s1}T_{s2}} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \delta(\Omega_1 - k_1 \Omega_{s1}, \Omega_2 - k_2 \Omega_{s2}),
$$

(2.23)

with sampling frequencies $\Omega_{s1} = 2\pi/T_{s1}$ and $\Omega_{s2} = 2\pi/T_{s2}$. This leads to the Fourier domain sampled function

$$
X_s(\Omega_1, \Omega_2) = X(\Omega_1, \Omega_2) * \tilde{\Delta}(\Omega_1, \Omega_2)
$$

$$
= \frac{2\pi}{T_{s1}T_{s2}} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} X(\Omega_1, \Omega_2) * \delta(\Omega_1 - k_1 \Omega_{s1}, \Omega_2 - k_2 \Omega_{s2})
$$

$$
= \frac{2\pi}{T_{s1}T_{s2}} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} X(\Omega_1 - k_1 \Omega_{s1}, \Omega_2 - k_2 \Omega_{s2}).
$$

(2.24)

The result is based on a specific property of the Dirac delta function: that the convolution of a function with a shifted delta function produces a shifted version of the function. The collection of these shifted copies provides the baseband scaled version of $X(\Omega_1, \Omega_2)$ along with an infinite number of Fourier images, which can overlap with one another causing aliasing.

The Fourier representation for sampling, (2.24), is most familiarly used with the classic result of the Whittaker-Shannon-Kotelnikov (WKS) sampling theorem, which demonstrates the theoretical perfect recovery of a bandlimited signal. Specifically, consider a 2D signal with the Fourier domain property of

$$
X(\Omega_1, \Omega_2) = 0 \quad \forall |\Omega_1| > \beta_1, |\Omega_2| > \beta_2,
$$

(2.25)

i.e., rectangularly bandlimited with bandlimit $(\beta_1, \beta_2)$. From (2.24) it can be seen that the individual Fourier images will not overlap (i.e., be free from aliasing).

---

2This terminology comes from classic signal processing and is unrelated to the “images” generally considered in image processing.
provided that $\beta_1 \leq \Omega_{s1}/2$ and $\beta_2 \leq \Omega_{s2}/2$, which will be referred to as the rectangular Nyquist bandlimiting condition\textsuperscript{3}. The principal result of the WKS sampling theorem is that the original signal $X(\Omega_1, \Omega_2)$ is perfectly reconstructible from its sampled version $X_s(\Omega_1, \Omega_2)$ if $X(\Omega_1, \Omega_2)$ is sufficiently bandlimited to be alias free. The reconstruction is found simply through an ideal lowpass filter,

$$
H_{\text{rec}}(\Omega_1, \Omega_2) = \begin{cases} 
\frac{T_s1 T_s2}{2\pi} & \text{for } 2\Omega_1 \leq \Omega_{s1}, 2\Omega_2 \leq \Omega_{s2} \\
0 & \text{elsewhere.}
\end{cases}
$$

(2.26)

This is the classic result of uniform sampling and signal reconstruction, which helps to form the theoretical justification that makes possible the digital processing of analog signals. More detailed examinations of this standard sampling result and related topics are widely available in many textbooks and collected works \cite{5,12,16}.

Unfortunately true image content is never ideally bandlimited. The magnitude of $X(\Omega_1, \Omega_2)$ will tend to decrease to nearly nothing at high spatial frequencies (relative to content), to the extent that the image can be sampled sufficiently to appear perfect to a human observer. An alternative condition to (2.25) is a practical bandlimit

$$
X(\Omega_1, \Omega_2) < \epsilon_n \quad \forall |\Omega_1| > \beta_1, |\Omega_2| > \beta_2,
$$

(2.27)

where $\epsilon_n$ is some threshold sufficiently below the noise introduced by the system. For this case, the effects of aliasing can be discounted as insignificant in comparison to that of additive noise. However, this does prevent theoretically alias free sampling. One solution for guaranteeing alias-free sampling is to apply the ideal reconstruction filter (2.26) to the image prior to sampling, referred to as an anti-aliasing filter. This acts as a linear distortion on the image, and has the effect of completely eliminating all frequencies falling outside the rectangular Nyquist support region. In exchange, all frequencies within the support are entirely free from aliasing degradation. In practice, the anti-aliasing filter does improve perceptual quality and reduce certain measures of quantitative error. Since the magnitude of

\textsuperscript{3}Unlike the 1D case, it is possible to construct alternative contiguous lowpass support regions which will still be alias free without meeting the rectangular Nyquist condition. A simple example which can easily be visualized is the parallelogram support. In both the 1D and 2D cases, non-contiguous or non-symmetric regions can be determined which are alias free but do not meet the classic Nyquist bandlimiting condition.
the frequency response tends to decrease as spatial frequency increases (detailed discussion on this is found in Chapter 3), it is generally worthwhile to retain these lower frequencies at the expense of the higher ones. It has also been shown that, from the standpoint of minimizing mean-squared error (MSE) of the reconstructed continuous signal, it is a better choice to rely on a non-ideal anti-aliasing filter with a response that is adjusted based on the magnitude of \(X(\Omega_1, \Omega_2)\) [19]. In this sense, the anti-aliasing filter acts as a prefilter that distorts all frequencies of the signal and requires a matching non-ideal reconstruction filter. This introduces aliasing error to all frequency components, but minimizes MSE.

In practice, neither of these implementations for anti-aliasing are used. The ideal lowpass filter (2.26), is infinite impose response (IIR) and thus not realizable, and adjusting a filter based on source content is impractical. Instead, a compromise between these two is employed, which is a static non-ideal filter. This filter is both finite impulse response (FIR) and generally shaped to accommodate the typical magnitudes of images, specifically as a low-pass dominated function. An example is the Gaussian function, given through a truncated, scaled, and sampled version of (2.8). Many other approaches such as bilinear or bicubic filtering [14] are common. As with distortion modelling, both anti-aliasing and subsequent post-processing can be made more efficient if shift-invariant functions are used. However, there can also be reason to use shift-varying functions, such as in the case of certain textured content.

Through (2.24) it is shown that the Fourier transform of the impulsive function (2.22) is a periodic function with rectangular periodicity \((\Omega_{s1}, \Omega_{s2})\). Using the discrete signal \(x[n_1, n_2]\) in substitute for \(x_s(t_1, t_2)\) allows use of the discrete-space Fourier transform representation of (2.2) and (2.3). Since \(x_s(t_1, t_2) = 0\) outside the sample instants, the two representations are composed of identical content in the spatial domain. Likewise, scaling the function \(X_s(\Omega_1, \Omega_2)\) and normalizing its periodicity to \((2\pi, 2\pi)\) leads to the discrete-space Fourier transform \(X(\omega_1, \omega_2)\). If \(x_s(t_1, t_2)\) is of finite non-zero duration, or is periodic, direct use of the DFT representation through (2.5) and (2.6) can be used.

The typical goal of digital image resolution enhancement is not to recover
a continuous signal from its sampled version, but to determine a discrete HR representation from a discrete LR representation. There are two ways for this problem to be considered: either the LR image is degraded from the desired HR image which is selected as the ground truth, or both the LR and HR images are degraded from the true continuous image. These cases will be respectively labeled discrete-to-discrete (D→D) and continuous-to-discrete (C→D) modelling.

D→D modelling certainly presents a computational advantage, since the entirety of image information is guaranteed contained in a finite representation. However, the validity of the D→D approach can break down subject to the image content and the relative sampling ratios of the two representation. The standard linear system used for purely discrete sample rate conversion is shown in Fig. 2.2 (depicted for the 1D case, but easily extended for multidimensional signals), which passes the input $x[n]$ through an $L$-fold expander, followed by an interpolation filter $h_{\text{INT}}[\cdot]$, and finally an $M$-fold decimator to produce the output $x[m]$. Multirate operations limit $M$ and $L$ to positive integers [20]. If the input signal is sampled from the continuous version according to $x[n] = x(nT_s)$, then the output is ideally equivalent to having sampled the continuous signal via $x[m] = x(mM/LT_s)$. Assuming $M > L$ (resolution reduction) and that $T_s$ meets the Nyquist interval criterion, it is possible to achieve this ideal rate transformation with an IIR $h_i[\cdot]$. Although the original signal will not be bandlimited and an IIR filter is practically infeasible, it is possible to consider the original signal as practically bandlimited (2.27) and make use of an FIR approximation to the ideal IIR filter. However, if $T_s$ represents sub-Nyquist sampling, then the D→D model can be insufficient as it will be impossible to undo the effects of aliasing through Fig. 2.2. Finally, since $M$ and $L$ are limited to integers, the system cannot accommodate an irrational change in sampling rate\(^4\). Irrational resampling rates can be accomplished with a general time-varying system alternative to the periodically time-varying Fig. 2.2.

\(^4\)In some ways this turns out to be less of a concern for image processing than for traditional time-domain signal processing applications. Using sufficiently large $M$ and $L$, an irrational rate change can be approximated (although there are certainly some concerns associated with implementation). Since digital images are finite length, a moderately accurate approximation can be sufficient for the entirety of the signal being considered. In contrast, long duration time-domain signals will eventually suffer from a breakdown in the validity of the rational approximation.
The alternative choice of C→D modelling does not lend itself as easily to implementation, but also lacks some of the theoretical limitations of the D→D approach. For instance, irrational rate change is possible to model. Perhaps the main advantage of using a continuous model is that a representation of the entirety of the scene is assumed known in the continuous model. If the HR discrete representation is assumed to contain all known information of the true image, this implies that it is alias free. In actuality, there may be some non-insignificant level of aliasing in the HR representation. If so, then there is generally no method to directly obtain the LR representation from the HR representation, except in very specific cases where the HR sampling rate in each dimension is divisible by the LR sampling rate in the respective dimension, which corresponds to regular partitioning of the individual HR aliasing bands (referring to Fig. 2.2, this would be no expansion operation, or $L = 1$, leading to $x[m] = x(mMT_s) = x[Mn]$ in the absence of an anti-aliasing filter). By relating both resolutions to the continuous image any relative change can be examined. However this is not sufficient for practical solutions as under both the C→D and D→D models any resolution enhancement is still an inverse problem in which the known LR representation is underdetermined for recovering the desired HR representation.

To summarize, the main advantage of a C→D model is that it provides a more accurate representation of the true content than the HR sampling, and the main advantage of a D→D model is that it allows for simple digital implementation. These advantages can be combined through use of a very high resolution (VHR) discrete version of the continuous image, which is assumed to be alias free, i.e., practically bandlimited (2.27). This enables a solution based entirely on discrete-processing for ease of implementation. Since the practical image will be of finite length, all representations can be assumed periodic, allowing for a DFT-based solution. Additionally, the VHR representation completely contains all information.
to produce both the HR and LR representations, but does not require the HR representation to be capable of producing the LR representation. Discussion of content representation models is the subject of Chapter 3. More detailed discussion related to the sampling relationship between different resolutions of images is presented in [21].

**Practical Image Sampling**

Use of a rectangular field of the Dirac delta functions is mathematically convenient for modelling the sampling process, but is not a realistic portrayal of the actual physical sampling done by the digital image acquisition device. At the base level, the acquired image information is often in a much different state. In typical applications the source information is converted to the standard image pixel array directly by the electronics, but advanced applications can sometimes use the source information directly to improve performance. In super-resolution applications it is often sufficient to assume the given pixel array was ideally sampled, depending on the acquisition characteristics and the desired level of resolution enhancement. It is also often the case that the given LR data is known without specifics of the acquisition process.

One of the key differences between the sampling model discussed above and practical sampling systems is the shape of the sample pulse. It is of course impossible to sample with an ideal delta function, although the photodetector device is hopefully small enough to sufficiently approximate the ideal function for the application (consider also that image display will essentially suffer from the same problem due to the shape of display device pixels). The detector surface will be a 2D shape converting the total level of incoming light into an electrical signal. A more appropriate model for information collected by the detector is the integral of the continuous function $x(t_1, t_2)$ on the 2D detector surface, rather than an ideal sampling function. Substituting this change into the sampling function
(2.22) provides
\[
x_s(t_1, t_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \left( \int_{T_{n_1, n_2}} x(t_1, t_2) dt_1 dt_2 \right) \delta(t_1 - n_1 T_{s_1}, t_2 - n_2 T_{s_2}),
\]
(2.28)
where \(T_{n_1, n_2}\) is the surface region for the \((n_1, n_2)\)th detector. Note that even if the acquisition is non-impulsive, this approach still produces an impulsive sampled function, \(x_s(t_1, t_2)\). If the area integral is scaled by the area of \(T_{n_1, n_2}\), the function (2.28) will approach the ideal case of impulsive sampling (2.22) as the area approaches zero.

A second key difference is the layout of the sampling pattern. A rectangular sampling pattern was assumed above. This is convenient mathematically because the pattern is easily separated into horizontal and vertical 1D functions. It is also easy to translate from the detector array to an output digital image that is composed of rectangularly spaced pixels. However, if the detector array is not uniformly sampled, or not sampled with the same density as pixels in the produced image, there must be some translation between the source information and desired representation. One of the most common examples of non-rectangular sampling is a hexagonal sampling pattern. A regular hexagonal sampling pattern is the most efficient pattern for alias-free sampling of a radially bandlimited signal (see Chapter 12.3 [20]). A hexagonal pattern can also offer similar efficiency in the arrangement of individual photodetector cells on the detection surface. A detailed examination of hexagonal sampling is provided in [22]. Since a non-rectangular sampling will still require an eventual rectangular representation for standard display, it is necessary for an interpolative conversion step to exist.

**Sampling for Color Signals**

So far all theoretical treatment has focused on a monochrome signal. However, the bulk of modern imaging (especially common consumer applications) uses trichromatic imaging in which three distinct color channels are examined at once. Typically, this more complicated approach is designed based on the presence of three different types of color sensors found in the human eye [14] to better convey
information of the scene. This gives the traditional red-green-blue (RGB) representation. Alternative representations of the RGB scheme such as luminance-chrominance are also commonly used (e.g., YCrCb for coding applications). In the broadest sense, much of the theory can be expanded to general hyper-spectral imaging, which makes use of large numbers of separate channels, many of which fall outside the visible spectrum. While trichromatic signals are typically easily translated into a representation which is directly meaningful under human observation, translation of hyper-spectral imaging to a visually meaningful representation can be a more complicated problem that frequently requires ad hoc approaches. This dissertation focuses on monochromatic and trichromatic signaling, but many of the approaches in modelling and reconstruction can be extended to the general multi-channel case. The remainder of this section will deal specifically with RGB processing.

It is typically assumed that the three channels are each sampled independently and with coincident sampling locations. That is, the same sampling lattice is applied without transformation (e.g., translational shifting) to the red, green, and blue components of the scene. In fact, the entire degradation model of Fig. 2.1 is usually considered applied independently to each of the three channels’ content. Although the content present in the different color channels is frequently highly correlated (in fact, many methods including some of those presented in this dissertation rely on this correlation), it is not necessarily required. Consequently, this gives three distinct sampled representations, $x_R[n_1, n_2]$, $x_G[n_1, n_2]$, and $x_B[n_1, n_2]$, each of which is obtained by applying (2.22) to the respective color dimension.

Unfortunately, obtaining such a representation directly can be costly. A three-component sampling of a single pixel requires three separated photodetector cells with standard technology. Directly obtaining three samples of a specific location on a scene requires a complicated and costly acquisition process of splitting the incoming light onto three different sensor locations. The approach is also more susceptible to alignment errors (e.g., sub-pixel shifts), which can require post-acquisition digital processing to correct. The standard method for obtaining color images uses distinct non-overlapping sampling locations for each of the three color
components. This is done by assigning each of the cells of a single photodetector to acquire only one of the color components. A light filter is then applied prior to the detector cell such that only light of the designated wavelength passes through. Consequently, the initial unprocessed representations of $x_R[n_1, n_2]$, $x_G[n_1, n_2]$, and $x_B[n_1, n_2]$ are all sparsely sampled, and only one of the three is able to be non-zero for a specific $[n_1, n_2]$. The color designation of the individual cells is typically provided by a small periodic pattern, such as the Bayer filtering pattern [23] which is $[2, 2]$-periodic and displayed in Fig. 2.3. This particular mosaic pattern (and many similar ones) provides a higher proportion of green samples than red or blue, and green filtered sensors were designated by Bayer to provide the “luminance” component while the others provide two “chrominance” components.

![Bayer color filter mosaicing pattern](image)

Figure 2.3: Bayer color filter mosaicing pattern, representing the color of light admitted through the filter to individual CCD elements. Periodic repetition of this pattern leads to regular sampling in all three color spaces, where the green “luminance” color space receives twice the number of samples as either of the red or blue “chrominance” color spaces.

Using the Bayer pattern, all three color spaces are sub-Nyquist sampled (although less so for green than for red and blue), and some interpolation technique must be used to provide the complete representations of $x_R[n_1, n_2]$, $x_G[n_1, n_2]$, and $x_B[n_1, n_2]$. The problem, frequently referred to as demosaicing, can be distinguished from most traditional interpolation problems due to the fact that significant correlation exists between the samplings of the three color spaces, providing additional information for calculating the missing values. However, the demosaicing problem is not directly considered in this dissertation, and the individual pixels of color images and video are treated as having been directly obtained. While this generally will not introduce significant error, it is important to keep the existence of this interpolation in mind for post-acquisition processing like super-resolution (the
presence of interpolation errors can, for instance, be modelled as part of the noise component. Alternative approaches can be made to work directly with the source data prior to demosaicing, as in [24], where an approach for joint super-resolution and demosaicing was presented.

2.4 Classic Image Restoration

The classic image restoration problem serves as foundation for the general super-resolution problem. The key difference in a purely-restorative problem is that no increase in resolution is sought. Additionally, the restorative problem generally relies on only a single source image, removing the need for registration. Just as there are many different methods available for super-resolution, there are a large variety of methods used in the restorative problem, many of which can be modified for resolution enhancement. There are many alternatives to the types of methods discussed here which can be found in the literature.

Removing the requirement to change the image’s resolution, the effects of sampling can be removed from consideration, leaving the image as degraded through distortion and noise only, or

\[ y[n_1, n_2] = x[n_1, n_2] \ast h[n_1, n_2] + w[n_1, n_2] \]  

(2.29)
in the standard case of LSI distortion and additive noise. It is assumed in this dissertation that \( h[n_1, n_2] \) and the statistics of \( w[n_1, n_2] \) are known or can be estimated. The more complicated blind restoration problem is discussed in [6,25,26].

2.4.1 LMMSE (Wiener) Restoration

One of the most commonly considered approaches for image restoration is linear minimum mean-squared error (LMMSE) filtering, also known as Wiener filtering. The approach requires a second-order statistical model for the uncorrupted image, \( R_{xx}[\tau_1, \tau_2] \), and noise, \( R_{ww}[\tau_1, \tau_2] \), along with the cross-statistics of the two, \( R_{xw}[\tau_1, \tau_2] \). The method seeks to determine a linear filter operating on \( y[n_1, n_2] \) to
produce the $\hat{x}[n_1, n_2]$ minimizing the quantity

$$\text{MSE} = E[|x[n_1, n_2] - \hat{x}[n_1, n_2]|^2]. \quad (2.30)$$

A block diagram representation for this is shown in Fig. 2.4, which will provide a restored output given through

$$\hat{x}[n_1, n_2] = y[n_1, n_2] * f[n_1, n_2]$$
$$\quad = x[n_1, n_2] * h[n_1, n_2] * f[n_1, n_2] + w[n_1, n_2] * f[n_1, n_2]. \quad (2.31)$$

Figure 2.4: LMMSE degradation and restoration block diagram. Signals and filters are labeled for the spatial domain, equivalent frequency domain version is applicable.

The solution is first examined in the Fourier domain, which makes use of the PSDs $S_{xx}(e^{j\omega_1}, e^{j\omega_2})$ and $S_{ww}(e^{j\omega_1}, e^{j\omega_2})$, and the cross spectral density $S_{xw}(e^{j\omega_1}, e^{j\omega_2})$. The optimal reconstruction is provided by

$$\hat{X}_{\text{opt}}(e^{j\omega_1}, e^{j\omega_2}) = X(e^{j\omega_1}, e^{j\omega_2})H(e^{j\omega_1}, e^{j\omega_2})F_{\text{opt}}(e^{j\omega_1}, e^{j\omega_2})$$
$$\quad + W(e^{j\omega_1}, e^{j\omega_2})F_{\text{opt}}(e^{j\omega_1}, e^{j\omega_2}), \quad (2.32)$$

which is equivalent to (2.31). Similarly, frequency domain versions of (2.29) and (2.30) are used. The solution\(^5\) for determining the LMMSE optimal filter is as follows.

From (2.30), the error spectral density (ESD) is given by

$$S_{ee}(e^{j\omega_1}, e^{j\omega_2}) = E[|X(e^{j\omega_1}, e^{j\omega_2}) - \hat{X}(e^{j\omega_1}, e^{j\omega_2})|^2]. \quad (2.33)$$

\(^5\)Note that the approach used here is a non-causal solution. The causal solution (a more complex approach that involves breaking the terms into their causal and anti-causal components [18]), is not especially critical for problems involving finite length images, and is thus ignored here.
Applying (2.32) to (2.33), and hereinafter dropping the \((e^{j\omega_1}, e^{j\omega_2})\) notation for brevity, the ESD is written as

\[ S_{ee} = E[|X - XHF + WF|^2]. \] (2.34)

This quadratic is expanded to obtain

\[ S_{ee} = (1-HF)(1-HF)^*S_{xx} + (1-HF)F^*S_{xw} + F(1-HF)^*S_{wx} + FF^*S_{ww}. \] (2.35)

The optimal solution \(F_{opt}\) minimizing \(S_{ee}\) is found through setting the derivative of (2.35) with respect to \(F\) equal to zero. This derivative is found to be

\[ \frac{\partial S_{ee}}{\partial F} = F^*([H]^2S_{xx} + S_{ww} - 2Re\{HS_{xw}\}) - HS_{xx} + S_{wx}. \] (2.36)

Setting to zero and solving for \(F\) determines the optimum value of

\[ F_{opt} = \frac{H^*S_{xx} - S_{xw}}{|H|^2S_{xx} + S_{ww} - 2Re\{HS_{xw}\}}. \] (2.37)

From Parseval’s theorem, minimizing \(S_{ee}\) for all frequencies is equivalent to minimizing the MSE of a stationary function in the spatial domain, meaning that \(F_{opt}\) optimized per-frequency is optimal for the series as a whole, making (2.37) a valid solution for minimizing (2.30). The signal and noise are often assumed to be statistically independent, leading to the simplified solution

\[ F_{opt} = \frac{H^*S_{xx}}{|H|^2S_{xx} + S_{ww}}. \] (2.38)

The result is also used to determine the optimal ESD by reapplying the solution (2.37) to (2.35). For the case of statistical independence between the signal and noise, the resulting expression is simplified greatly resulting in

\[ S_{ee,opt} = (1-HF_{opt})(1-HF_{opt})^*S_{xx} + F_{opt}F_{opt}^*S_{ww} \]

\[ = \left| 1 - \frac{|H|^2S_{xx}}{|H|^2S_{xx} + S_{ww}} \right|^2 S_{xx} + \frac{|H|^2S_{xx}^2}{(|H|^2S_{xx} + S_{ww})^2} S_{ww} \]

\[ = \frac{S_{xx}S_{ww}}{|H|^2S_{xx} + S_{ww}}. \] (2.39)

A similar, more complicated equivalent can be found in the absence of statistical independence. Applying the expression for \(S_{ee,opt}\) to Parseval’s theorem provides a closed form representation for the minimized MSE. Additionally, examining \(S_{ee,opt}\) in the spectral domain indicates the error power as a function of frequency.
2.4.2 Regularization

Regularization is a technique to solve ill-posed inverse problems. The approach seeks to determine a unique solution to an ill-posed problem by applying some assumption on the characteristics of the solution. Many different forms of regularization exist depending on the assumption used. A commonly used example is Tikhonov regularization [27] which attempts to solve the standard linear system \( Ax = b \) by introducing a penalty on solutions differing from some specified desired characteristics. The approach determines the estimate

\[
\hat{x} = \arg\min_x \left( \|Ax - b\|_2^2 + \|\Gamma x\|_2^2 \right),
\]

a minimization problem giving equal weight to both the solution error and the regularization term (and can be effectively weighted by scaling \( \Gamma \)). Selection of the regularization matrix \( \Gamma \) is used to indicate the desired properties of the solution. For instance, a selection of \( \Gamma \) that amplifies the higher frequencies of the solution (e.g., a Laplacian operator) will result in a lower penalty for smoother functions. Such constraints can also serve some utility in maintaining the stability of an iterative numerical solution approach to the minimization problem. A closed form solution to (2.40) is found through

\[
\hat{x} = (A^T A + \Gamma^T \Gamma)^{-1} A^T b.
\]

Note that when \( \Gamma = 0 \), the problem (2.40) and its solution (2.41) are equivalent to least-squares estimation.

A second regularization example is the total-variation (TV) regularization, which has found application in image restoration [28–31]. This approach relies on the estimator

\[
\hat{x} = \arg\min_x \left( \|Ax - b\|_2^2 + \|\nabla x\|_1 \right),
\]

where \( \nabla \) is the discrete gradient operator [14]. Unlike in the previous case, the regularization term here is based on the \( l_1 \) norm, computed as the sum of magnitudes. One of the properties of TV regularization making it attractive for image processing applications is that it tends to preserve image edges through processing [32]. However, the method is a nonlinear problem requiring an iterative solution.
A more detailed discussion on the topic of regularization for image restoration is provided in Chapter 3.6 of [6]. Among other topics, this source covers additional methods for regularization and practicableness such as parameter selection and iterative implementation. Recent examples of successful regularization solutions for the super-resolution problem are found in [31].

2.4.3 Richardson-Lucy

The Richardson-Lucy (RL) algorithm [33, 34] is an iterative deconvolution algorithm that converges on the maximum-likelihood (ML) solution [35]. The ML solution $\hat{x}$ is the estimate for the unknown $x$ maximizing the probability density function of the observed data $y$ given $x$, or

$$\hat{x} = \arg\max_x p_{y|x}(y|x).$$  \hspace{1cm} (2.43)

The RL algorithm requires a specific, known PSF model for the distortion in order to define the conditional PDF. A generalization of the RL algorithm is found in the expectation-maximization (EM) algorithm [36], which provides an estimation method in the case of unknown parameters affecting observation (e.g., an unknown PSF) or, more generally, missing observations—which is clearly of interest for resolution enhancement problems.

The RL algorithm is most commonly considered for classic image deblurring applications. The known spatially-variant PSF is given by $h_{mn}$, indicating the proportion of the undistorted image intensity at location $n$ becoming a component of the distorted image at location $m$. From $\hat{x}_n^{(k)}$, the current estimate of the undistorted image, the current estimate of the distorted image is given by

$$\hat{y}_m^{(k)} = \sum_n \hat{x}_n^{(k)} h_{mn}. \hspace{1cm} (2.44)$$

Assuming a normalized PSF ($\sum_m h_{mn} = 1 \forall m$), the updated estimate of the undistorted image is found through

$$\hat{x}_n^{(k+1)} = \hat{x}_n^{(k)} \sum_m \left( \frac{y_m}{\hat{y}_m^{(k)}} \right) h_{mn}, \hspace{1cm} (2.45)$$
where $y_m$ is the observed degraded image. The iteration of (2.44) and (2.45) converges on the ML solution. In an ideal noise-free environment, the optimal estimate of $\hat{x}_{n}^{(k)} = x_n$ provides a distortion estimate $\hat{y}_{m}^{(k)} = y_m$, leading to update of $\hat{x}_{n}^{(k+1)} = \hat{x}_{n}^{(k)}$ in (2.45).

Additional discussion of the RL algorithm and the effects of noise, implementation with resolution enhancement, and practical considerations are provided in [35].

### 2.5 Registration

Registration is required to align multiple LR images or frames, essentially providing the locations of all pixel samples on the continuous scene. The registration model is differentiated from the sampling model in that it seeks to relate two or more groups of samples to each other, rather than model the loss of information caused by the sampling process. The registration model then accounts for the motion between two images/frames (either local motion of a temporally changing scene, or overall motion of the acquisition device). Relative to the standard degradation process depicted in Fig. 2.1, the motion model is generally placed prior to the distortion, acting on the original image before any degradation. In many video super-resolution algorithms, this placement of the motion becomes problematic to model, requiring assumptions limiting the form of motion (see Chapter 5). For the still image case, motion between the individual LR images can be assumed to have been caused by camera motion, and can be modelled globally. The presence of local temporal motion existing over multiple frames of video introduces a more complicated registration problem that typically cannot be fully described with global models.

Registration is one of the most critical sub-problems of super-resolution, and accounts for a substantial amount of research apart from the wider enhancement problem. Minor errors in the registration estimate can easily lead to substantial errors in the super-resolved estimate. Whereas errors in the noise and distortion models generally lead to an over- or under-sharpened product, errors
in the registration typically cause severe speckling artifacts of objects within the image, especially at edges or in areas of texture. An example of the effects of registration error on super-resolution is found in Fig. 2.5, where a super-resolution interpolation was performed using some misregistered data. Artifacts in these two examples take the form of hatch marks or checkerboarding. The errors in these examples occur in small block patches due to a block-based registration which was used.

![Figure 2.5: Example effects of registration error.](image)

Unfortunately, error is found to be a result of all practical estimation techniques. Discussion of registration error for image problems is found in [37] and related works. Generally speaking, the levels of error necessary to have a noticeable effect on the reconstruction quality are easily reached, although this does of course depend on the quality and type of registration being performed. Global registration is less susceptible to error than local methods (provided the actual motion is global), since there is a larger portion of data being used to perform the estimate.

The prevalence of registration error and its highly detrimental effect on reconstruction quality, highlights the importance of a robust reconstruction technique. There are a couple solutions commonly employed (additional details are found in later chapters). The first solution is simply to use a directly robust reconstruction algorithm. An example of this can be the regularization method discussed above. The nature of registration artifacts, as illustrated in Fig. 2.5, is to be textured with significant variation between adjacent pixels, which is commonly
out of place on the background of standard content (and typically much less noticeable in the alternative case as it is existing on a similar camouflaging textured content). A regularization approach that assigns a higher cost to such registration artifacts (i.e., both the Tikhonov and TV approaches) will determine a solution reducing their presence. Of course this can also have some detrimental effects of reconstruction of content with similar features as the artifacts. Other approaches also determine a reconstruction with reduced artifacts, e.g., the smoothing response of a noise-reducing Wiener filtering solution. Alternatives such as direct spatial-domain bilinear filtering, which do not provide any guard against erroneous modelling are generally inapplicable for practical super-resolution problems.

The second solution for registration artifact reduction is to make use of an additional processing step. One such possibility is a post-reconstruction step performing artifact detection and removal. In an iterative reconstruction technique, the post-reconstruction step can be introduced in the processing loop, acting as a part of the iterative solution. Another approach, which was used in [38], is to identify portions of the registered content likely to introduce registration artifacts (e.g., content with significant contrast differences from their surroundings) and remove them prior to the reconstruction step. Such additional processing steps are best suited for removing the most significant registration error artifacts.

The remainder of this subsection briefly explores specific registration techniques. This is really a very large topic in itself, and the discussion here only touches on a very limited portion of the total research in registration and motion modelling. The reconstruction approaches appearing later in the dissertation will assume registration information is provided after having been separately estimated.

2.5.1 Global Registration

Global registration is not suitable for all scenarios. The standard model requires multiple images acquired simultaneously or successive images operating on an unmoving scene. In either case, it is assumed that all acquired samples come from a single unchanging 2D continuous scene. Although global registration can be useful for certain portions of dynamic scenes sampled at different times,
there will be portions of content that do not fit within the global model. Global registration of such dynamic scenes can be used to detect regions of local motion, and can be useful for limited restoration and enhancement scenarios. For example, a static background with global camera panning can be restored using a global registration model with the dynamic foreground portions ignored or treated with a secondary motion model. The reconstruction approach in [39] provides an adept method for such a scenario, as it allows for poorly modelled data to be discarded while still improving the result using samples designated as “accurate,” regardless of registration model.

The standard approach used for global motion is the affine model, which relates the pixel locations of the output image $y = [y_h y_v]^T$ to those of the input image $x = [x_h x_v]^T$ via

$$
y = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} x + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = A x + b. \tag{2.46}$$

The components in $b$ provide the translation (horizontal and vertical shifts) of the content, while $A$ indicates the non-shift linear motion component. $A$ is often used to consider scaling and rotation components only. Pure scaling is provided by diagonal $A$, and pure rotation (by angle $\theta$) is given by $a_{11} = a_{22} = \cos(\theta)$ and $a_{12} = -a_{21} = -\sin(\theta)$. Arbitrary $A$ describes a general shear linear transformation.

Selection of the global motion model (2.46) thus requires estimation of six parameters using data from two images. A variety of techniques exist for doing this. Of the frequency domain techniques, a frequently used method determines the translational model through phase plane correlation, through

$$\phi[n_1, n_2] = \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{x[n_1, n_2]\}\mathcal{F}\{y[n_1, n_2]\}}{|\mathcal{F}\{x[n_1, n_2]\}|^2} \right\}, \tag{2.47}$$

the normalized correlation between images $x[n_1, n_2]$ and $y[n_1, n_2]$. Under ideal circumstances, purely translational motion (which must be cyclical for a DFT implementation) results in $\phi[n_1, n_2] = \delta[n_1 - b_1, n_2 - b_2]$, i.e., the Dirac delta function positioned according to the motion vector $b$. Non-ideal circumstances will still
produce $\phi[n_1, n_2]$ containing a dominant peak corresponding to the motion vector. Sub-pixel translational motion can be identified as a spread of the delta function’s energy to multiple spatial indexes, or accommodated through a higher resolution implementation of (2.47). The rotation-scaling components can be determined via a phase plane correlation on a log-polar mapping of the data [40]. Efficient FFT-based methods for estimating motion through the correlation plane technique are provided in [41,42]. An alternative approach for rotation estimation is found in [43], which estimated rotation as a cyclical shift to the function of average signal energy at different angular directions (interestingly, this is similar to the measure (3.10) used for content modelling). Outside the frequency domain approaches, methods are typically based mainly on the spatial domain data, although some form of data reduction is frequently introduced to reduce the abundance of information present in two complete images, for instance the use of tomographic projections in [44].

2.5.2 Local Registration

The distinction between global and local motion modelling is in some cases debatable, for instance use of a global motion mesh (in which each pixel has arbitrary motion potential within the confines of a globally smooth function) or unusual non-linear “global” models which do not conform to simple linear global approaches, e.g.,

$$\begin{bmatrix}
y_h \\
y_v
\end{bmatrix} = \begin{bmatrix}
a_{11}x_h^2 + a_{12}x_v + b_1 \\
a_{22}x_v + b_2
\end{bmatrix}.$$

The proposed reconstruction methods appearing later in this dissertation either require a very specific global model (e.g., pure-translational) or can accommodate arbitrary motion on a per-pixel level. The focus here is on different models that try to account for the different forms of motion found in real-world scenarios. Local motion is most often associated with video scenarios, in which the content changes between frames based on temporal motion, whereas global motion models are better suited for still scenes in which different images are produced via changes in orientation of the acquisition device(s).

The most commonly used model for local motion is a block-based transla-
tional model. This approach has many positive aspects, being simple to understand, generally inexpensive to implement, and used by all video codecs. However, it unfortunately is also incapable of modelling true temporal motion.

The method divides a frame into multiple blocks, then determines a separate two-parameter motion vector (a translational model is assumed) for each block by matching it to an earlier frame. Different criteria can be employed to determine the "best" match, typically an MSE or sum of absolute differences (SAD) is calculated. Matching the block to the best possible choice from the entire frame requires an evaluation at each possible position, requiring a large number of computations per block evaluation. Further, when considering sub-pixel motion estimation, the required number of evaluations increases geometrically with scaling of accuracy. Several approaches have been used to reduce the total computation requirement, however frequently at the potential cost of accuracy. Among the more common approaches for computational reduction are range limitations, i.e., an assumption the block’s absolute motion is limited to a specified maximum, and hierarchical search patterns. An hierarchical search first determines the best block fit over a coarse pattern of possible locations (e.g., at every fourth possible position in each dimension) within a large window area. Subsequently the search is refined to evaluating locations on a finer pattern within a smaller window centered about the best estimate from the prior stage. This process is repeated until the search is operating at the finest search pattern considered by the motion estimation process and the final best fit is determined. This approach is considerably faster than a global exhaustive search, but can clearly miss the true globally optimal point entirely. The method is best suited if the error criterion is a slowly changing function that does not have any local minima that are not global minima. More detailed discussion is found in Chapter 3.10 of [6].

Even in the case of optimal motion vectors, there are still several obvious problems with a block based approach. The most significant of these is the fact that true motion is not block based. Blocks are typically described using a purely translational model which, while certainly of less concern for local motion approximation than global, is unable to describe true motion. In video coding sce-
narios, where block-based motion is most frequently used, a residual error signal is encoded to describe the difference between the motion-predicted frame and the true frame. However, the super-resolution scenario contains no residual information, and partial-block errors can severely reduce reconstruction quality. Another problem with block motion is that the quality of the estimated motion vector is notoriously poor. Methods typically rely on some form of a “best-fit” algorithm to place a block within another frame. While a global search can find the best fit based on some objective criterion (e.g., SAD), individual blocks’ “best fits” often reduce the complete estimation quality (this is further compounded over multiple frames, and can lead to obvious temporal inconsistencies). Some techniques, such as motion vector smoothing [45], can improve this.

Due to many of the deficiencies in commonly used and simple block-based motion modelling, more advanced methods have been developed. One alternative approach is object-based motion [46, 47], which identifies specific objects (with arbitrary perimeter) within a sequence and tracks their motion. While an improvement over the block-based motion, there are still limitations to the method, mainly in describing content that changes over time (rather than simply moves). Models allowing “free-flow” motion [48, 49] are preferable in that the motion of individual pixels can be accounted for, however these approaches do suffer from an increase in difficulty and required computation for their estimation. Such methods work to determine a vector field that describes the per-pixel motion, which can more accurately describe real temporal motion found in video sequences. Ongoing work in motion field research is investigating solutions in some of the more complicated problem scenarios, such as the presence of occlusion.
3 Content Modelling

3.1 Introduction

Image and video content can be modelled in a large variety of ways. However, much more so than the degradation models discussed in the previous chapter, the choice of content model depends on the reconstruction algorithm used. This chapter provides some general discussion of content modelling, but is focused mainly on the models which are of use to the algorithms presented in later chapters.

The purpose of content modelling in this dissertation is to provide a representation for the characteristics of the desired reconstruction to aid in estimation or restoration. The methods can also be beneficial for additional applications such as classification or encoding. For ill-posed super-resolution problems, some form of modelling is necessary, providing some criterion that gives a solution to an otherwise unsolvable problem. For the purposes of the methods presented in this chapter, the degradation models are either known or are irrelevant for constructing a content model.

As mentioned, the specific type of model needed depends on the enhancement algorithm used. In fact, the reconstruction algorithm dictates the form of the model. Any non-trivial reconstruction scenario will require some model for cases where the problem is ill-posed. However, in many cases the use of a content model is less obvious. This chapter defines a distinction between two different forms of content modelling: inherent and specified. This first case, inherent models, take the form of some assumption within the reconstruction process (e.g., a bandlimited signal, stationary random process, regularization term, etc.). While the exact
form of the inherent model can often be adjusted, it is generally specified prior to observation, and the characteristics of the reconstruction conform to it. For this reason inherent models are typically not too explicit, and permit reasonable reconstruction for a large variety of content, without favoring any specific type of content.

In many ways, specified models work in the opposite fashion. They are determined subsequent to observation, based directly on the characteristics of the LR images. For this reason specified models are more costly to employ, perhaps prohibitively so depending on the application. Nevertheless, there are clear advantages to tuning the characteristics of the reconstruction based directly on the observations. There is also a some element of the inherent model in the specified model: the specification is required to fall within the boundaries of some assumed form.

This chapter first examines inherent models, followed by a discussion of specified models, which are the main interest. Focus is primarily on second-order statistical models which are of interest for the reconstruction methods of the later chapters. Two important cases examined are: PSD models, which are used under a stationary signal assumption, and a localized correlation approach, which is non-stationary and can be varied for different regional content. For both these cases, novel models are presented along with discussion of alternative prior approaches.

### 3.2 Inherent Models

An inherent model is defined here as a description of the content characteristics which takes the form of some assumption that makes the reconstruction problem solvable. In all examples presented in this dissertation these models are simply given and assumed to be accurate. Some of these models are certainly superior to others, and the quality of the results they produce depends much on the actual accuracy.

Some examples were already discussed above in the portion on regularization based solutions to the classic restoration problem, although they were not
explicitly identified as such at the time. Consider an image restoration problem which makes use of a Tikhonov regularization (2.40) to find a solution (the non-specific arguments following can be extended to other forms of regularization as well). As it is known that the energy found in the high-frequency components of images is typically much lower than that found in the lower frequencies, a regularization term $\Gamma$ is selected such that $\|\Gamma x\|^2_2$ increases as the image content varies from low-frequency dominated to high-frequency dominated. This regularization term then determines cost along with the error term $\|Ax - b\|^2_2$, where $b$ is the observed degraded image and $A$ is the degradation process acting on the original image. Minimizing the cost with respect to $x$ produces the estimated uncorrupted image, $\hat{x}$.

The quality of the result in this example is highly dependent on the accuracy of the assumption that the desired image is low-frequency energy dominated (and more specifically, that it is so in the way described by $\Gamma$, and not some alternative model penalizing a different form of high-frequency dominance). It is certainly possible to construct pathological design content for which a given regularization method will fail, but an approach can be developed to work reasonably well with standard content. Many approaches have successfully made use of a regularization-based reconstruction [31, 50–52], and it is arguably the most successful of the approaches based on non-specified models. However, even in the best cases, some localized variance in reconstruction quality should be expected with an assumed model. Although a locally adjustable regularization term can be used, this would be considered to fall under the scope of a specified model to be discussed in the next subsection.

While regularization methods are probably the most commonly examined and successful inherent models for image restoration and enhancement problems, there are additional types of models that have been used. Some of the early approaches used for super-resolution relied on a simple bandlimited model, based on the inherent assumption that the original image contained no high-frequency content [3, 53, 54]. In an ideal situation of zero noise and sufficient sampling density (at least equal to the Nyquist rate), these different methods will produce the same
bandlimited reconstruction. In fact, the assumption guarantees the reconstruction result to be bandlimited. While this works well for bandlimited images, practical images are unfortunately not ideally bandlimited and their reconstruction can suffer under such an assumption, most notably containing ringing artifacts (Gibbs phenomena) [55] of the sort illustrated in Fig. 3.1.

![Figure 3.1: Example Gibbs phenomena.](image)

These examples of inherent models, a regularization term and a bandlimited assumption, are just two of many different types of inherent content models that have been employed in super-resolution algorithms. Other approaches include simple spatial domain interpolation approaches that use bilinear or bicubic formula (the HR reconstruction is “inherently” assumed to follow the characteristics of the interpolator). More advanced spatial domain reconstruction approaches like [56] also fall into this category. Other reconstruction methods like projection onto convex sets [57,58] or iterative back projection [59] include an assumed structure to the reconstruction. Several of the reconstruction approaches presented in the next two chapters are also based on some content model assumption.

### 3.3 Specified Models

Unlike the inherent models considered, the specified models are estimated directly from the image content. This typically makes them more accurate than inherent models, but also more costly to obtain and potentially less robust (depending a lot on the form of the model and the approach used to determine it).
The desired characteristics of a specified model and its estimation are:

1. The model should be highly accurate when estimated directly from the original, undegraded image/frame. Frequently, a simplified parametric model is used, allowing the content characteristics to be efficiently represented with a small number of parameters. This can be useful for model encoding purposes and is often necessary for cases where the model must be estimated from the degraded content (in which case the required information is lacking from a purely nonparametric modelling). If a parametric model is used, it must accurately provide the information required for the restoration/enhancement algorithm. Some performance loss is expected as the model is simplified, but too much loss makes the model ineffective.

2. The estimation process should be robust in the presence of degradation. In the practical case the model will need to be estimated from the degraded image/frame (or set of images/frame sequence). However, the desired model is intended to portray the characteristics of the original, undegraded image/frame. There are a couple different approaches that can be taken to resolve this problem. Knowledge of the degradation process is required, to the extent of having a general model for how the original data is corrupted. The standard distortion-sampling-noise degradation model (Fig. 2.1) is used in this work.

One approach for recovering a model of the uncorrupted content from the corrupted data is to first determine an estimate for the recovery from which the model can be derived (this approach relies on the fulfilling the prior criterion, that the model estimated from the original content is accurate). Unfortunately this is a somewhat circular approach, since the specified model is desired for the purpose of removing the degradation. Nevertheless, there is some value to this technique, due to the fact that the model does not need to be derived from the best reconstruction (this is especially true for parametric models which lose information of content detail and retain only the broader modelling trends). Consequently, it is possible to first use a poor...
reconstruction process (e.g., an efficient one relying on an inherent model) in order to determine a result from which a specified model is obtained for use with a second reconstruction process that is higher in quality. Depending on the computation constraints, it is also possible to repeat this technique and iteratively refine the model and reconstruction.

The other method for determining the model from the corrupted data is to simply use the degraded content. The principle disadvantage of this approach is that the degradations will contribute to a more significant portion of the model. The effects of this depend largely on the specific types of degradation found in the images and the model being used. Under this approach, the effects of resolution reduction (sampling) will typically have a more detrimental effect than those of noise and distortion. The primary problem with a resolution reduction is in obtaining the high-resolution content model from the low-resolution content. While it is possible (although generally inadvisable) to simply ignore noise and distortion through a direct model estimation, the resolution reduction has to be taken into account so that the recovered model is of the correct resolution. This is one of the more significant problems in modelling for image super-resolution. Several different solution methods for this problem are presented below.

3. The model estimation process should be made computationally efficient. This is often of lesser importance than the other criteria, and depends on the application. However, in almost all practical cases there are computation constraints which favor fast algorithms. Non-iterative approaches and models with relatively low numbers of parameters are favored.

Similarly to registration modelling (Chapter 2.5), specified content models can be categorized as global or local. The advantages and disadvantages of each form are also similar. Global models are generally more cheaply produced as they require a single evaluation for the entirety of content (although this single evaluation does typically cost more than any of the individual local evaluations). Their definition is also more robust against content degradations since there is a
larger portion of content being used to determine the model. In contrast, localized models are relatively cheap to individually estimate due to the small amount of content, but multiple estimations must be performed for the entire image. The estimate is also less robust to degradation since there is less source data. However, the major advantage of localized content models is the fact that they can provide better descriptions of the local content than global models, which are forced to combine the content of the image in a single description. With sufficiently accurate estimation, localized models are able to outperform global models, but specified global models still offer improved performance over inherent models which offer no content-based adjustment.

The focus of the work presented in this subsection is on second-order statistical (SOS) models. This is the most frequently used of specified modelling approaches for super-resolution (as well as other applications, e.g., restoration, enhancement, encoding, etc.). SOS models are favored in image processing applications for some of the same reasons they are in traditional signal processing. There is a long history of theory and applications based on SOS modelling, and the approach is relatively simple, making the method well understood by the image processing community. Computationally, SOS models are relatively inexpensive, especially in the stationary case where efficient FFT techniques can be used. Finally, SOS models are sufficient for providing the required information for MSE-based evaluation, which is a frequently used criterion for signal and image optimization problems (e.g., Wiener filtering discussed in Section 2.4).

Organization of this topic is as follows. Global SOS modelling techniques are presented in Section 3.3.1 and local models are presented in 3.3.2. Both these sections contain original work on models and model estimation techniques which are used for the super-resolution techniques presented in later chapters. Alternative SOS models and methods for their estimation are also presented.

3.3.1 Global Correlation and PSD Modelling

Discussion here deals strictly with discrete-space signals. If needed, functionally equivalent continuous-space versions can be determined using similar pro-
cedures. The first- and second-order statistics of a stationary 2D signal $x[n_1, n_2]$ are given by equations analogous to (2.14) and (2.15), providing $\mu_x$ and $R_{xx}[\tau_1, \tau_2]$. Without loss of generality, it is herein assumed that the means of all considered signals are zero. The PSD of a stationary signal is provided through the discrete-space Fourier transform of $R_{xx}[\tau_1, \tau_2]$, giving $S_{xx}(e^{j\omega_1}, e^{j\omega_2})$. From the properties of correlation and the Fourier transform, the autocorrelation of a real signal is real and even symmetric and its PSD is non-negative and even symmetric.

In the context of digital images, the frequently used term “PSD” is a misnomer, as the practical image is a deterministic signal and not a random process. For a deterministic signal, the more appropriate term is energy spectral density (ESD), which turns out to be in many ways functionally equivalent to the PSD. In the context of this work, it is convenient to treat an image’s ESD as a PSD and assume the signal is random. The most significant problem with this assumption is that the signal is actually not stationary, a fact which can be readily observed in many images by noting the very different types of content existing in different regions (e.g., smooth vs. textured). Reconstruction formulas relying on the assumption of stationarity will have locally variable performance levels depending on the local content features. The best solution for this problem is to abandon the stationary model entirely and instead rely on a localized correlation model, as discussed in Section 3.3.2. However, in spite of this weakness, treating the deterministic signal as stationary and the ESD as a PSD can be useful as a reconstruction tool.

As shown in [60], the ESD of a signal is given by

$$
\bar{S}_{xx}(e^{j\omega}) = |X(e^{j\omega})|^2,
$$

i.e., the squared magnitude of $X(e^{j\omega})$, the discrete-space Fourier transform of the deterministic signal $x[n]$. Single dimensional signals are used here for brevity; extension to 2D signals is straightforward. The bar notation is used here to dis-
tistinguish the ESD from the PSD. The ESD (3.1) can be decomposed into

\[ \bar{S}_{xx}(e^{j\omega}) = X(e^{j\omega})X^*(e^{j\omega}) = \left( \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \right) \left( \sum_{m=-\infty}^{\infty} x^*[m]e^{j\omega m} \right) = \sum_{n=-\infty}^{\infty} x[n] \sum_{\tau=-\infty}^{\infty} x^*[n - \tau] e^{j\omega(n - \tau)} \]

using the substitution \( m = n - \tau \). Thus, \( \bar{S}_{xx}(e^{j\omega}) \) is the discrete-space Fourier transform of \( \bar{R}_{xx}[\tau] = \sum_{n=-\infty}^{\infty} x[n]x^*[n - \tau] \),

which can be viewed as a deterministic autocorrelation of \( x[n] \). For the case of finite length signals (such as images), the summations over \( n \) and \( m \) in (3.2) can be truncated to include only the non-zero elements. Interestingly, this truncated definition of the ESD is identical through scaling to one of the standard PSD estimates: the periodogram \([60,61]\). The length-\( N \) periodogram spectral estimate is defined as

\[ \hat{S}_{xx}(e^{j\omega}) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n]e^{-j\omega n} \right|^2, \]

which is exactly the definition of the ESD of a length-\( N \) deterministic signal, excluding scaling (which is of little relevance in most applications, so long as all components are scaled proportionally). Similarly, the correlogram \([62]\) spectral estimate,

\[ \hat{S}_{xx}(e^{j\omega}) = \sum_{\tau=-(N-1)}^{(N-1)} \hat{R}_{xx}[\tau]e^{-j\omega\tau}, \]

shows an equivalent scaled form to (3.2), where \( \hat{R}_{xx}[\tau] \) is the biased estimate of the autocorrelation sequence given by

\[ \hat{R}_{xx}[\tau] = \frac{1}{N} \sum_{n=-(N-1)}^{N-1} x[n]x^*[n - \tau], \]
which is a scaled equivalent to (3.3) for length-$N$ sequences. Finally, note that a DFT equivalent can be constructed under the assumption of signal periodicity.

The fact that these relatively simple PSD estimates are functionally equivalent to the definition of the ESD helps to justify the treatment of deterministic signals as though they are stationary random signals. From the standpoint of spectral estimation, the periodogram and correlogram are the simplest of the non-parametric methods. More advanced methods such as the windowed periodogram or the Blackman-Tukey [60] estimator are also applicable. Applied to images, these methods provide a means for determining a detailed non-parametric spectral model which can be considered the “ground truth.” Parametric models will typically be less detailed but can also provide more utility, as will be discussed in later portions of this section.

**Boundary Effects**

When dealing with images, some extra consideration must be taken to account for the boundary effects. Consider a discrete spectral estimate obtained using the DFT of a given image. That is, a 2D discrete form of (3.1) using (2.5) to determine the non-parametric spectrum

\[
S_{xx}[k_1, k_2] = |X[k_1, k_2]|^2.
\]  

Since the magnitude of the DFT is invariant with circular shifts (i.e., shifts maintaining the periodicity of $x[n_1, n_2]$), this spectral estimate can contain artifacts caused by cross-boundary discontinuities. Specifically, a contrast mismatch between the top and bottom borders or between the left and right borders will produce in the PSD estimate the same sort of features as a horizontal and vertical edge features. To illustrate, consider the two images shown in Fig. 3.2, in which the original image (a) was subjected to a circular shift to produce (b). As these images produce the same periodogram estimate, the obvious discontinuities that contribute to the spectrum of (b) affect that of (a) as well.

The specific effects of a sudden jump in contrast across the boundary depend on the image content. For example, a large difference in contrast between top
Figure 3.2: Circular shift equivalent images.

and bottom borders of an image containing a significant number of horizontal edges is less detrimental to the estimate since the artifact will camouflage in the content. The effects are also highly determined by the content existing at the borders. However, the typical effect is that a horizontal edge induces a vertical discontinuity, contributing to a magnitude increase of frequencies along the vertical axis of the spectrum (and vice versa for vertical edges). This problem is illustrated by two example images shown in Fig. 3.3. The intensities of both images change sinusoidally with the horizontal index and are constant with respect to the vertical index. For an infinite-span continuous 2D signal with this property, the spectrum will contain two Dirac delta functions on the horizontal axis, corresponding to the frequency of the sinusoidal intensity change (a DC term will also exist if the content is not zero-mean). However, this does not necessarily carry through to the DFT-based spectrum due to sampling and windowing. Thus, while the horizontal amplitude function in Fig. 3.3(a) governing the image shown in Fig. 3.3(b) produces the “expected” spectrum in Fig. 3.3(c) via the standard periodogram estimate, the slight change in sinusoidal frequency of the amplitude function found in Fig. 3.3(d) and image in Fig. 3.3(e) produces the “unexpected” spectrum in Fig. 3.3(f). Both images contain a very small level of white noise, included so that the magnitude of the spectrum is always non-zero and can therefore be viewed at a logarithmic
scale as in (c) and (f). What is seen in Fig. 3.3(f) is that the entire horizontal axis of the spectrum contains significant magnitude levels (which decay further from the origin). Although this is a synthetic example, this spectral artifact is found in standard images as well. The cause of this is the discontinuity found in the periodic repetition of Fig. 3.3(d), which is absent in Fig. 3.3(a), illustrating how windowing can have very different effects on different frequencies. Apart from the horizontal axis, the spectrum is entirely unaffected in this simple case (since the images are constant in the vertical dimension). In a standard image, artifacts should be expected on both axes.

![Figure 3.3](image-url)

Figure 3.3: Example axis artifacts. Two windowed sinusoid functions of different frequencies (a,d) dictate the horizontal amplitudes in images (b,e) leading to very different periodogram estimated spectra (c,f).

Fortunately, this problem is relatively simple to correct once identified. The solution simply seeks to remove the sharp difference in cross-border contrast. Three approaches for this are considered here: removal with interpolation, windowing, and boundary blurring.

The first approach, removal with interpolation, is very straightforward.
Simply, the values of $S_{xx}[k_1, k_2]$ on the axes are removed from the estimate. The missing values are then interpolated from the neighboring portions of the spectrum. Although this is highly effective in removing the axis artifacts, it also removes non-artifact portions of the spectrum. Using an interpolation to replace these portions is not necessarily fully effective for recovering the original. This is especially true in many practical images which, being framed by humans, often contain many natural horizontal and vertical edges. This correction, in removing the artificial edges that are caused by the boundaries, also removes these desired edges. Thus, while direct and effective at artifact removal, the removal with interpolation approach is less desirable than other approaches which seek to confine the areas targeted by artifact removal to the image borders.

The windowing method first pre-multiplies the image $x[n_1, n_2]$ with the function $w[n_1, n_2]$ reducing the intensity of the image at regions further from the center. Nearly all standard window designs [5, 60] have a peak value of one at the center and decay symmetrically as distance from the origin increases. The resulting spectral estimate is then the (circular) convolution in the spectral domain between $S_{xx}[k_1, k_2]$ and $S_{ww}[k_1, k_2] = |w[n_1, n_2]|^2$. In the spatial domain, selecting a window that tapers toward zero at the borders provides a smooth transition for the borders of the product $x[n_1, n_2]w[n_1, n_2]$. Windows that do not taper toward zero (e.g., low-$\beta$ valued Kaiser windows [5]) should be avoided for this application. In the spectral domain, the features of the unadulterated PSD estimate will be distorted from convolution with $S_{ww}[k_1, k_2]$, but the axis artifacts will be highly mitigated–essentially spread through the convolution to the neighboring frequencies. A relatively smooth $w[n_1, n_2]$ is required since a sharp drop from 1 to 0 in the spatial domain leads to unacceptable levels of spread in the spectral domain convolution term (via increases in sidelobe prominence). However, such a smooth window does have the undesirable consequence of significantly diminishing the contribution of content located near the image edges, since the contrast there is largely flattened. This does not present much of a problem for stationary signal PSD modelling, since the statistics at the mostly-retained center of the windowed observation are identical to the statistics at the mostly-removed edges. It does
however have a highly detrimental effect on a deterministic signal’s ESD model, the features of which vary locally. This concern can be eliminated by performing multiple estimates, each circularly shifted by a different amount prior to windowing, and averaging the results together to obtain a final estimate in which all portions of the image are given equal weighting under all portions of the window (this approach is in some ways similar to the Welch spectral estimate [63], which relies on overlapping data segments).

The final method for reducing border contrast artifacts is to apply a spatially variant circular convolution to the image prior to spectral estimation. Although this approach lacks the convenient mathematical precision of windowing, it is more attractive in that it better retains the features of the original data, without resorting to multiple sub-estimates. The principle of the approach is as follows. Since the axis artifact is caused by a potentially artificial cross-border contrast jump, the jump is reduced by blurring the image so that there is a smooth transition between the boundaries. The extreme example of a spatially invariant blur leads to a corresponding Fourier domain transfer function which (for typical blurs) acts as a low pass filter on the $S_{kk}[k_1, k_2]$. This significantly reduces the axis artifacts, but also degrades many of the high frequencies of the entire spectral estimate. It also unnecessarily targets the center regions of the image. The obvious solution for this is to confine the region of blurring to the image borders. This retains the original content of the central portions of the image for the estimate, but the boundary transition between the blurred and unaffected regions retains some of the same types of artifacts that the preprocessing is trying to remove (although of reduced significance). To further improve the data for spectral estimation, the level of blurring is slowly reduced from a maximal cross-boundary distortion at the border pixels to a lesser distortion some distance in from the border (e.g., 10% of the way toward the center), after which the blurring is nonexistent. This can be viewed as a form of windowing, in which the window is applied to the spread of the distortion (e.g., variance $\sigma^2$ for a Gaussian blur) rather than to the data. The Matlab function edgetaper uses a modification producing a linear weighting of the original and blurred versions in which the edge pixels of the processed image
are dominated by the blurred version and the central pixels are dominated by the original, with a smooth transition of the proportion of each version used from the edge to center.

The application of these methods for artifact reduction are shown in Fig. 3.4. The problematic image from Fig. 3.3(e) is used. Application of the windowing method is used to produce the image in Fig. 3.4(a) which leads to the periodogram spectral estimate in Fig. 3.4(b). The window used here is a 2D separable Hamming window [5]. The Matlab edgetaper implementation for boundary blurring is used to produce Fig. 3.4(c), which provides the spectral estimate shown in Fig. 3.4(d). Both approaches reduce the axis artifact from Fig. 3.3(f). Some problems with the approaches are an obvious spreading of spectral content from the ideal case, although the content used in this example is an extreme case and similar effects are greatly reduced for practical images.

![Figure 3.4: Example axis artifact reduction. Image from Fig. 3.3(e) is modified via Hamming window (a) and boundary blurring (c) to produce respective periodogram estimates in (b) and (d).](image-url)
One of the principal goals of the presented preprocessing techniques has been to retain as much of the image content as possible, while still removing the artifacts. The main problem involved here is that the preprocessing necessarily degrades the content. A modification to the above approaches is to extend the span of the image with artificially generated content. The purpose of such an extension is to keep the entire image in the mostly unaffected central portion, while the mostly degraded boundary transition content is composed of pixels not actually part of the true image. While this can be an effective technique in some cases, there are two main problems with adding content: it changes the size of the image and it introduces artificial content to the estimate.

Changing the size of the image is not a direct problem for the estimation process, but it can introduce problems in later reconstruction stages which make use of the estimate. Practical implementations will use a DFT-based estimate, providing a sampled spectral representation from the sampled image. Increasing the size of the image by introducing extra samples beyond the original image will produce a corresponding increase to the number of samples in the estimate $S_{kk}[k_1, k_2]$. However, since the sampling rate in the spatial domain is not being changed, the extra spectral samples are not padded onto the existing ones. Instead, the effective sampling rate of the spectrum is adjusted\(^1\). The implication of this is that, except for the effects of adding synthetic data to increase the image size, there is no additional information added to the spectral estimate and it is merely the representation of the data which is changed (and represented less efficiently). The main problem with this is that for later DFT based processing, it is extremely inconvenient to have the spectrum sampled at a different rate than the image. This requires either the spectral estimate to be resampled at a lower rate (likely reducing its accuracy) or the image to be padded with extra synthetic content to match the size. If the image is padded, there can be additional difficulties in the implementation for resolution enhancement applications, since the padding of data in the low resolution must be made to correspond to that of a high-resolution

\(^1\)As consequence of Fourier duality it is shown that, for either domain, increasing the span over which the samples exist at a fixed rate results in an increase to the sampling rate over a fixed span in the other domain.
spectral estimate. However, these difficulties are by no means insurmountable and effective synthetic padding solution is relatively easy to construct.

The second problem with adding content for spectral estimation is simply that the added content does modify the estimate. Since the intended reason for adding content is to improve the quality of the estimate, the real test of utility is to determine if the improvement provided from reducing image degradation is greater than any reduction in spectral accuracy caused by adding the synthetic data. To avoid introducing additional artifacts, the synthetic pixels should all be based on the image pixel they are located near, and should strive not to introduce significant false structure. A common method for synthesizing the padded content is to extend the known image symmetrically along the border [64]. However, this does introduce some false correlation between pixels across the boundary and can also falsely replicate diagonal edges. If an initial simple correlation model is first estimated, it can be used to predict the pixels. Techniques borrowed from inpainting can also be used [65]. In general, the fact that the synthetic image portions are largely made insignificant by the boundary artifact preprocess means that there is not a significant effect. Some level of estimation error will remain, but it is small. As discussed later on, using a parametric model derived from the non-parametric estimate tends to lose much of the detail. If the non-parametric spectrum is used for this purpose, the main goal should be to avoid major artifacts which might carry through to the parametric estimate.

**Standard Characteristics of Image Spectra**

With a non-parametric spectral estimation technique in place, it is possible to examine the characteristics of typical image spectra. The eventual goal is to develop general parametric models for image spectra based on the common features found. The main focus is on standard optical digital images. The observations made here should not be considered valid for different classes of images (or general 2D signals). For instance, the spectra examined in certain multi-dimensional array signalling problems [66] should be expected to vary so significantly from those discussed here that an entirely different basis for parametric modelling is required.
(see later comments on auto-regressive spectral models). However, the spectra of certain non-optical images may be categorized as those discussed here. For instance, many sonar or ultrasound and medical imaging applications will produce content with similar spatial features as common optical image, and can be considered under this framework. The validity of these observations is highly dependent on content (which in turn is most often dependent on the application), meaning that in some cases there can be optical images that do not fit well with these observations. However, aside from artificial pathological scenarios, most common images are good candidates for the following discussion.

To illustrate the following discussion, two example images are provided in Fig. 3.5, along with corresponding spectra in Fig. 3.6. The spectra are displayed on a logarithmic scale in order to be visually meaningful, since the $S_{xx}(e^{j\omega_1}, e^{j\omega_2})$ experiences such significant decay as to drop several orders of magnitude over a short range (for this reason, all spectra in this work will be examined on a log scale so that their details are perceptible). Based on these examples, along with many other examined images, it has been noted [38,67,68] that the image spectra tend to decay radially from a peak value at the origin. The dominance of radial features provides motivation for examining the spectra in polar coordinate representation, denoted as $S_{xx}(\Omega_r, \Omega_\theta)$ in the continuous case. A polar representation can be examined in the discrete case as well, but the rectangular sampling pattern will induce a change in the span of support on $\Omega_r$ with varying $\Omega_\theta$, meaning the bandlimiting condition for alias-free sampling requires that

$$S_{xx}(\Omega_r, \Omega_\theta) = 0 \forall \Omega_r > \frac{\pi}{\min(|\cos(\Omega_\theta)|, |\sin(\Omega_\theta)|)}; \quad (3.8)$$

assuming a normalized rectangular sampling frequency. The polar representation is used here to simplify the later modelling process.

Both spectra in Fig. 3.6 display a trend of radial decay with a peak value at $S_{xx}(0, 0)$. In both cases it can be noted that the rate of radial decay varies as a function of $\Omega_\theta$. The angular function of decay observed in the spectrum of the Lena image (a) tends to vary relatively slowly. In contrast, the decay observed in the Pentagon image (b) has several noticeable spikes of slower decay overlayed on an
Figure 3.5: Example images for discussion. Pentagon image obtained from USC database [69].

Figure 3.6: Spectra of images in Fig. 3.5.
otherwise relatively smooth function. These features, referred to as “spectral rays,”
correspond to the straight line features found in the architecture of the Pentagon.
Straight lines existing in the spatial domain at a given angle will produce a spectral
ray on a perpendicular direction. Similar spectral features are commonly found
in many images, due to the prevalence of straight lines in human created objects,
which motivates one of the parametric spectral models to be introduced. Other
interesting spectral features can be occasionally found in some images, but the most
common feature are the spectral rays (existing in images with significant straight
line features) and the underlying smoother radial decay (found in all standard
images).

Further investigation of the spectral features is found by separating the func-
tion $S_{xx}(\Omega_r, \Omega_\theta)$ into weighted angular and radial averages. The log$_{10}$ weighting is
maintained, again so that the information contained in the calculated averages is
visually meaningful. Denoting a spectrum meeting the bandlimited criterion (3.8)
as $S_{BL}(\Omega_r, \Omega_\theta)$, the radial function is determined as the weighted angular average
\[
\bar{S}(\Omega_r) = \frac{1}{\pi \Omega_r} \int_0^\pi \log_{10}(S_{BL}(\Omega_r, \Omega_\theta)) d\Omega_\theta, \tag{3.9}
\]
and the angular function as the weighted radial average
\[
\bar{S}(\Omega_\theta) = \int_0^\pi \log_{10}(S_{BL}(\Omega_r, \Omega_\theta)) d\Omega_r. \tag{3.10}
\]
The symmetry of the PSD makes it only necessary to consider (3.9) and (3.10)
over half the angular span. Both computations consider only the spectral
components falling inside the $\Omega_r < \pi$. At higher radial frequencies $S_{BL}(\Omega_r, \Omega_\theta)$ is
guaranteed to be zero in some angular directions according to the (3.8) criterion,
and will lead to the inclusion of components outside the normalized period in
the sampled case. Note that the measurements do not include the “corners” of
$S_{BL}(\Omega_r, \Omega_\theta)$, excluding $(4 - \pi)/4 \approx 21.46\%$ of the area from (3.9) and (3.10).
However, the energy contained in these corner regions is very small compared to
the total energy due to the steep radial decay or image spectra (e.g., in Fig. 3.6
accounting for only approximately .008\% (a) and .039\% (b) of the total energy).
It is possible to modify the measures to include these corner regions but this will
lead to virtually no change in $\bar{S}(\Omega_\theta)$, and a frequently inaccurate $\bar{S}(\Omega_r)$ (due to low signal-to-noise ratio (SNR) and reduced amount of data). Additionally, later parametric modelling techniques relying on these calculations do not need the additional data.

Numerical computation of (3.9) and (3.10) is determined from a DFT-based non-parametric spectral estimate. Since the standard DFT provides a rectangular sampling, the spectral function is first resampled onto a polar mesh, giving uniformly spaced samples on $\Omega_r$ and $\Omega_\theta$, so that (3.9) and (3.10) can be approximated simply through averaging (more advanced numerical integration techniques can be applied, but are deemed unnecessary for the applications in this dissertation). The examples from Fig. 3.6 produce the $\bar{S}(\Omega_r)$ measurements shown in Fig. 3.7 and $\bar{S}(\Omega_\theta)$ measurements shown in Fig. 3.8. Examining the angular function first, there are very pronounced differences between the measure of the Lena image (a), which displays slight variation overlaying a relatively slowly changing curve, and of the pentagon image (b), that contains significant spikes corresponding to the spectral rays (along with lower magnitude spikes from lesser straight-line features). In contrast, there is little difference in the shapes of the decay curves in Fig. 3.7. Some slight variation will be found between images, but a similar decaying function is almost always found. The primary differences can be accounted for mainly as differences in each image’s rate of decay and its initial value at $\bar{S}(\Omega_r = 0)$. More detailed measures are discussed later on, e.g., allowing $\bar{S}(\Omega_r)$ to be measured over a finer angular range to improve accuracy over the global measure.

Estimation from Degraded Content

In restoration and enhancement applications, the purpose of a spectral estimate is to provide a statistical model of the original image in order to improve reconstruction from the degraded observation. The main practical problem with this is that the estimate cannot be estimated from the original, and has to be determined instead from the observations, meaning the degradations are affecting the spectral model. Using the standard degradation model (Fig. 2.1), the spectrum is affected by distortion, noise, and aliasing (sampling). The effects of these
degradations on the spectrum follow from the discussion in Section 5.3. Denoting the original image spectrum as $S_{xx}(e^{j\omega_1}, e^{j\omega_2})$ and the degraded spectrum as $S_{yy}(e^{j\omega_1}, e^{j\omega_2})$, the effects of the degradation are easily determined:

**Distortion** by a linear transfer function $H(e^{j\omega_1}, e^{j\omega_2})$, leads to the relationship

$$S_{yy}(e^{j\omega_1}, e^{j\omega_2}) = S_{xx}(e^{j\omega_1}, e^{j\omega_2})|H(e^{j\omega_1}, e^{j\omega_2})|^2. \quad (3.11)$$

**Additive noise** with a power spectral density $S_{ww}(e^{j\omega_1}, e^{j\omega_2})$ leads to

$$S_{yy}(e^{j\omega_1}, e^{j\omega_2}) = S_{xx}(e^{j\omega_1}, e^{j\omega_2}) + S_{ww}(e^{j\omega_1}, e^{j\omega_2}). \quad (3.12)$$

**Sampling** a continuous signal with spectrum $S_{xx}(\Omega_1, \Omega_2)$ leads to the spectrum

$$S_{yy}(\Omega_1, \Omega_2) = \frac{2\pi}{T_1 T_2} \sum_{k_1 = -\infty}^{\infty} \sum_{k_2 = -\infty}^{\infty} S_{xx}(\Omega_1 - k_1 \Omega_{s1}, \Omega_2 - k_2 \Omega_{s2}), \quad (3.13)$$

which can be normalized in periodicity to provide $S_{yy}(e^{j\omega_1}, e^{j\omega_2})$. 

Figure 3.7: Plot of the radial measurement, given by the weighted angular average of (3.9) for the images of Fig. 3.5.
Theoretically, the effects of additive noise and distortion are easily reversed (provided $|H(e^{j\omega_1}, e^{j\omega_2})|^2$ is free from zeros), allowing $S_{xx}(e^{j\omega_1}, e^{j\omega_2})$ to be recovered from $S_{yy}(e^{j\omega_1}, e^{j\omega_2})$. For a non-bandlimited signal, sampling leads to aliasing which cannot be reversed, requiring some assumption to determine $S_{xx}(e^{j\omega_1}, e^{j\omega_2})$. Practically, there can be problems in reversing the effects of additive noise and distortion as well, mainly due to the fact that $S_{yy}(e^{j\omega_1}, e^{j\omega_2})$ is provided from a non-ideal estimation process. As such, the “obvious” choices for recovery, $S_{xx}(e^{j\omega_1}, e^{j\omega_2}) = S_{yy}(e^{j\omega_1}, e^{j\omega_2})/|H(e^{j\omega_1}, e^{j\omega_2})|^2$ and $S_{xx}(e^{j\omega_1}, e^{j\omega_2}) = S_{yy}(e^{j\omega_1}, e^{j\omega_2}) - S_{ww}(e^{j\omega_1}, e^{j\omega_2})$, can lead to unlikely or impossible results. For instance, the case of an observation $S_{yy}(e^{j\omega_1}, e^{j\omega_2}) < S_{ww}(e^{j\omega_1}, e^{j\omega_2})$ would imply an impossible negative value of $S_{xx}(e^{j\omega_1}, e^{j\omega_2})^2$. Similarly, a direct removal of distortion is problematic with values of $|H(e^{j\omega_1}, e^{j\omega_2})|^2$ approaching zero, leading to a poorly conditioned reconstruction that is disastrous when combined with variance in the estimate $S_{yy}(e^{j\omega_1}, e^{j\omega_2})$.

This problem occurs very easily due to the high variance common in spectral estimation, a fact which can be easily verified in simulation.
The effects of sampling are likewise problematic for recovering the undegraded spectrum, posing an ill-posed problem due to the uncertainty of aliasing, which is both theoretically and practically irrecoverable.

Because of these problems, a practical spectral estimate relies on the observations to spectral characteristics made above, which can be used as a guideline for shaping the undegraded spectrum in the face of an otherwise ill-posed and ill-conditioned problem. Note first that the dominant characteristic of typical image spectra was shown to be a sharp radial decay. This implies, as was confirmed in examples above, that the lowpass energy greatly outweighs the energy of the rest of the spectrum (this is true under nearly any reasonable definition of lowpass). Fortunately, the lowpass portions are also highly resilient against typical degradations. For the case of distortion, typical transfer functions act as lowpass filters, meaning \( |H(e^{j\omega_1}, e^{j\omega_2})| \) is close to 1 for the lower frequencies, leaving them relatively unaltered. Moving away from the origin, the transfer function dampens the higher frequencies more significantly (this general description is most applicable to lens or atmospheric blurring, or functions modelled with a Gaussian filter; the effects of motion blur can be different as shown in Chapter 2.3.1). Additive noise does not significantly affect the lowpass portion of the image spectrum because its power in these regions is much lower than the signal power. For example, a unit variance white Gaussian noise field with size equal to that of the images in Fig. 3.5 will have a numerically calculated \( \hat{S}(\Omega_r) \) measure (3.9) that is roughly flat over \( \Omega_r \) at a value of about 5.2, well below the signal power for all frequencies. Increasing the variance \( (\sigma^2 = 100) \) gives a measure \( \hat{S}(\Omega_r) \approx 7.2 \), which is about the value of the signal’s measure in Fig. 3.9(a) at \( \Omega_r = \pi/2 \). This corresponds to a relatively significant level of noise, as illustrated in Fig. 3.9. Even these higher levels of noise will have little effect on the lower frequencies of \( S_{xx}(e^{j\omega_1}, e^{j\omega_2}) \) (for example, in Fig. 3.9(a), the measure is an order of magnitude higher than the \( \sigma^2 = 100 \) noise for all \( \Omega_r < \pi/4 \)). However, due to a low SNR in the highpass regions, the noise does still have a detrimental effect on the higher frequencies.

Finally, it can be demonstrated that aliasing does not significantly alter the lowpass frequencies of the estimated spectrum. As in the case of additive
noise, this is based on the low-power corruption having little effect when applied to high power corruption. The aliasing is determined according to (3.13), giving a $S_{yy}(\Omega_1, \Omega_2)$ that is periodic with period $(\Omega_{s1}, \Omega_{s2})$. Assuming the original spectrum decays radially, the dominant portion of periodic $S_{yy}(\Omega_1, \Omega_2)$ is the baseband of $S_{xx}(\Omega_1, \Omega_2)$, i.e., frequencies in the rectangular support region $([-\Omega_{s1}/2, \Omega_{s1}/2], [-\Omega_{s2}/2, \Omega_{s2}/2])$. This quality of this assumption does vary somewhat based on location in the region, being most valid near the origin and entirely invalid at the edges of the support region. To illustrate the problem near the edges of the region, consider that frequency $S_{xx}(\Omega_{s1}/2 - \epsilon, 0)$ is corrupted with $S_{xx}(-\Omega_{s1}/2 - \epsilon, 0) = S_{xx}(\Omega_{s1}/2 + \epsilon, 0)$ (in addition to many other aliasing components). Expecting $S_{xx}(\Omega_{s1}/2 + \epsilon, 0) \approx S_{xx}(\Omega_{s1}/2 - \epsilon, 0)$ for sufficiently small $\epsilon$, the sampled spectrum $S_{yy}(\Omega_{s1}/2 - \epsilon, 0)$ is clearly not dominated by the baseband. However, in contrast, for sufficiently large $(\Omega_{s1}, \Omega_{s2})$ relative to the decay of $S_{xx}(\Omega_1, \Omega_2)$, it can be assumed that

$$S_{xx}(0, 0) \gg S_{xx}(k_1\Omega_{s1}, k_2\Omega_{s2}) \quad \forall \quad k_1, k_2 \neq 0,$$  

(3.14)

implying $S_{yy}(\Omega_1, \Omega_2) \approx S_{xx}(\Omega_1, \Omega_2)$ near the origin. The validity of this assumption is increased if anti-aliasing filtering is employed prior to sampling.

One important exception in the case of images is sampling of a signal with spectral rays (e.g., the spectra in Fig. 3.6 (b)). Without adequate anti-aliasing
filtering, spectral rays outside the \([-\Omega s_1/2, \Omega s_1/2], [-\Omega s_2/2, \Omega s_2/2]\) baseband can dominate portions of the baseband through aliasing. This is shown in the example cameraman image and its spectrum in Fig. 3.10. Fortunately, the aliased spectral rays do not have a significant impact on the measurements (3.9) and (3.10) (with the exception of spectral rays along the axis), allowing these averages to better retain their functionality when applied to undersampled images.

![Figure 3.10: Cameraman image (a) and spectrum (b) which demonstrates aliasing of spectra ray features with insufficient anti-aliasing filtering.](image)

The main conclusion of this discussion is that while various forms of degradation are detrimental to the higher frequencies of \(S_{yy}(\Omega_1, \Omega_2)\), they all tend to leave the lower frequencies relatively unaffected. This will be an important result for the remainder of the section, allowing the parametric spectral models to be constructed from degraded images based on their non-parametric spectral estimates. A model of the entire spectrum of an original image can be obtained from its degraded observation, provided that:

1. the low-frequency content can be accurately estimated and portrayed.
2. the high-frequency content can be predicted from low-frequency content with reasonable accuracy.

The result of this section demonstrates that the first of these conditions is possible. Using this in combination with the prior observations on image spectral

---

Interestingly, the image used in this example—the widely available standard cameraman image measuring 256 × 256 pixels (found with MATLAB software package and included in many other image databases)—is unaltered, bringing into question the quality of anti-aliasing used to produce this standard.
characteristics, which showed a relation trend between low- and high-frequency component using the polar coordinate averages (3.9) and (3.10), allows the second condition to be achieved. The realization of this is provided in the next section using novel parametric spectral models, the parameters of which are adjusted based on the non-parametric spectral model of the observed image.

**Parametric Modelling for Image Spectra**

The examination of parametric spectral models will begin with prior work before moving onto the new results. Early 2D parametric spectral modelling was based on modification of traditional 1D techniques. The principal problem with these approaches was that image spectra do not conform well to the parametric models (although some of these approaches can be very useful for other 2D spectra applications, e.g., array processing). A secondary problem involves the high computation involved in accurate multidimensional spectral processing, which can make some of the approaches unattractive. New methods presented here attempt to improve the modelling situation for images by providing parametric models specifically designed based on standard image spectral characteristics. Prior approaches are described first.

Some discussion in [11] examines some approaches which are most appropriate for scenarios with limited data. These approaches include the maximum likelihood method [70], and the maximum entropy method\(^4\). While these methods are of use in beamforming applications [66], the models do not fit the characteristics of images. Additionally, the abundance of information available in image applications (even highly degraded images contain a huge number of pixels in comparison to typical sensor arrays) removes the necessity of using these methods. Similar statements hold for methods based on line spectral assumptions (signals composed purely of complex sinusoidal components) [60].

The most common parametric spectral model is the auto-regressive moving-average (ARMA) model [60]. The ARMA model is usually considered for 1D:\(^4\)Note that, unlike for 1D signals, the lack of a closed form solution in the 2D means the maximum entropy method is not equivalent to an auto-regressive model estimate [11].
signals, but discussed here from the 2D standpoint. The spectrum is given by passing white noise to a rational transfer function, which can be represented by the Z-transform $H(z_1, z_2) = B(z_1, z_2)/A(z_1, z_2)$ [11]. The spectrum is then given by the transfer function

$$S_{xx}(e^{j\omega_1}, e^{j\omega_2}) = \sigma^2 \left| \frac{B(e^{j\omega_1}, e^{j\omega_2})}{A(e^{j\omega_1}, e^{j\omega_2})} \right|^2,$$

(3.15)

where $\sigma^2$ denotes the power of the input noise. An auto-regressive model (a simplified form of (3.15) where $B(z_1, z_2) = 1$) has been used in certain image spectral modelling applications. Even without the numerator term, selection of the coefficients in $A(z_1, z_2)$ can be problematic. As demonstrated in [11], the problem is expressed as

$$x[n_1, n_2] = -\sum\sum_{(k_1, k_2) \in A} a[k_1, k_2]x[n_1 - k_1, n_2 - k_2] + w[n_1, n_2],$$

(3.16)

where $w[n_1, n_2]$ is the white input process and $A$ is the region of non-zero coefficients in $A(z_1, z_2)$, excluding the tap at $a[0, 0]$, which is normalized to 1. Many of the standard methods for estimating AR parameters in the 1D case (e.g., Yule-Walker [71, 72] or Burg [73] methods) can be easily extended to the 2D case. One of the central problems with AR modelling is the selection of the region $A$. The shape of the estimated spectrum can vary significantly depending on the assumed size of the region. Some reports exist of improved estimation when reconstructions using different choices $A$ are combined, for example through [11]

$$S_{xx}(e^{j\omega_1}, e^{j\omega_2}) = \frac{1}{1/\hat{S}_{xx}^{(1)}(e^{j\omega_1}, e^{j\omega_2}) + 1/\hat{S}_{xx}^{(2)}(e^{j\omega_1}, e^{j\omega_2})}.$$

(3.17)

Although the application of AR models to standard images is of questionable validity and accuracy, they nevertheless have been employed in some applications (e.g., chapter 3.5 of [6]).

**Smooth Decay Model**

More successful approaches determined a spectral model based directly on the features of the image. The early methods [74, 75] for this used a rotationally-
invariant (or isotropic) spectral model based on an observation of isotropic auto-correlation function given by

$$R_{xx}(\tau_1, \tau_2) = \exp \left[ -\alpha (\tau_1^2 + \tau_2^2)^{1/2} \right].$$

(3.18)

The Fourier transform of this auto-correlation function determines the PSD

$$S_{xx}(\Omega_1, \Omega_2) = 2\pi \alpha \left( \alpha^2 + \Omega_1^2 + \Omega_2^2 \right)^{-3/2}.$$  

(3.19)

These expressions are valid for $\alpha > 0$. The polar coordinate representation for (3.19) is

$$S_{xx}(\Omega_r) = 2\pi \alpha \left( \alpha^2 + \Omega_r^2 \right)^{-3/2},$$

(3.20)

which is independent of angle $\Omega_\theta$. For sample image spectra the function can be bandlimited to avoid aliasing concerns. A plot of the function for different values of $\alpha$ is shown on a logarithmic scale in Fig. 3.11. As will be discussed later on, estimated values of $\alpha$ are typically well below 1, leading to the sharp decay curves shown in Fig. 3.11. As observed, for $\alpha \ll \Omega_r$, the denominator of (3.20) is dominated by the radial frequency and $\alpha$ acts primarily as a scaling term. At the lowest frequencies the shape of the curve does vary from the steep decay found at higher frequencies, according to the relative values of $\alpha$ and $\Omega_r$. This is depicted in the detail box in the upper-right of Fig. 3.11. These results are of course scalable on $\Omega_r$ (frequency dilation/contraction) and the entire function (3.20) can include a scaling term. This makes it relatively simple to match a version of the curve (3.20) to the radial function (3.9), e.g., the measured values shown in Fig. 3.7.

Allowing for scaling adjustment, the representation (3.20) requires only two parameters: the scaling term and $\alpha$. Tuning this model for a specific image thus requires optimization over only these two parameters. However, noting the observed characteristics of typical image spectra discussed previously–specifically, the change in radial decay with respect to angular direction–it is clearly advantageous to allow the parametric spectral model to vary as a function of $\Omega_\theta$. This was the approach taken in [38, 68], which led to the non-isotropic function

$$S(\Omega_r, \Omega_\theta) = 2\pi \alpha(\Omega_\theta) \left( \alpha^2(\Omega_\theta) + \Omega_r^2 \right)^{-3/2}.$$  

(3.21)
This modification simply allows the decay parameter $\alpha$ in (3.20) to vary with $\Omega_\theta$. In a parameterized version, the variable $\alpha(\Omega_\theta)$ is simply defined for samples over $[0, \pi)$, and interpolated to provide the continuous 1D function for (3.21). The appropriate number and locations of $\alpha(\Omega_\theta)$ samples is content dependent\(^5\), but a reasonable approximation of the continuous function is possible with a relatively small number of individual parameters. Note that allowing the decay parameter $\alpha$ to change as a function of $\Omega_\theta$ does lead to a discontinuity of $S(\Omega_r, \Omega_\theta)$ at $(0, 0)$, with a changing value based on the direction of approach. The function should not be defined by (3.21) at the origin, and should instead be defined through a separate parameter (determined through direct measurement of the image) or selected based on the average value of $\alpha(\Omega_\theta)$. Note that these alternative selections of $S(0, 0)$ will still be discontinuous at the origin, but the fact that a practical spectrum is sampled removes the concern. The results in this dissertation base $S(0, 0)$ on direct measurements of the image and sample $\alpha(\Omega_\theta)$ uniformly, making use of a cubic spline interpolation to determine the continuous function.

Selection of specific $\alpha(\Omega_\theta)$ samples are based on matching the 1D function

\(^5\text{Although not explored here, one approach that may be worth consideration is to select sample locations based on the angular function } S(\Omega_\theta) \text{ defined in (3.10).}\)
(3.20) to the measurement similar to that defined in (3.9). However, in order to consider only the radial decay on a segment of $\Omega_\theta$, the integral is performed over a small slice of the full spectrum. This measurement is found multiple times for different radial segments of $S_{xx}(\Omega_r, \Omega_\theta)$, and each measurement is used to determine a single sample of $\alpha(\Omega_\theta)$. For example, to determine $K$ uniformly spaced samples of $\alpha(\Omega_\theta)$, the spectrum can be divided into $K$ contiguous segments, each spanning $\pi/K$ radians. (As mentioned in [38], it might also be of interest to allow some overlap of these segments, giving each a greater span and introducing some redundancy in the measurements.) The $k$th localized measurement is thus determined through

$$\bar{S}_k(\Omega_r) = \frac{1}{(u_k - l_k)\Omega_r} \int_{l_k}^{u_k} \log_{10}(S_{BL}(\Omega_r, \Omega_\theta)) d\Omega_\theta,$$

(3.22)

with upper and lower bounds $u_k$ and $l_k$. This segment measurement is then used to determine the sample $\alpha((u_k - l_k)/2)$, selected as the parameter best fitting the measurement to (3.20).

Parameter selection for the $k$th segment is determined by selecting decay parameter $\alpha_k$ and scaling constant $c$ that minimize the function

$$\int_{0^+}^{\pi} \left| \bar{S}_k(\Omega_r) - \log_{10}(cS(\Omega_r)) \right|^2 d\Omega_r.$$

(3.23)

This is the mean squared error between the measurement (3.22) and the scaled function (3.20). The optimal parameters for minimizing (3.23) will be determined through numerical optimization. The lower bound is set to $0^+$ to exclude $\Omega_r = 0$ from the calculation, due to the problems with discontinuity there. In practice with a discrete spectrum, this excludes a single value from the calculation. This also prevents some problems which can occur in numerical calculation of $\alpha_k$, since the position of the curve can differ considerably at low values of $\Omega_r$, as was illustrated in the detail box of Fig. 3.11.

The numerical calculation of optimal parameters is performed using the iterative function $\text{fmincon}$ found in the MATLAB Optimization Toolbox [76], operating on a discrete representation for (3.23). Only boundary constraints are used (to keep $\alpha_k$ and $c$ within reasonable limits determined from initial observations) and the gradient and Hessian are automatically calculated numerically by
the function. In initial testing, it was found that allowing the constant $c$ to vary as a function of $\Omega_\theta$ introduced too much variance in the final parametric estimate, leading to an unusable spectral model. Consequently, a single scaling term is first determined for the entire function, after which only the parameter $\alpha_k$ needs to be determined for each zone (this also means only $K + 1$ parameters are required to completely define the model). To further simplify the optimization process, a global selection for $\alpha$ is first determined by minimizing a global, scale-normalized version of (3.23), given by

$$\int_{\Omega_\theta} \left| \left[ \bar{S}(\Omega_r) - \bar{S}(0^+) \right] - \left[ \log_{10}(S(\Omega_r)) - \log_{10}(S(0^+)) \right] \right|^2 d\Omega_r. \tag{3.24}$$

This minimization determines a global value of $\alpha$ independent of global scale $c$, which is then determined through (3.23) using the fixed global $\alpha$. Once the global $c$ is determined, local selection of $\alpha_k$ is determined for each angular zone using (3.23) with the fixed $c$ value. Following this, the continuous function $\alpha(\Omega_\theta)$ is determined via a smooth interpolation.

To summarize the process, the parametric spectral model is determined from a non-parametric estimate through the following steps:

1. Global measure $\bar{S}(\Omega_r)$ is computed from (3.9).
2. Global decay parameter $\alpha$ is estimated via numerical optimization of (3.24).
3. Global scaling constant $c$ is estimated via numerical optimization of (3.23) with fixed $\alpha$ from previous step.
4. For each of $K$ zones, local decay parameter $\alpha_k$ is estimated via numerical optimization of (3.23) with fixed $c$ from previous step.
5. $K$ samples of $\alpha_k$ used to interpolate continuous function $\alpha(\Omega_\theta)$, used to determine $S(\Omega_r, \Omega_\theta)$ through (3.21), scaled by $c$.

This procedure replaces the initial problem of minimizing a single 2D function of several variables with the lower complexity problem of minimizing multiple 1D functions of a single variable. The above procedure requires $K + 2$ distinct 1D
optimizations. The main strength of this method is its use with degraded images. As discussed above, standard degradations tend to distort the higher frequencies of the spectrum more significantly, and leave the lower frequencies relatively unaffected. The upper bound of the integrals in (3.23) and (3.24) can be adjusted to account for this, allowing a parametric estimate to be determined without having to include the less reliable data. The region of $S(\Omega_r, \Omega_\theta)$ can then be extended to include the desired frequencies from the parametric model. This provides an interpolated correlation model without spectral images (which cannot be fully suppressed using traditional linear interpolation).

**Spectral Ray Model**

The spectral ray model is a parametric model designed to account for straight line features leading to significant radial spectral characteristics as shown in the example spectra of Fig. 3.6(b) and Fig. 3.10(b). This approach was originally presented in [67] and improved upon in [38], which also demonstrated an approach for combining the spectral ray model with the smooth decay model described above. As described in [38], the spectral ray model is based on the Fourier transform pair

$$e^{-\beta|\tau_1|} \iff \frac{2\beta}{\beta^2 + \Omega_1^2}, \quad \beta > 0. \quad (3.25)$$

Considered as a 2D function, the fact that the correlation domain representation on the left hand side of (3.25) is constant on $\tau_2$ means that the transform is zero for all $\Omega_2 \neq 0$. This gives a spectral ray on the $\Omega_1$ axis, with a decay shape controlled by $\beta$. Observed spectral rays will of course not precisely correspond to this simple function, which is intended to provide the general trend of the features rather than an ideal description. Similar to the smooth decay model discussed above, the model can be used to describe high frequency spectral features based on low frequency observations, making it a useful method for predicting original spectra from observations of degraded images.

For spectral rays not on the $\Omega_1$ axis, the representation (3.25) can be rotated
by an angle $\phi$, which gives the correlation function

$$R(\beta, \phi) = \exp \left( -\beta \left| \cos(\phi) \tau_1 + \sin(\phi) \tau_2 \right| \right),$$  \hspace{1cm} (3.26)

and the correspondingly rotated spectral function

$$S(\beta, \phi) = \left\{ \begin{array}{ll}
\frac{2\beta}{\beta^2 + (\cos(\phi)\Omega_1 + \sin(\phi)\Omega_2)^2} & \text{for } \tan(\phi) = \frac{\Omega_2}{\Omega_1} \\
0 & \text{otherwise.}
\end{array} \right.$$  \hspace{1cm} (3.27)

Thus, two parameters are required for each individual spectral ray: direction $\phi$ and decay $\beta$.

Multiple spectral rays are considered simultaneously using a correlation domain multiplication of individual functions (3.26). With $L$ distinct spectral rays, this provides the composite spectral ray model

$$\tilde{R}(\tau_1, \tau_2) = \prod_{l=1}^{L} R(\beta_l, \phi_l).$$  \hspace{1cm} (3.28)

The equivalent spectral model, referred to as $\tilde{S}(\Omega_1, \Omega_2)$, is found through the Fourier transform of (3.28), which is the 2D convolution of the $L$ terms $S(\beta_l, \phi_l)$. The structure of the $\tilde{S}(\Omega_1, \Omega_2)$ can be described based on this convolution, and is most easily considered for the case of $L = 2$. When two line functions are convolved in a 2D space, the result is simply the replication of the first line from the intersection outward along the second, being weighted by the value of the second at each point. With sufficiently rapid decay (which will be the case for typical image spectra rays), the higher frequencies–points on each ray far from the origin–are nearly zero. Only near the origin, where $S(\beta, \phi)$ reaches its maximum value of $2/\beta$, will the replication of one line along the second be significant. Hence, the observed structure of the convolution is dominated by an apparent overlay of the two rays, subject of course to some spreading from the convolution with significant low-frequency components. A similar pattern follows for the general case of an arbitrary number of spectral rays, meaning $\tilde{S}(\Omega_1, \Omega_2)$ tends to retain the features of the individual rays through the 2D convolution, although for sufficiently large $L$ the combined spreading from multiple convolutions will eventually reduce some of
this structure. Some spreading of the spectral rays is also observed in the examples Fig. 3.6(b) and Fig. 3.10(b).

The originally conceived spectral ray model [67] used only the representation (3.28) to provide the spectrum. Individual values for $\beta_l$ and $\phi_l$, along with the number of spectral rays ($L$), were determined using an iterative numerical algorithm that attempted to fit the parametric model of the observed spectrum based on the radial average measurements $\hat{S}(\Omega_\theta)$ defined by (3.10). This approach did have a tendency to overemphasize features on spectral rays’ direction, and was improved upon in [38] by combining it with the smooth model. The combined approach determines the smooth model first (using the method described in the previous section), then determines if any spectral rays are needed to improve the model accuracy. If so, the combined approach iteratively determines the correct number and parameters for the required spectral rays, which are then overlaid on the smooth model.

For instance, consider the $\hat{S}(\Omega_\theta)$ measure depicted in Fig. 3.8(b). After a smooth model is constructed for this image, unaccounted peaks are selected based on the difference between the $\hat{S}(\Omega_\theta)$ measures of the observed and (smooth decay) modelled spectra. The peak values of this difference are examined against the local average of the difference (a six-degree span was used in [38]). If above a pre-specified threshold, these peaks are deemed to be significant enough and a $\phi_k$ is selected at the corresponding value of $\Omega_\theta$. Once $\phi_k$ are selected, the corresponding $\beta_k$ parameters must be determined. This is done using an iterative algorithm which is somewhat computationally costly, but capable of determining the parameters.

Each iteration determines $\hat{S}(\Omega_1, \Omega_2)$, the combined smooth decay model with spectral rays added. The spectral ray component is determined in the correlation domain via (3.28) using the current set of $\beta_k$ parameters. The spectral domain equivalent is found from the DFT of the correlation function multiplied by a 2D Tukey cosine-tapered window. The measurement (3.10) is then applied to the combined spectrum $\hat{S}(\Omega_1, \Omega_2)$, producing $\hat{\hat{S}}(\Omega_\theta)$. This model measurement is then compared to the true angular measurement $\hat{S}(\Omega_\theta)$ to refine the $\beta_k$ parameters. Unfortunately, $\hat{\hat{S}}(\Omega_\theta)$ will vary nonlinearly with the $\beta_k$ parameters, and a change
to any $\beta_k$ will alter $\hat{S}(\Omega_\theta)$ for all $\Omega_\theta$, not just at $\Omega_\theta = \phi_k$ (this is due to the complexity of the convolutional relationship between multiple spectral rays). However, an increase in a particular $\beta_k$ will primarily result in an increase to the measurement $\hat{S}(\phi_k)$, with lesser changes along other directions. This provides justification for a fairly simple algorithm: each $\beta_k$ is increased by stepsize $\mu_k$ if $\hat{S}(\phi_k) < \bar{S}(\phi_k)$, and decreased by $\mu_k$ if $\hat{S}(\phi_k) > \bar{S}(\phi_k)$. If sign $[\hat{S}(\phi_k) - \bar{S}(\phi_k)]$ changes from one iteration to the next, the stepsize $\mu_k$ is replaced with $\mu_k/2$ to narrow the search on $\beta_k$. In simulations the values of $\hat{S}(\phi_k) - \bar{S}(\phi_k)$ collectively converge reasonably close to zero within twenty iterations, although the values at other $\Omega_\theta$ will differ. A detailed example is provided in the next section.

**Example**

An example combining the proposed smooth decay and spectral ray models is presented here. Individual examinations of earlier versions of these methods were presented in [67,68]. This example and discussion were presented in [38]. The combined model will be applied to the test image shown in Fig. 3.12. This original $400 \times 400$ pixel image is subjected to Gaussian blurring ($\sigma = 1$ pixel), reduced in resolution to a size of $200 \times 200$ pixels, then corrupted by additive white Gaussian noise (AWGN) with a variance of 50. To provide a comparison, the MMSE restoration is found using six distinct spectral models. The models considered are: the periodogram estimate (windowed, edge-blurred) from the original image, this smooth decay model—both with and without spectral ray components—numerically determined from the periodogram estimate of the original image, the smooth decay model (again, both with and without spectral rays) numerically determined from the periodogram estimate of the LR degraded version of the image, and finally the periodogram estimate of the degraded spectrum under a bandlimited assumption (essentially setting unknown higher frequencies to zero).

Performance is evaluated using two distinct measurements of reconstruction quality: peak signal-to-noise ratio (PSNR) and an image quality index (QI)
Figure 3.12: Test image used to evaluate the spectral models.

proposed in [77]. The PSNR is defined as

$$\text{PSNR} = 10 \log_{10} \left( \frac{x_{\text{max}}^2}{\text{MSE}} \right),$$  \hspace{1cm} (3.29)

where $x_{\text{max}}$ is the maximum possible pixel value (typically 255) and MSE is simply the measured mean-squared error between the original and restored image. Use of the PSNR (or MSE) as an image quality evaluation tool has been scrutinized for some time, but its simplicity, accuracy as a pure difference evaluator, and comparability with MMSE-based reconstruction tools have always managed to keep it at the forefront of traditional evaluation standards. Arguably, the best evaluation is based on time-consuming and costly human appraisal, but these results are problematically subjective. However it is well established that many basic numerical evaluation measures, like PSNR or MSE, do not necessarily match well to perceptual quality evaluation ([6] Chapter 8.2). The QI measurement proposed in [77] is a numerical evaluation intended to better match human perception of difference.
between two images, and is defined by

\[ QI = \frac{4\sigma_{xy}\bar{x}\bar{y}}{\left(\sigma_x^2 + \sigma_y^2\right)(\bar{x}^2 + \bar{y}^2)}, \quad (3.30) \]

where \( \bar{x} \) and \( \bar{y} \) are the sample means of the two images, \( \sigma_x^2 \) and \( \sigma_y^2 \) are the unbiased sample variances of the two, and \( \sigma_{xy} \) is the unbiased covariance between the two images. The measurement was derived by combining a correlation measurement with luminance and contrast distortion measurements. Possible values for the QI measurement can vary between \([-1, 1]\), with 1 indicating perfect image recovery.

The reconstruction performances under each of these distinct spectral models are shown in Fig. 3.13 as a function of the number of angular zones used to calculate the smooth model, i.e., the number of separate evaluations of (3.22). There is a general consistency between the PSNR results shown in Fig. 3.13(a) and the QI results in Fig. 3.13(b). As expected, the periodogram spectral estimate from the original image produces the highest quality recovery and the bandlimited PSD produces the lowest. In terms of PSNR the proposed parametric models perform roughly midway between the two extremes, but closer to the upper limit under the QI measurement. The performance of the parametric model determined from the degraded image is just slightly lower than as determined from the original image, indicating a degree of robustness against degradation. The results also show that the reconstruction performance tends to increase as the number of angular zones is increased, although there is an exception of a slight decrease for the three-zone parametric model constructed from the original image. This occasionally happens for a model using a low number of zones due to the difficulty in providing a single parameter which is accurate over a large angular span, and can be eliminated by increasing the number of angular zones which allows for greater localized accuracy. Similarly, overlaying the spectral ray model will also tend to improve measured performance since the spectral rays each introduce spikes to only a very narrow portion of the angular function (3.10) and cannot be accurately described by the smooth model employing small number of wide angular zones. It follows that the

\(^6\)Limited testing in their paper indicates the QI successfully meets this objective, but it should by no means be considered free from the standard deficiencies of numerical evaluation techniques.
benefit provided by introducing spectral ray components also tends to decrease as the number of zones grows large, as the decreasing width of the individual zones allows for greater localized accuracy. While the improvements from spectral rays decreases as the number of angular zones increases, they are still of interest for applications where a low parameter transmission/storage capacity prohibits the use of a very large number of angular zones.

Figure 3.13: PSNR (a) and QI (b) for the MMSE recovery of the image shown in Fig. 3.12 using different PSD models for reconstruction filter design.

To further illustrate, the radial average functions (3.10) of the spectral models are shown in Fig. 3.14 for the original PSD and the smooth model, both with and without spectral rays. The first two sub-figures show the measurements from the original and degraded images using 10 angular zones, while the final two sub-figures both use 80 angular zones. The spectral rays help to provide a more accurate modelling with a low number of zones, but when there is a large number of angular zones they do not contribute much beyond the underlying smooth model.
Figure 3.14: Weighted angular averages for the image in Fig. 3.12 (thin black curve) and its smooth (thick blue curve) and combined spectral ray (dashed red curve) models. (a) Original image, 10 zones. (b) Degraded image, 10 zones. (c) Original image, 80 zones. (d) Degraded image, 80 zones.
3.3.2 Local Correlation Modelling

The principal problem with the global modelling techniques discussed above in Section 3.3.1 is simply that most images do not have globally constant features or statistics. While the above methods were focused on spectral modelling, much of this discussion can be applied to alternative global modelling techniques. Examining an image via its energy spectrum effectively averages the contribution of all the features into a single model, providing statistics that on average work best for all content considered simultaneously, but are never optimal for any specific locality. The alternative approach used in this section is to develop correlation models which are only considered valid over a smaller region. Changing the correlation model based on the local content provides a better representation for the image as a whole.

Although theoretically superior to global modelling, a localized approach does present more complications to overcome. First, the models must be defined from only a relatively small amount of observation data. While this is advantageous in that the determined model will be based only on local statistics and not on the average statistics of the entire image, the problem with using such a small amount of data is that the variance can lead to bad models, in the worst case possibly even leading to an ill-conditioned or unstable reconstruction. Thus, extra care must be taken to ensure the model remains stable. A second complication is the potential for significant variance between adjacent local models which, depending on the reconstruction method used, can lead to patching artifacts in which the difference in reconstruction of adjacent localities is visible (even if the quality of each region would not individually be questioned by the observer). A final complication is the potential for significant increase in required computation. While a global model uses all the observed data, it only does so once. In contrast, local modelling requires solving many smaller problems, which can potentially require more operations. Additionally, constructing a model for a specific locality often requires information based on observations from a larger area surrounding the central portion of interest. This can lead to a high level of redundancy in the overall computation, with each observation being used multiple times over different local models. These various
complications are not by any means insurmountable, but must be considered in the development and utilization of a model.

The non-stationary correlation model $E[x[n_1, n_2]x^*[m_1, m_2]]$ is considered here. The lack of stationarity means that the standard Fourier domain processing methods used above are not applicable, which will require a spatial domain solution. For simplicity, the model considered here can be viewed as locally stationary, allowing the standard $R_{xx}[\tau_1, \tau_2]$ notation to be used, so long as it is understood that the function $R_{xx}[\tau_1, \tau_2]$ must be redefined for each unique local region considered.

In the remainder of this section the function $R_{xx}[\tau_1, \tau_2]$ refers to the correlation of the observed LR signal, and the function $R_{yy}[\tau_1, \tau_2]$ refers to the correlation of the desired HR signal. For reconstruction of the original image, $R_{yy}[\tau_1, \tau_2]$ is needed. However, only $R_{xx}[\tau_1, \tau_2]$ is directly observed. Thus, similarly to the global scenario, the principal problem for local correlation estimation is that the statistics of the original (or, high-resolution) data must be determined from the observed LR data. This leads to many of the same problems of estimating from degraded content as described above. The remainder of this section discusses various specific methods that have been developed to determine the original HR statistics from the observed LR data. These methods typically make use of only a single LR frame to determine the model from, rather than the collection of all LR data. The reason for this is due to the difficulty and increase in computational complexity of applying these estimation problems to a set of non-uniformly sampled pixels. It is simply more feasible to determine a model based on a single uniformly sampled LR data set. However, modifications can be employed to some approaches to take advantage of the additional information present in the entire non-uniformly sampled pixel set.

**High-Resolution Covariance Estimation from Geometric Duality**

The approach employed by Li and Orchard in [78] made use of a property therein referred to as “geometric duality.” This property is used to determine
an HR covariance\textsuperscript{7}, \( R_{yy}[\tau_1, \tau_2] \), directly from an LR covariance, \( R_{xx}[\tau_1, \tau_2] \). The approach was developed for a simpler problem than the general super-resolution case, and consequently only required finding \( R_{yy}[\tau_1, \tau_2] \) for double the resolution (in each dimension) of \( R_{xx}[\tau_1, \tau_2] \). That is, the HR sampling grid for \( R_{yy}[\tau_1, \tau_2] \) was based on half the sampling period of \( R_{xx}[\tau_1, \tau_2] \), which simplified later calculations.

In a simpler illustrative case, the authors considered a 1D where the covariance was assumed related to sampling distance by \( R(\tau) = e^{-\tau^2/2\sigma^2} \), which allowed establishing the relationship \( R(d) = R(2d)^{1/4} \), where the sampling distance \( d \) is then used to relate the covariance of the low- and high-resolution representations. A similar approach is used for the 2D case. If the HR grid is sampled on the integers, the LR grid can be considered as sampled on the even indexes. The property of duality can then be used to find the covariance value including the odd indexes, essentially by relating them to the measurable covariance values. For example, a known measurement between pixels at \([n, m]\) and \([n+2, m+2]\) is used to find the estimate between \([n, m]\) and \([n+1, m+1]\). The can be extended to include the rest of the required correlation values based on regularity of the indexes. Some illustrative diagrams are shown in the first two figures of \[78\].

The principal disadvantages of this approach are that the resulting structure is limited by the regularity between the desired HR grid and the observed data. The method cannot be applied to general nonuniformly sampled scenarios. The method also lacks an ability to account for the degradations (distortion and noise) in the measurement. There is also the innate assumption of geometric duality which, while usually applicable, is not guaranteed to fit all data scenarios.

Local Weighting of Stationary Model

The approach provided by Hardie in \[79\] was based on adjusting a stationary correlation model given by the continuous function

\[
R(\tau_1, \tau_2) = \sigma^2 d^\rho \sqrt{\tau_1^2 + \tau_2^2},
\]

\textsuperscript{7}The term covariance used in \[78\] in this case is equivalent to this dissertation’s use of autocorrelation.
which is simply a scaled form of (3.18), where $\rho = e^{-\alpha}$. Finite power signals require $\rho < 1$. The scaling term $\sigma_d^2$ in (3.31) is simply the variance of the desired image, which is also accounted for in the practical versions of (3.18). Model parameters $\sigma_d^2$ and $\rho$ can be selected from the data or assumed. The continuous function $R(\tau_1, \tau_2)$ provides the statistical model of the original image and any necessary statistical measurements (e.g., the correlation function of the degraded sequence and the cross-correlation between the desired HR pixels and observed LR pixels) can be determined by applying the appropriate degradations to this continuous function. For instance, passing the original signal $y(t_1, t_2)$ through a distortion function given by $h(t_1, t_2)$ to produce $x(t_1, t_2)$ leads to the distorted correlation function

$$R_{xx}(\tau_1, \tau_2) = R_{yy}(\tau_1, \tau_2) * h(\tau_1, \tau_2) * h(-\tau_1, -\tau_2).$$

(3.32)

This allows statistical content models to be constructed under the standard degradation model Fig. 2.1.

Modification of (3.31) for a non-stationary equivalent is provided by adjusting the variance term locally, giving

$$R_i(\tau_1, \tau_2) = \sigma_{d_i}^2 \rho \sqrt{\tau_1^2 + \tau_2^2},$$

(3.33)
in which $\sigma_{d_i}^2$ is the local variance. The decay factor $\rho$ is kept constant. The model then requires estimation of $\sigma_{d_i}^2$ based on the observed data. Based on (3.32), this local variance of the original image is related to the local variance of the distorted image (denoted as $\sigma_{f_i}^2$) through

$$\sigma_{d_i}^2 = \frac{1}{C(\rho)} \sigma_{f_i}^2,$$

(3.34)

where

$$C(\rho) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho \sqrt{\tau_1^2 + \tau_2^2} (h(\tau_1, \tau_2) * h(-\tau_1, -\tau_2)) d\tau_1 d\tau_2.$$  (3.35)

This in turn requires estimation of the distorted variance $\sigma_{f_i}^2$. An estimate can be made directly from the local observations, which provide sample variance $\hat{\sigma}_{x_i}^2$. Assuming known independent additive noise of $\sigma_n^2$, the estimated distorted variance can be found through

$$\hat{\sigma}_{f_i}^2 = \hat{\sigma}_{x_i}^2 - \sigma_n,$$

(3.36)
which is sufficient for finding an estimate of $R_{yy}(\tau_1, \tau_2)$, the local undegraded correlation function.

The main strength of this method is that it is a computationally efficient and fairly direct estimation technique. Since it is based only on a the sample variance of the observed data, it can be found from general nonuniformly sampled data. Hardie also provided a polynomial mapping approach to find $\sigma^2_d$ from the sample variance, an alternative to the linear mapping of (3.35). The main approach does however require a reasonably accurate model for the degradation process (which is typically fine in practice). There is also some limitation in terms of the scope of the model as the local correlation estimate is isotropic and does not account for edge features as in [78]. The implementation in [79] also does not allow the decay parameter $\rho$ to vary, which does limit the correlation model. Improvement of the method to a varying $\rho$ is stated as ongoing work.

**Local Correlation Estimation from Fourier Description**

A method proposed in [39] is based in part on the earlier modelling approaches in [38]. The method seeks to determine a uniformly spaced HR correlation estimate from a uniformly spaced LR correlation estimate. Unlike the approach in [78], the difference in resolution between the LR and HR grids can be arbitrary (including irrational scaling factors), so the property of “geometric duality” cannot be taken advantage of. The method uses a Fourier description, allowed from the assumption of local stationarity. The basic approach used is simply to determine a local, non-parametric spectral description of the LR data, translate this description to an HR description, then apply the inverse Fourier transform to find the HR correlation model. The first and final of these steps are fairly trivial, and the second step can be easily performed based on some of the prior results for stationary models.

Assume a resolution enhancement factor of $s_1 \times s_2$ is desired. The first step determines the LR non-parametric spectral estimate from a small block of data (stationarity is assume over the entire block, but applied in reconstruction only to a reduced portion existing in the center, such that there are no concerns
of boundary artifacts in the reconstruction). Without loss of generality, a square block of \( L \times L \) pixel is used to find the local LR spectrum. The standard boundary-blurred periodogram approach (as described in Section 3.3.1) is used, distorting the edges of the block to prevent axis artifacts in the spectrum. Since relatively small amounts of data are used in this case, the span of blurring is much less than that used in the global model case. This provides a local spectrum in an \( L \times L \) representation.

The next step, translating this LR spectrum to the HR spectrum, is done in the spectral domain. The HR local spectrum is a block of size \( s_1L \times s_2L \), rounded to the nearest integer. The complete \( L \times L \) LR FFT block is replicated onto the central \( L \times L \) portion of the HR FFT block, in both cases centered on the DC component. This description assumes \( L \) is odd, otherwise the LR FFT block must be adjusted so the HR FFT block is symmetric. It is important to note that proportionality between rate scaling and frequency contraction can be lost for non-integer \( s_1 \) and \( s_2 \), due to rounding required for an integer HR FFT block size. In this case the magnitude samples are still replicated rather than interpolated or reevaluated with a fractionally spaced FFT. This does potentially introduce a slight error on the spectral model due to misalignment of DFT samples in the HR and LR representations. However the error is small, does not require interpolation to correct, and does not significantly affect reconstruction in practice (most importantly, the trend of radial decay is preserved as the measured LR spectrum is centered at the origin in the HR model).

This provides the base frequencies of the HR FFT block. There is unfortunately not much information on the higher frequencies that can be determined from the source data. They are approximated from the isotropic model (3.20) that uses decay parameter \( \alpha \) and radial frequency \( \Omega_r \), and applied directly to the DFT samples of the higher frequencies. Generally, a scaling parameter is also needed, so that the magnitude of the central block matches the added spectral content (similar to the use of \( \sigma_d^2 \) in Hardie’s approach (3.33)). However, noting that for \( \Omega_r >> \alpha \) the decay parameter and a scaling term are functionally identical in (3.20), the shape of the added frequencies can be predetermined from a single function, and
then scaled based on the measured LR block. Scaling is selected simply based on
the mean value of some of the known border samples of the LR FFT, essentially
providing a smoother transition between the low and high frequency portions of
the HR FFT (the main goal is to avoid a sharp magnitude transition on the bound-
ary). Alternative approaches are certainly valid as the reconstruction performance
turns out to be fairly invariant to smaller changes in the scaling.

The third and final step simply returns the spectral model to the correlation
domain through the inverse FFT. The desired HR correlation values can then be
supplied to the reconstruction algorithm.

This overall approach is more computationally intensive than the previous
two, so effort is made to reduce the required number of operations. The following
analysis was provided in [39]. Since the approach requires a separate correlation
estimate for each locality (although this can be modified based on an approach
such as regional classification), use of an efficient FFT processing reducing the
required computation is desirable.

As described above, a 2D FFT is used to determine the LR frequency do-
main representation from the boundary-blurred LR data block. This result is then
plugged into the low frequencies of the HR frequency representation which is com-
bined with high frequencies modelled through scaled (3.20). Generally speaking,
the most significant portion of the correlation estimation is expected to be the HR
inverse FFT. In a simplified case, consider an $N \times N$ data block with a correspond-
ing $O(N^2 \log(N^2))$ computational load [11]. A scaling factor of $m \times m$ corresponds
to an approximate $mN \times mN$ HR block, the direct processing of which requires
a $O(m^2 N^2 \log(m^2 N^2))$ load. However, a couple aspects of the HR Fourier represen-
tation can be exploited to reduce the required computation. First, the low and
high frequency components can be processed separately. The high frequencies are
all determined by a scaling of (3.20). Thus, the inverse FFT of these high frequen-
cies can be found once for a specified correlation block size and scaling factor, then
applied to all blocks with the appropriate local scaling (the $N \times N$ block of low
frequencies having been replaced with zeros for this computation).

The remaining $N \times N$ low frequency block is then zero padded to (approx-
imate) size $mN \times mN$ and its inverse calculated to determine each HR correlation block’s unique low-frequency component. Taking advantage of separability, one representation for the calculation of a 2D inverse discrete Fourier transform (IDFT) is

$$q = FQF^H,$$  \hspace{1cm} (3.37)

where $q$ and $Q$ are the spatial and Fourier domain equivalents and $F$ is the transform matrix. Since the calculation is transforming a PSD to a correlation function, $Q$ and $q$ are real and even symmetric. Breaking the transformation matrix into its separate real and imaginary parts, $F = F_r + jF_i$, the IDFT is represented as

$$q = (F_r + jF_i)Q(F_r + jF_i)^H$$

$$= F_rQF_r^T + F_iQF_i^T + \text{imaginary components.}$$  \hspace{1cm} (3.38)

The fact that $q$ is real implies that there are no imaginary components. Additionally, it can be shown that $F_i$ has an odd symmetry and is zero-valued along the axes. Combining this with the even symmetry of $Q$ requires $F_iQF_i^T$ to equal zero. Consequently, the IDFT can be determined through the simplified calculation

$$q = F_rQF_r^T.$$  \hspace{1cm} (3.39)

Furthermore, only a smaller $N \times N$ portion of $Q$ is non-zero, allowing significant portions of $F_r$ to be discarded prior to calculation. The calculation of an $mN \times mN$ correlation matrix thus requires the computation load for multiplying real matrices of size $(mN \times N)(N \times N)(N \times mN)$. Further exploiting the symmetry of $q$ (roughly halving the required number of output values to be computed), a naive matrix multiplication requires approximately $O((m^2 + m)N^3/2)$ real multiplications. It may certainly be possible to further reduce the required number of computations by modifying a 2D FFT approach as an alternative to the DFT matrix implementation discussed above. In experiments so far, a simple MATLAB implementation using (3.39) has been significantly faster (roughly an order of magnitude for typical values of $m$ and $N$) than a direct zero-padded FFT, although this would not necessarily be expected to remain the same in an optimized hardware or software implementation.
The main advantages of this method for correlation modelling is that the local statistics are taken into account (improving the accuracy) and the representation does not require a specific regularity between the LR and HR representations, as with the method in [78]. Although not considered here, the use of a Fourier based method also easily allows for degradations such as distortion and noise in the observed data to be corrected during the process of estimation. This is a similar feature to the method in [79]. The main disadvantage of the approach is extra computation time, although effort has been made to reduce the required number of computations.

3.4 Acknowledgement

4 Still Image Super-Resolution Reconstruction

4.1 Introduction

With the various necessary models in place, the subject of this dissertation now moves into super-resolution reconstruction, the process which uses the observed low-resolution (LR) data and the models for degradation and content to recover an estimate for the desired high-resolution (HR) content. The reconstruction approach considered here is intended to be fully separate from the modelling scenario; while the previously obtained models influence the reconstruction, the reconstruction in no way directly influences the modelling. Most complete super-resolution methods use this approach, although there are a minority of approaches which work to simultaneously obtain a model and reconstruction [80]. A complete super-resolution system can also be designed based on a recursive procedure, in which the final reconstruction is used to refine the model to obtain the next refinement of the reconstruction. However, for the philosophy of this dissertation, using the reconstruction result to subsequently update the model does not constitute a joint approach, as the reconstruction assumes an externally provided model.

This dissertation also distinguishes between resolution enhancement applied to a single image and that applied to a set of multiple images. While the first case is certainly a problem of interest, it is a very different and simplified problem from the multi-image case, requiring no consideration for fusing the content of the images together. The true power of multi-image super-resolution is in combining the non-
redundant content from multiple observations to provide an improved estimate of the high-resolution. Resolution enhancement for the case of a single still-image involves estimating based on the model, but never increases the information beyond that of a single LR image (instead, it merely changes the representation of that information to provide a more visually meaningful representation). There is some inconsistency in the literature as to what specific scenarios validate use of the term “super-resolution,” but in this dissertation the term is applied only to methods that can be applied to the multi-image scenario. Single image scenarios, lacking the problem of data fusion, are referred to using the more general terminology of “resolution enhancement.”

As much as data fusion for multiple LR images potentially provides, it is also the most significant source of troubles in the super-resolution problem. Practical use requires sufficiently accurate registration modelling, as well as possibly a method of dealing with registration errors. The reconstruction method may also limit the permissible forms of registration. A more detailed examination of this is provided in discussion of specific methods below.

The reconstruction problem is considered in this chapter for the still-image case. There can be some confusion on the distinction between approaches for still images and those for video, but this chapter considers two important limitations which are required for the still image case. The first of these is the assumption that the observations come from an unchanging scene or, equivalently, are simultaneously obtained from a dynamic scene. In both cases, the content of the observations can be interpreted as having come from a single source (taking the form of a continuous function or high-resolution discrete model). Accordingly, while a separate degradation model is required to describe each LR observation, only a single model for the HR content is needed. The second limitation for the still image case involves the registration model. It is generally assumed that each observation is a standard digital image, uniformly sampled from the underlying continuous 2D scene. In the absence of unusual and bizarre spatial warping of the content, the registration can be modelled through the global affine transform (2.46). This provides the collection of LR pixels on the continuous 2D space using
a very compact representation.

Combining these limitations leads to a very precise description of the individual pixel locations. Violations of either of these limitations can lead to reconstruction errors that result from either misregistration of the given content or attempting to incorporate unregisterable content in the solution. In many practical scenarios such violations are impossible, since simplified models are often insufficient for describing the nature of the true observation. This can either be resolved by the reconstruction method (through some form of automatic correction of discrepancies between the model and true observations), or the resulting super-resolved images can be put through some form of post-processing to correct errors.

The organization of this chapter is as follows. First, a brief review of the history of still-image super-resolution is provided in Section 4.2. A detailed examination of Fourier-based reconstruction approaches is provided in Section 4.3, along with the new reconstruction approach proposed in this dissertation. A brief discussion of some alternative reconstruction approaches is provided in Section 4.4. The algorithms are applied to two reconstruction scenarios in Section 4.5 to evaluate the performance.

### 4.2 Problem History

This description of prior results in still-image super-resolution is largely borrowed from [38]. There are many different reconstruction approaches that have been considered over the years, and more detailed description of several methods can be found in [4], along with a well populated catalog of references up to 2003.

Some of the earliest initial approaches [3,53] to the problem of image super-resolution were formulated in the frequency domain and assumed a bandlimited HR content model. Under this assumption, aliasing causes the individual frequencies of the LR image’s Fourier transform to be determined by the weighted sum of a specific finite set of frequencies for the HR image’s Fourier transform. This weighting varies for each LR image as a function of blurring and translational
shifting. It is then possible to construct the linear matrix equation $y = Wx$, which relates the known vector of LR image Fourier transforms, $y$, to the product of the known weighting matrix, $W$, and the unknown vector of aliased HR image frequencies, $x$. Given a sufficient number of LR images (corresponding to an average sampling rate at least equal to the Nyquist rate), this matrix equation is at least critically determined and the HR image information contained in $x$ can be found.

Another early solution approach designed a set of LSI filters to operate on the sampled images [54], in what was a 2D formulation of the generalized sampling expansion (GSE) [81, 82]. While eventually determining spatial domain filters, the frequency domain representation of these filters provides the same reconstruction solution as the methods above. In the view of this dissertation, the approach is considered a variant within the family of frequency domain reconstructions. Further, the frequency domain GSE can be considered the classic formulation for super-resolution under a critically-sampled bandlimited assumption. Reconstruction from an undersampled set as in [83] is based on a generalization of the GSE [84] which determines the MMSE reconstruction under an alternative (non-bandlimited) reconstruction space. Examinations of alternative reconstruction spaces can also be found in [85, 86]. Another recent approach considering Fourier domain reconstruction was provided in [87].

The approach proposed in this chapter is best seen as a continuation of these earlier Fourier-based approaches, seeking to provide some of the advantages found in frequency domain solutions, while at the same time improving on some of the deficiencies found in the earlier approach (most importantly, the removal of a bandlimited requirement). Current state-of-the-art alternative methods typically rely on entirely different (non-Fourier) approaches for reconstruction. Many of the more recent methods have moved away from Fourier methods due to the previously understood limitations in content modelling, as well as the necessary limitations in degradations modelling (specifically, the use of linear shift-invariant distortion functions and problems that occur in the presence of non-translational registration models). A brief listing of many of these alternative approaches follows. This is a
fairly representative selection, but by no means complete.

Typically, these reconstruction approaches have a greater computational cost in comparison to frequency domain methods. For instance, the technique of projection onto convex sets (e.g., [88]) generally requires a large number of costly iterations to converge to an acceptable solution. An iterative approach was also used in [80], in this case to jointly optimize a maximum a posteriori (MAP) estimate of the HR image and the LR images’ registration. An advantage of this joint MAP formulation is that it does not require a separate process to determine the registration. This approach is one of the family of methods based on regularization, the principle of introducing some assumption (e.g., smoothness) which provides a solution to an otherwise ill-posed problem. Another paper [89] provided an approach for blind estimation of the LR images’ distortion functions and used these estimates to produce reconstructing functions in a solution based on a generalization of the GSE. Other approaches such as [56,90] sought improvements in computational efficiency. The approach [31] was also more computationally efficient than many of the earlier methods and made use of regularization for the purpose of reducing sensitivity to errors in estimation of the LR images’ degradation. Another interesting improvement was found in [91], which operated directly on transformed versions of the LR images, and allowed improved modelling of the degradation introduced by lossy compression algorithms. Another recent approach [92] provided a joint estimation of registration parameters, degradation models, and the high-resolution image through a stochastic estimation framework.

4.3 Fourier Based Reconstruction

The family of Fourier based techniques is based on an understanding of the specific patterns of aliasing that the image undergoes when it is undersampled. To avoid introducing additional complications, it is generally assumed that the individual LR observations are sampled at the same rate, implying the same aliasing pattern is present in each observation. In the absence of noise, the only distinguishing factor between each LR image is the initial transfer function, $H_i(\Omega_1, \Omega_2)$,
applied prior to sampling. Subscript \( i \) distinguishes the transfer function for the \( i \)th observation. To illustrate the reconstruction procedure, consider the simpler case of a 1D signal bandlimited to \((-\beta, \beta)\). From the sampling formulas (2.24) and (2.25), an alias free representation requires a sampling rate at least equal to the Nyquist rate \( \Omega_N = 2\beta \). Consider instead the signal sampled at half the Nyquist rate, which introduces aliasing. Represented in the frequency domain and ignoring any scaling constant, the observed sampled signal is thus given in the Fourier domain by

\[
Y_i(\Omega) = H_i(\Omega)X(\Omega) + H_i(\Omega - \beta)X(\Omega - \beta), \quad 0 \leq \Omega \leq \beta.
\]

(4.1)

Since \( Y_i(\Omega) \) is periodic with period \( \beta \) this interval is sufficient for describing the entire signal. At some specific frequency \( \Omega_\alpha \), a pair of observations with different distortion functions provides a pair of linear equations that give

\[
\begin{bmatrix}
Y_0(\Omega_\alpha) \\
Y_1(\Omega_\alpha)
\end{bmatrix} =
\begin{bmatrix}
H_0(\Omega_\alpha) & H_0(\Omega_\alpha - \beta) \\
H_1(\Omega_\alpha) & H_1(\Omega_\alpha - \beta)
\end{bmatrix}
\begin{bmatrix}
X(\Omega_\alpha) \\
X(\Omega_\alpha - \beta)
\end{bmatrix},
\]

(4.2)

which can be written as a linear matrix equation \( \mathbf{y} = \mathbf{Hx} \). Provided \( \mathbf{H} \) is non-singular (introducing conditions on the distortion functions), a solution can be determined for \( X(\Omega_\alpha) \) and \( X(\Omega_\alpha - \beta) \). Applying this formula to all frequencies \( 0 \leq \Omega_\alpha \leq \beta \) reconstructs the complete signal \( X(\Omega) \) from the pair of undersampled observations. From (4.2), the solution takes the form \( \mathbf{x} = \mathbf{H}^{-1}\mathbf{y} \), which can be used to determine linear reconstruction filters according to

\[
\mathbf{H}^{-1} = \mathbf{F} =
\begin{bmatrix}
F_0(\Omega_\alpha) & F_1(\Omega_\alpha) \\
F_0(\Omega_\alpha - \beta) & F_1(\Omega_\alpha - \beta)
\end{bmatrix}.
\]

(4.3)

Again, complete representation of the reconstruction filters is given by considering (4.3) for all frequencies \( 0 \leq \Omega_\alpha \leq \beta \), which is essentially an interpretation of the reconstruction result provided by Papoulis' GSE [81].

At different sampling rates, and applied to 2D images, the general representation of (4.2) stays the same, but the number of columns of \( \mathbf{H} \) must change with the number frequencies aliased together, and the number of rows of \( \mathbf{H} \) must change.
with the number of observations. It follows that keeping the number of observations and the degree of aliasing equal will maintain the possibility of an invertible $H$. When this is not the case, the solution cannot be found. In early frequency domain solutions to super-resolution [3, 53], this introduced a bandwidth limitation for the image based on the number of available observations. Higher-resolution representations can only be determined in this case by setting the undeterminable higher frequencies to zero, which introduces rippling artifacts in the reconstruction. This is the primary limitation of a deterministic representation for Fourier domain super-resolution.

### 4.3.1 MMSE Solution

To overcome the limitations of the deterministic model, the approach presented here considers a WSS stochastic signal model for frequency domain super-resolution. This provides an alternative to the deterministic approach of the traditional GSE. The proposed method also easily accommodates stochastic models for additive noise, and can be viewed as essentially an extension of the classic Wiener image restoration problem discussed in Section 2.4, modified to incorporate multiple observations and resolution scaling through a filter bank framework. Much of the discussion here is borrowed from [38], which largely focused on a solution to the problem of super-resolution from a set of jointly undersampled images.

It is assumed here that the PSD of the desired undistorted HR image is known or has been estimated. It is also assumed that linear distortions, registration information, and stationary models for any additive noise are known as well. Several different modelling approaches that have been discussed in the prior chapters can be used, and the reconstruction process takes the models as determined from some external process. The general problem is then considered through a filter bank structure in which the analysis bank is used to model the generation of a set of degraded LR images from a single HR original and the synthesis bank is used to reconstruct a super-resolved output.

\footnote{The term “jointly undersampled” refers to a set of images that is collectively at a sub-Nyquist average rate.}
The design presented here is a modified version of the MMSE generalized undersampling reconstruction originally developed in [84]. Full discussion of this solution to the undersampled stochastic filter bank is provided in the Appendix. For image processing applications, a 2D equivalent is considered with separable decimation and expansion blocks. The HR input $x[n_1, n_2]$ is a discrete model with resolution equal to that of the desired super-resolved image. Its PSD $S_{xx}(e^{j\omega_1}, e^{j\omega_2})$ must be estimated in the practical case. The $c$th of $C$ individual LR images, $w_c[n_1, n_2]$, is obtained from $x[n_1, n_2]$ passed through a sequence of degradations: linear filtering by $H_c(e^{j\omega_1}, e^{j\omega_2})$, resolution reduction via $(D_1, D_2)$-fold rectangular decimation, and additive combination with stationary noise $v[n_1, n_2]$. Each unique $H_c(e^{j\omega_1}, e^{j\omega_2})$ models both blurring and a translational shift.

To obtain the HR output $y[n_1, n_2]$, each of the $C$ LR images is increased in resolution via $(D_1, D_2)$-fold interpolation, then passed through a linear filter operation $F_c(e^{j\omega_1}, e^{j\omega_2})$. The filtering is performed in the frequency domain via the following procedure. First, the discrete Fourier transforms (DFTs) of the LR $w_c[n_1, n_2]$ are first computed, then tiled to form their expanded versions. Each tiled version is then multiplied with the corresponding filter’s frequency response (the filters are designed directly in the frequency domain). Finally, the inverse DFT of the sum of these results is found, providing the super-resolved image. Symmetric extensions [64] of the LR images can be used to reduce DFT-produced border artifacts in the final result. The complete filter bank model is shown in Fig. 4.1 where, for brevity, the spatial and frequency indexing terms are omitted from the signal and filter labels.

The structure of Fig. 4.1 differs from the standard maximally decimated filter bank in that the number of channels is not necessarily equal to the decimation factor. One of the primary interests will be for the over-decimated case, in which $C < D = D_1D_2$, thereby prohibiting perfect reconstructing of a full-band input. The linear MMSE reconstruction solution adapted from [84] examines the input signal and filters in the frequency domain as partitioned into sub-bands by the decimation operations. These $D = D_1D_2$ rectangular sub-bands are each of area $2\pi/D_1 \times 2\pi/D_2$. A sub-band notation will be used where $H_1(W^d)$ refers to the
Figure 4.1: Filter bank model for super-resolution. A single HR image is degraded via the analysis bank to produce $C$ LR images, which are processed and combined in the synthesis bank to reconstruct the super-resolved output.

$d$th sub-band of the function $H_1(e^{j\omega_1}, e^{j\omega_2})$ shifted in frequency to occupy the base range

$$
-\pi/D_1 \leq \omega_1 < \pi/D_1,
-\pi/D_2 \leq \omega_2 < \pi/D_2.
$$

(4.4)

All HR signals and filters are considered with this sub-band notation and are only defined over the range of (4.4). The ordering of the $D$ sub-bands is inconsequential provided it is consistent for all matrices.

The solution will use the alias component (AC) matrix representations for the analysis bank,

$$
H = \begin{bmatrix}
H_1(W^1) & H_2(W^1) & \cdots & H_C(W^1) \\
H_1(W^2) & H_2(W^2) & \cdots & H_C(W^2) \\
\vdots & \vdots & \ddots & \vdots \\
H_1(W^D) & H_2(W^D) & \cdots & H_C(W^D)
\end{bmatrix},
$$

(4.5)

and the synthesis bank,

$$
F = \begin{bmatrix}
F_1(W^1) & F_2(W^1) & \cdots & F_C(W^1) \\
F_1(W^2) & F_2(W^2) & \cdots & F_C(W^2) \\
\vdots & \vdots & \ddots & \vdots \\
F_1(W^D) & F_2(W^D) & \cdots & F_C(W^D)
\end{bmatrix}.
$$

(4.6)
The PSD of the WSS input model is defined through the diagonal matrix

\[
S_{xx} = \begin{bmatrix}
S_{xx}(W^1) & 0 & \cdots & 0 \\
0 & S_{xx}(W^2) & \cdots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & S_{xx}(W^D)
\end{bmatrix}.
\] (4.7)

Since they are composed of the shifted sub-band components, these three matrices are only defined over the range of (4.4).

The additive noise components are assumed to be statistically independent from the image. The cross-spectrum of the \(k\)th and \(l\)th noise components is contracted from the full normalized frequency range to that of a single sub-band, and denoted as

\[
\tilde{S}_{n_k,n_l} = S_{n_k,n_l}(e^{j\omega_1}, e^{j\omega_2}).
\] (4.8)

The contracted noise cross-spectral density matrix is then defined as

\[
N = \begin{bmatrix}
\tilde{S}_{n_1,n_1} & \cdots & \tilde{S}_{n_1,n_C} \\
\vdots & \ddots & \vdots \\
\tilde{S}_{n_C,n_1} & \cdots & \tilde{S}_{n_C,n_C}
\end{bmatrix}.
\] (4.9)

As with the previous matrices, \(N\) is only defined over the range (4.4), in this case due to the frequency contraction. For the most common case of statistically independent white noises, (4.9) will be a constant diagonal matrix.

Frequency domain descriptions of the linear MMSE reconstruction filters are found by minimizing the output error with respect to the synthesis AC matrix (4.6). From [84], the optimal solution is found to be

\[
F_{opt} = DS_{xx}H^*(DN + H^TS_{xx}H^*)^{-1}.
\] (4.10)

In practice, the solution’s inverse is guaranteed to exist by both the presence of noise, and the combination of a non-zero PSD and unique shifts of the LR images. Although the solution (4.10) is defined as a continuous function of frequency within the sub-band range (4.4), in the practical case it is determined for a finite number of frequency samples. The solution is therefore determined for frequencies
corresponding to those of the HR image’s DFT. For example, if the desired output image has a DFT of \( Q \times Q \) uniformly spaced samples, each of its sub-bands will contain \( Q^2/D \) uniformly spaced samples. This requires (4.10) to be computed \( Q^2/D \) times, with each calculation simultaneously determining one frequency sample for each sub-band of each synthesis filter. Therefore, while the computational cost of the actual reconstruction is low, the cost of determining the reconstruction filters is generally not. However, since the reconstruction solution is determined from repetition of a single expression, it may be possible to significantly reduce the total computation time through some pipelined calculation process.

4.3.2 Fundamental Limitations of Super-Resolution in the Frequency Domain

Although the MMSE reconstruction does improve upon earlier frequency domain solutions by removing the requirement of a bandlimited signal model, there are some unavoidable limitations in any frequency domain solution.

The first limitation is the reliance on fairly detailed models. The solution requires a spectral model for the image content, and frequency domain models for the noise power, distortion, and registration. These models can certainly be estimated and provided using the approaches in the previous chapters, but this will introduce additional computational requirements, which can be undesirable in some scenarios. In contrast, some alternative approaches (details in Section 4.4) rely on inherent models, decreasing the required computation as well as the potential specificity of the model. Additionally, since the models must often be estimated from the degraded content, they are potentially inaccurate which can produce low quality or meaningless image reconstructions. Typically, the reconstruction is most robust against distortion and noise modelling errors, somewhat robust against errors in the spectral model, and quite sensitive to errors in the registration model.

The second limitation for frequency domain reconstruction techniques is a consequence of the filter bank structure Fig. 4.1, and the scenarios to which it
can be applied. Unlike the model limitation described above, which essentially limited the quality of the reconstruction based on the accuracy of the model, this limitation absolutely restricts certain scenarios from even being considered. For instance, since each LR image is modelled as passed through an LSI filter followed by a rectangular decimation operation, the structure can only be applied to sets of images with translational motion differences. Similarly, the distortion functions must be LSI, and the signal and noise are assumed to be WSS signals. While such rigid limitations will strictly prohibit the solution from being applied to certain scenarios, it is often possible to modify the scenario and force it to conform to these limits. In the practical case the very assumption than an image is WSS is an example of modifying the scenario to conform to the reconstruction formulation. As noted in the previous chapter, images are not stationary random processes, although a WSS model can describe the average content in a somewhat general sense.

Of all these restrictions, the most significant one is the requirement of a purely translational motion model. Fortunately, it is sometimes possible to modify the data to conform to this required model by manipulating the observed date. This is done through preprocessing of the individual LR images by applying transforms to determine (in the ideal case) an equivalent set of images which can be modelled through an LSI analysis filter bank. This may require reversing rotation or sheering differences and performing minor adjustments to scale. This sort of preprocessing can be applied to any super-resolution algorithm, and is in fact necessary for GSE based techniques (e.g., see such consideration in [89, 93]). Note that the preprocessing will introduce error, which can be modelled as contributing to the additive noise, but will have a detrimental effect on the final quality. In addition to preprocessing, the errors in reconstruction can be cleaned up with a postprocessing phase.
4.4 Alternative Approaches

The two categories of approaches covered in this section represent just some of the many classes of reconstruction methods available. Discussion in [4] provides additional discussion of many approaches, including projection onto convex sets, iterative back projection, and other hybrid approaches.

4.4.1 Iterative Interpolation and Deblurring via Regularization

The approach presented in [31] is representative of many similar approaches making use of regularization based iterative solution. References within [31] as well as more recent articles from those authors provide other instances of similar methods.

The iterative regularization approach relies on a two-step method: first a data fusion that places the individual pixels from the observed LR images onto the HR grid, followed by the iterative reconstruction process. Several modifications are helpful for the basic data fusion step, e.g., interpolation for jointly undersampled data sets, or median filtering for the case of joint oversampling. The result of this is to produce a “fused” HR image labelled $\hat{X}_0$, which is blurred, noisy, and contains the effects of registration errors. The recovery of an uncorrupted HR image is found through the subsequent iterative refinement with regularization.

As discussed in the literature, there are many different regularization approaches which can be applied, with [31] considering primarily $L_1$ and $L_2$ approaches. The authors ultimately settle on an approach referred to as bilateral total-variation (BTV), which is a modification of the standard total variation regularization term in (2.42). The modified BTV regularization approach has been experimentally shown to better preserve edges and finer details in iterative reconstruction, helping to reduce the effects of over-regularization. Using an $L_1$ similarity measure, this results in the regularization problem

$$\hat{x} = \arg\min_x \left( \sum_{k=1}^K \| A_k x - y_k \|_1 + \lambda \Gamma_{BTV}(x) \right),$$

(4.11)
where $\mathbf{x}$ is a vector representation of the HR image, $\mathbf{y}_k$ of the $k$th observed image, $\Gamma_{BT\text{V}}$ is the BTV regularization cost function controlled by regularization parameter $\lambda$, and $\mathbf{A}_k$ is a matrix representation for the individual distortion, motion, and sampling each LR image is subjected to. The solution to (4.11) is determined through the iterative steepest descent algorithm, using the data-fused $\hat{X}_0$ to form the initial value of $\mathbf{x}$. Additional implementation details are discussed in [31].

### 4.4.2 Direct Spatial Domain Interpolation

There is a wide variety of spatial domain interpolation approaches, and for the most part their performance is typically too low for many applications. Similarly to other methods, they rely first on a data positioning step, which seeks to combine the registered data from multiple images onto a high-resolution field of sampling positions. Unlike the data fusion step discussed in Section 4.4.1, the data positioning step can generally allow for arbitrary sampling positioning, rather than forcing all samples onto the positions corresponding to the resolution rate of the desired reconstructed image. This is a slight advantage which is typically overshadowed by the disadvantage of high susceptibility to registration errors. Further, a secondary deconvolution step is often required after the HR interpolation to undo the effects of distortion. An example spatial domain reconstruction approach based on Delaunay triangulation is given in [56].

### 4.5 Example Reconstruction Results

The results presented in this section were originally reported in [83] and [38].

#### 4.5.1 First Test

An initial demonstration of the capabilities of the proposed LMMSE reconstruction approach is performed using the airplane test image in Fig. 4.2. A simple scenario is constructed to illustrate the significant advantages of the proposed approach over prior frequency domain approaches assuming a bandlimited model.
The individual LR images will be uncorrupted by noise and free from blurring. The only differences between the LR frames will be known translational shifts. The spectrum of the scene is assumed known in this example, and determined from the original image using the periodogram estimate. In a more practical approach, the spectrum would have to be estimated through the methods in Chapter 3, reducing overall performance (see the subsequent examples).

Figure 4.2: Test image used to demonstrate improvement over prior bandlimited frequency domain reconstructions.

Four individual LR images are obtained by decimating the model scene of Fig. 4.2 by a factor of 4 horizontally and vertically. The super-resolved image is found at the model’s original resolution, representing a total undersampling factor of 4. Relative to the first LR image, the others have respective horizontal and vertical shifts of (2, 1), (3, 2), and (1, 3) pixels.

Fig. 4.3 shows a selected portion of the original scene in (a) and one of the four LR frames in (b), along with two super-resolved versions (c) and (d). The scene is assumed bandlimited and critically sampled to obtain (c), which contains significant ringing, a feature corresponding to an image being bandlimited. The proposed approach is then used with the known spectral model to obtain (d), which has significantly decreased ringing and increased readability of the plane’s number.
The PSNRs of the complete super-resolved images are 31.23 dB for reconstruction under the bandlimited assumption and 34.56 dB with the known image spectrum.

Figure 4.3: Magnified portions of test scene (a), one of four LR frames at 1/16th resolution (b), frequency-domain reconstruction under bandlimited assumption (c), the proposed LMMSE approach assuming full knowledge of the HR spectrum (d).

4.5.2 Second Test

The first examination of super-resolution performance seeks to reconstruct a known HR image from a simulated LR acquisition scenario. In this test, the
original “boat” image of Fig. 4.4(a) is first blurred by a Gaussian point spread function (PSF) with spatial width $\sigma = 0.6$ pixels. Four distinct LR images are taken from this blurred original using differing spatial shifts followed by a $\left(4,4\right)$ decimation, meaning each LR image is of $1/16$th the original size and the set of all four LR images provides only 25% of the original sample count. Relative to the sampling of the first LR image, the remaining three have translational offsets of $(1,2)$, $(2,2)$, and $(3,1)$ pixels (at the HR sampling rate). Finally, each LR image is corrupted by AWGN with a variance of 20. Fig. 4.4(b) shows the bicubic interpolation of one of these LR images to the original HR image size. The original image is of size $512 \times 512$.

The parameters affecting the blurring and noise are assumed to be unknown during the reconstruction phase. This imitates the conditions in practical super-resolution scenarios and allows a fair comparison of different reconstruction approaches, some of which do not use as specifically defined distortion and noise models as the approach presented in Section 4.3.1 does. The translational offset for each LR image is assumed known to all reconstruction methods.

This proposed MMSE reconstruction method has the best performance when the original HR image’s PSD is known. This result is shown in Fig. 4.4(c). Assuming the PSD is unknown, it is estimated from the LR set using a combined model (as discussed in the previous subsection) with 80 angular zones and the option of overlayed spectral rays. This estimate is found by numerically tuning the model parameters to best fit the PSD of the HR image reconstructed using a spatial interpolation. This first stage reconstruction is used only to recover the general image features so the PSD model can be found. Since the parametric PSD model only describes an approximation of the radial features, it can still be reasonably estimated from a lower-quality HR recovery. This model is used to produce the reconstructed image in Fig. 4.4(d). This reconstruction required no outside information except for the registration parameters. A depiction of some of these PSD functions is provided in Fig. 4.5, which shows the original HR PSD, the model PSD, and the PSD of one of the LR images.

For comparison, two alternative reconstruction methods’ results are exam-
Figure 4.4: Super-resolution comparisons for several methods from a set of four degraded LR images with known registration but unknown blurring and noise. Depicted images are: (a) original boat image, (b) one of four degraded LR images (resolution increase using bicubic interpolation for display), (c) MMSE reconstruction using original image PSD, (d) MMSE reconstruction using parametric PSD model estimated from the LR image set, (e) Delaunay triangulation based cubic interpolation using the MATLAB `griddata()` function, and (f) iterative interpolation/deblurring using $L_2$ regularization.
Figure 4.5: Comparison of spectra used in the second super-resolution test. Spectra are put through a log10 scaling to produce visually discernable details for the figure. The PSD of the original image is shown in (a), followed by the model obtained from the LR data in (b), and the PSD of one of the undersampled LR images in (c).

The first, shown in Fig. 4.4(e), is a Delaunay triangulation based cubic interpolation obtained through the MATLAB \texttt{griddata()} function. A discussion of the features of Delaunay triangulation super-resolution is in [56]. A second comparison is shown in Fig. 4.4(f), which was produced using the iterative deblurring and interpolation algorithm presented in Section II.E of [31]. An $L_2$ regularization is used with parameters: regularization factor $\lambda = 1$, step-size $\beta = 2$, and 50 iterations. There were slight contrast differences in the results produced by the regularization approach. These were adjusted by scaling the resulting HR image for each experiment so its mean value equalled that of the HR original (this was done so that PSNR tests would better represent visual quality). The contrast scalings were barely noticeable to the human observer.

Performance measures comparing the reconstruction capabilities are shown in Fig. 4.6 and Fig. 4.7. Two measurements are examined: the PSNR, and the image quality index metric [77]. The amount of noise and blurring are varied for these results. It is important to note with these numerical results that the relative performances of different super-resolution methods are dependent on the image content and acquisition scenario. Performance was examined for the proposed

\footnote{Special thanks to the MDSP research group at the University of California, Santa Cruz for providing their software.}
MMSE approach using both the true PSD and the PSD modelled from the LR set. These primary tests assumed the noise and distortion were unknown. For comparison, performance was examined in the case of a complete model for the degradation process, which greatly increased performance for both metrics. The Delaunay-based cubic interpolation and $L_2$ regularization approaches were also examined. The regularization parameter $\lambda$ was kept at 1 for all tests. While it could be re-adjusted for each instance of noise, this optimized result never found more than about 0.03 dB improvement, even as the noise was increased.

![Figure 4.6: PSNR results for several methods of super-resolution with varying amounts of AWGN. Gaussian PSF with (a) $\sigma = 0.6$ and (b) $\sigma = 1.0$.](chart)

In terms of PSNR, the performance of all methods followed the same general trend. The iterative $L_2$ regularization approach had superior performance over the other methods in the case of increased PSF spread, but lower relative performance under a narrower spread, appearing to have benefited more from an increased anti-aliasing effect of the distortion filter. The reconstruction using the proposed MMSE method is only slightly reduced when the estimated PSD model is substituted for the true HR PSD. The quality index performance follows the same general trends as the PSNR performance.

The simulations were performed via MATLAB .m script files. While this
Figure 4.7: Quality index results for several methods of super-resolution with varying amounts of AWGN. Gaussian PSF with (a) $\sigma = 0.6$ and (b) $\sigma = 1.0$.

is a relatively slow implementation, it does provide some indication of required computation time. There are two main components to the algorithm: the numerical PSD modelling process and the design and use of the reconstruction filters. In this experiment, the PSD model took approximately 20 sec to compute (a significant portion of which involved fusing the initial LR data prior to running the numerical algorithm) and the reconstruction phase took approximately 4.5 seconds to determine the final super-resolved result. In a practical implementation these computation times could be significantly reduced.

### 4.5.3 Third Test

The third super-resolution evaluation is performed using a set of LR images with an inexpensive consumer digital camera. The camera was hand-held, resulting in minor shifts between the images. Although the registration is not perfectly represented using a simple translational motion model, the non-translational components were sufficiently small to not significantly degrade the resulting reconstruction. Registration vectors were determined between a designated first image and the remainder of the set and the modelled noise component is used to account
for registration inaccuracies. Motion was determined using an exhaustive best fit search to minimize the mean squared difference between an enlarged version of the first LR image and the remainder of the set. A bilinear resolution enhancement of 20×20 was used, providing motion accuracy to a resolution of 0.05 pixels for both horizontal and vertical shifts.

A common Gaussian PSF and constant white noise level were assumed for all the LR images. Based on an initial examination of reconstruction output, the variance of the PSF was selected to be $\sigma = 1.2$ pixels. The assumed noise power was kept small, accounting only for minor quantization and low levels of registration inaccuracy. The resolution was enhanced by a factor of $2 \times 2$ for all cases. The results using the method proposed in 4.3.1 are compared with the results from the iterative interpolation/deblurring method with an $L_1$ regularization term [31]. The number of LR images used was varied from as few as two to as many as thirteen. The HR PSD model was numerically estimated for each LR image in the set and averaged to provide the final PSD. Since the individual LR images mainly differ only in translational shift, there is little difference between the estimated models. In fact, nearly the same reconstruction result can be obtained using the model estimated from only a single image.

The camera used to obtain the LR image set had variable resolution options. This allowed acquisition of a single HR image for qualitative comparison. Due to non-ideal registration, accurate quantitative comparisons are not possible. This camera-obtained HR image is shown in Fig. 4.8(a). Note that in this simulation a color image was used, and the reconstruction processes were applied separately to all three color components. The bicubic interpolation of a single frame is shown in Fig. 4.8(b).

The proposed MMSE reconstruction was examined to reconstruct from a varying number of LR source images. Each color of the image is processed separately. Interestingly, there was very little difference in performance when the number of LR images was increased beyond two. This can be primarily attributed to errors in the image registration. Unfortunately, there is little that can be done for this, as the proposed method is more reliant on accurate modelling than its
Figure 4.8: Comparison of several images: (a) camera-obtained HR image, (b) $2 \times 2$ bicubic interpolation of individual LR image, (c) proposed MMSE reconstruction from 2 LR images, (d) proposed MMSE reconstruction followed by non-linear post-processing for artifact removal, (e) $L_1$ regularization reconstruction from 2 LR images, and (f) $L_1$ regularization reconstruction from 13 LR images.
competitors. It is possible to balance the error with an assumption of increased noise, but this can also lead to excess blurring in the result. Also present in the reconstruction is a slight textured artifact, most noticeable on smooth portions of the image. By adjusting the assumed noise and distortion, it is possible to trade between reconstruction sharpness and artifact reduction, but typically some portion of the artifact will remain unless the result is made very blurry. To compensate, a non-linear post-processing stage is introduced to reduce the presence of the texture artifact found in the practical scenario. A $3 \times 3$ median filter is used, which significantly reduces the appearance of the texture artifact. Not wanting to reduce sharpness in the reconstructed image, the median filter is not applied to edge pixels or pixels adjacent to the edges (which were determined through the Canny edge detection algorithm). The reconstruction is shown both before (c) and after (d) the post-processing enhancement.

The results are compared with the $L_1$ regularization reconstruction, for which each color component was also examined separately. Two separate cases are shown: two LR frames are used to produce the result in (e) and thirteen LR frames are used to produce the result in (f). Unlike the proposed MMSE approach, there is performance increase following an increase to the number of LR sources, in spite of the modelling errors. This is one of the main advantages of an $L_1$ regularization reconstruction. There is a tradeoff in the reconstruction, between reduction of a mild artifact and reconstruction sharpness. This tradeoff is primarily controlled by the regularization parameter, which is selected in this case to provide what is deemed to be the best visual quality.

Computation time for the proposed MMSE approach is relatively low. In the case of two LR images, each color component required 14 seconds for PSD estimation and 13 seconds for the calculation and use of the reconstruction filters. When thirteen distinct LR images were used, the respective computation times jumped to 87 and 35 seconds. The post-processing algorithm required approximately 2 seconds. In many cases the same models can be used for all different color components, reducing the total required computation.
4.6 Acknowledgement

 Portions of the text of this chapter are adapted from material that has been published as: R. S. Prendergast and T. Q. Nguyen, “Improving frequency domain super-resolution via undersampling model,” in Proc. IEEE Int. Conf. on Image Processing, 2005, and material submitted for publication as: R. S. Prendergast and T. Q. Nguyen, “Spectral modelling and Fourier domain recovery of high-resolution images from jointly undersampled image sets,” under review with IEEE Trans. Image Proc., 2008. The dissertation author was the primary researcher for these publications, and the co-author directed and supervised the research which forms the basis for this chapter.
5 Video Sequence
Super-Resolution Reconstruction

5.1 Introduction

Treatment of the video super-resolution problem is essentially the same as for the still-image case. The reconstruction process takes in a set of observations (in this case degraded LR frames), estimates or is provided with necessary modelling information, and determines an HR reconstruction. The reconstruction can be either an HR version of the sequence as a whole, or a single still HR frame. For the latter case, a combination of consecutive reconstructed frames provides the desired HR sequence, although there are some potential problems that can occur if the reconstruction for each frame is evaluated independently. For the most part, sequence reconstruction methods will be based on the repetition of a single-frame reconstruction process, mainly for reasons of computational practicality. The reconstruction cost of an individual HR frame is often fairly large, and attempting to recover multiple frames simultaneously can potentially increase this computational requirement to impractical levels. Additionally, online reconstruction methods benefit from being able to reconstruct a single frame at a time, allowing the resolution enhancement process to better operate with streaming data.

The main difference between video and the still-image case is how the individual pieces of LR content are related. The assumptions stated in the previous chapter dealing with still-image reconstruction required the registration differences between individual frames to be global (generally conforming to an affine model).
Additionally, the entire set of images was required to be either simultaneously obtained, or obtained from an unchanging scene. The combination of these two assumptions meant the desired HR image was composed of samples from a single continuous 2D function of the scene’s intensity, as were the pixels composing the set of LR images. For the video case, these assumptions are typically invalid. Instead, each LR frame is composed of samples taken from a 2D sequence that changes over time. It is also possible to consider a more general still-image scenario with the video super-resolution framework, for instance a set of still LR images subject to unusual (non-affine) spatial warping may be better treated as a video reconstruction problem, since the unusual motion violates the standard assumptions of the still-image scenario.

As a result of the temporal changes in video content, there is an additional degree of difficulty involved in determining how the information from the set of LR sources can be used to estimate the HR frame. Ideally, the registration process is capable of mapping the temporal changes from one frame onto another. However, this is often impractical or even theoretically impossible to accomplish. As a result, rather than estimating a single continuous $x(u_1, u_2)$ from a set of $C$ LR images $y_c[n_1, n_2] \in x(u_1, u_2)$, $1 \leq c \leq C$, the problem becomes (for the case of single frame reconstruction) estimating $x(u_1, u_2)$ from a set of $C$ LR frames $y_1[n_1, n_2] \in x(u_1, u_2)$ and $y_c[n_1, n_2] \in x^{(c)}(u_1, u_2)$, $2 \leq c \leq C$, where each individual $x^{(c)}(u_1, u_2)$ is (hopefully) similar enough to $x(u_1, u_2)$ to be meaningful.

Most of the differences between consecutive frames can usually be accounted for by local motion models. However, there are some portions of a frame which will not correspond to any portions from neighboring frames (e.g., in the case of occlusion). In such cases, the registration model can provide an accurate description for the motion of some pixels, and the best choice can be to simply identify certain pixels as unregisterable and not include them in the reconstruction. Consequently, the amount of LR data available to produce the HR frame estimate will vary significantly per locality, dependent on the content of the video sequence. Having to account for the difficulties in local motion and the existence of unregisterable data are the primary factors making video super-resolution a more complicated
problem than traditional still-image super-resolution. Note also that even without unregisterable content, the video scenario is also more likely to be subject to registration errors, since local motion is more difficult to accurately estimate than global motion. Many of the standard video processing techniques such as block-based models will invariably misregister some pixels. Alternative approaches like object-based motion [94], mesh-based motion [95], and the particle model for video motion [96] can improve some of these problems by attempting to construct a more natural model for the temporal motion in video relating content between frames. Details of these approaches were briefly discussed in Section 2.5.2.

A final concern that is present in the video scenario is maintaining temporal consistency of the reconstructed sequence. The problem of temporal inconsistency is that a sequence can be composed of frames which are individually of high quality (by some objective or subjective criterion), but which together form a low-quality sequence. Consider as an example introducing a slight random contrast adjustment to each frame. If a single still frame is examined, the contrast difference may be easily ignorable or even unnoticeable by the viewer. However, the sequence of frames with random contrast adjustments will induce a flickering effect, which will be very distracting to the viewer. Similarly, temporal consistency issues can apply to reconstruction quality, and may be caused by the motion and distortion models, content model, the availability of data, or the reconstruction process. The practical video super-resolution solution should seek to avoid temporal inconsistencies, either by considering multiple frames jointly in the reconstruction, or by employing a method ensuring the inherent consistency present in the LR sequence is maintained in the HR equivalent.

5.2 Prior Approaches

Before presenting the proposed novel approach for video super-resolution, this section considers two earlier methods and discusses their advantages and disadvantages.
5.2.1 State-Space Implementation

An approach presented in [97], drawing from earlier work in [98,99], is based on the forward state-space model given by the equations

\[ x(t) = F(t)x(t-1) + u(t), \]
\[ y(t) = D(t)H(t)x(t) + w(t). \]

The first equation of this model (5.1) describes the change in the HR frame \( x(t) \) based on motion \( F(t) \) applied to the previous frame \( x(t-1) \), along with some innovations content \( u(t) \). The LR observation \( y(t) \) of the current HR frame is given by applying distortion operation \( H(t) \) and resolution reduction \( D(t) \) to \( x(t) \), along with additive noise given by \( w(t) \). The data from the HR and LR frames are put into vectors, such that \( Q \), the length of the observation vector \( y(t) \), equals the number of pixels in the observed frame. Assuming an \( s \times s \) resolution increase, the HR vector contains \( s^2Q \) elements. \( D(t) \) introduces the resolution reduction, and is of size \( Q \times s^2Q \). The sizes of all other components in (5.1) and (5.2) follow. The matrices \( F(t), D(t), \) and \( H(t) \) are assumed known, as are the covariance matrices for \( u(t) \) and \( w(t) \).

The approach is further simplified by an assumption that all motion is translational and all distortion is spatially invariant, which imposes specific structure on \( H(t) \) and \( F(t) \), allowing them to commute. This allows the substitution \( z(t) = H(t)x(t) \) to produce an alternative equivalent to (5.1) and (5.2):

\[ z(t) = F(t)z(t-1) + v(t), \]
\[ y(t) = D(t)z(t) + w(t), \]

where \( v(t) = H(t)u(t) \) has a determinable covariance. As in other works of those authors [31], the complete super-resolution reconstruction problem is then considered through a two-phase process: first the data is fused together and any missing pixels are interpolated to provide a blurred version of the reconstruction, then a joint interpolation/deblurring operation is performed (via a regularization solution to a MAP problem) to determine the final undistorted version of the HR reconstruction.
The data fusion is found through the Kalman filtering solution to the system of equations (5.3) and (5.4). Specifics of the implementation are found in [97]. The approach also conveniently considers an approach combining the demosaicing problem with the super-resolution. This, along with the fairly efficient online Kalman filtering implementation does make the approach attractive for use at the front end of a video acquisition system. The use of a Kalman filtering solution is advantageous for the video scenario, allowing an easily updated implementation as the number of observed frames is increased. The most significant problem with the approach is its limitation to translational motion models. As mentioned in this chapter’s introduction, the purely translational model is not applicable to standard video problems. Even modelling temporal changes with local translational models and applying the reconstruction locally is a poor choice, since the translational model is not ideal, leading to registration errors and consequential reconstruction errors. The potential for error propagation—a subject not directly covered in [97]—also comes into question.

5.2.2 Adaptive Wiener Filter Implementation

The approach considered by Hardie in [79] makes use of a spatial domain Wiener filtering solution, which is very similar to the independently derived solution of this dissertation presented in the next section. The notation used here is from [79], which differs from the notation used in a similar way in the next section. The approach is based on a spatial domain implementation of the LMMSE solution, where the estimated output vector of HR pixels $\hat{d}$ is given by

$$\hat{d} = W^T g,$$  \hspace{1cm} (5.5)

where $g$ is a vector containing LR observations. (The noiseless equivalent observations are referred to as $f$.) $W$ is determined from the problem model as

$$W = R^{-1}P,$$  \hspace{1cm} (5.6)

where $R$ is the autocorrelation matrix of the observation vector, and $P$ is the cross-correlation matrix between the observation and ideal HR vector (i.e., the standard
LMMSE solution). Implementation is performed over a small sliding window, such that reconstruction of an entire frame requires many independent evaluations of (5.5) and (5.6). Accurate models for distortion, noise, and registration are assumed known.

The entries of $R$ are populated from auto- and cross-correlation functions determined directly from the local observations. The auto-correlation function of the desired HR frame portion is assumed given by

$$R_{dd}(\tau_1, \tau_2) = \sigma_d^2 \rho \sqrt{\tau_1^2 + \tau_2^2}, \quad (5.7)$$

using a fixed constant for $\rho$. The scaling term $\sigma_d^2$ is assumed to vary locally and is estimated based on the sample variance of the local observations, through the method discussed in Section 3.3.2, (3.31)- (3.35). Recall that for $\sigma_d^2 = 1$, the function (5.7) is equivalent to (3.18) with $\rho = e^{-\alpha}$, meaning the standard isotropic correlation function for images is again being used. The necessary correlation values are then determined through

$$R_{ff}(\tau_1, \tau_2) = R_{dd}(\tau_1, \tau_2) * h(\tau_1, \tau_2) * h(-\tau_1, -\tau_2) \quad (5.8)$$

and

$$R_{df}(\tau_1, \tau_2) = R_{dd}(\tau_1, \tau_2) * h(\tau_1, \tau_2), \quad (5.9)$$

where $h(\tau_1, \tau_2)$ is the known LSI distortion function, common to all LR observations. Accounting for distortion and noise, the reconstruction matrices are determined as

$$R = R_{ff}(\tau_1, \tau_2) + \sigma_n^2 I, \quad (5.10)$$

using a known $\sigma_n^2$ and

$$Q = R_{fd}(\tau_1, \tau_2). \quad (5.11)$$

Factoring out $\sigma_d^2$ appearing within $R_{dd}(\tau_1, \tau_2)$ of both (5.10) and (5.11), the solution (5.6) is rewritten as

$$W = \left( R_{ff}(\tau_1, \tau_2) + \frac{\sigma_n^2}{\sigma_d^2} I \right)^{-1} R_{fd}(\tau_1, \tau_2). \quad (5.12)$$

In effect, this is simply a stationary signal assumption with non-stationary noise, the model for which is varied inversely to the signal’s local estimated variance.
As in the case of [97], the implementation requires that the distortion model and motion model are able to commute, generally limiting motion to a translational model\(^1\). The solution uses a spatial domain correlation model to represent the content characteristics. If the degradation model is limited to purely translational motion, a periodicity appears in the sampling pattern of the complete set, which leads to the same samples from \(R_{ff}(\tau_1, \tau_2)\) and \(R_{fd}(\tau_1, \tau_2)\) being called throughout the reconstruction. As the final solution (5.12) only changes locally through a scaling of the noise model, this can significantly reduce the required number of operations. Further reduction is found by quantizing the allowed noise to several pre-selected levels, allowing \(R^{-1}\) to be precomputed and the entire reconstruction weighting function \(W\) stored in a look-up table. This significantly benefits efficient online solution computation, although the option is only available for cases of ideal globally translational motion with a common LSI PSF shared by all observed images. Details on some minor solution specifics and the required number of computations are provided in [79].

### 5.3 Proposed Video Super-Resolution Enhancement Algorithm

The proposed reconstruction algorithm is presented in this section. Much of this discussion is taken directly from [39]. It is assumed that any required modelling information is provided from external estimation algorithms. Most importantly, this includes a registration model providing a description of localized motion from one frame to the next. This will generally be obtained directly from the LR source or possibly derived in part from the sequence’s encoded motion information. Additional modelling information will typically include noise/quantization and distortion/blurring. These can also be modelled locally, or described with generic global models (the reconstruction is more robust against noise and distortion modelling errors than it is against registration errors). A locally varying

\(^1\)Other commutable pairs also exist, such as a rotationally invariant distortion function with a rotational motion.
statistical model for the frame content is also required, and is determined through a method described below. With the modelling information in place, the resolution enhancement algorithm can be performed.

5.3.1 Design Goals

- **Reconstruction quality.** Most importantly, the estimated HR sequence should be as close as possible in visual quality to the original (or ideal) HR sequence. This entails not only determining a high-quality estimate of each individual frame, but also maintaining the temporal consistency of the reconstructed HR sequence. Individual frame quality requires accurate estimation of the high-frequency content, lost in the LR frame due to aliasing (typically also first reduced via anti-aliasing filtering). Resolution reduction is a many-to-one process, providing an ill-posed problem in the general case. However, using multiple frames to reconstruct an individual frame provides multiple descriptions of the same content with slight sub-pixel motion differences, giving additional insight for estimating the aliased content. The reconstructed frame should appear crisp and detailed (not blurry), and avoid common resolution enhancement artifacts such as jagged edges, ringing, and over-sharpening. Maintaining temporal consistency is not an especially difficult problem for a resolution enhancement process. Simply, assuming consistency of the LR sequence, the algorithm need only replicate it in the HR estimate. Most importantly, the algorithm must seek to avoid inconsistencies in the modelling information from frame to frame (especially inconsistencies in frame registration).

- **Flexible and robust reconstruction.** Two aspects of flexibility are considered: required source information and output resolution. Flexibility of source information allows the algorithm to perform the resolution enhancement using a variable number of LR frames to determine the HR output. At the bare minimum, only a single LR frame is used to determine its HR equivalent, allowing for a meaningful frame estimate in the absence of ac-
accurate frame registration. With additional registration information, content from the neighboring frames are used to improve the result, subsequent to a consistency check with the current frame. The goal of output resolution flexibility allows the algorithm to produce an enhanced image at any desired magnification factor. Alternative approaches are often limited to simple ratio enhancement factors (e.g., 4/3, 2, 3, etc.). The horizontal and vertical enhancement factors can be treated separately, allowing for resolution enhancement with aspect ratio conversion. The overall flexibility is an essential part of providing a robust reconstruction, as the method must be able to adapt to cases where registration is inaccurate, possibly adjusting how much of the provided LR data is used.

- **Minimized computational complexity.** This final goal of minimized computation complexity is considered less consequential than the first two goals. Minimized computation is sought, but generally not at the expense of significantly reduced performance. However, for the practical scenario, computation must be sufficiently small for real-time operation. There generally exists a tradeoff between quality and complexity. Quality is improved through the use of additional information from neighboring frames and through more detailed modelling of the local content statistics. However, both these measures require additional computation in determining the enhanced frame. The algorithm is capable of using less information and less detailed models to reduce complexity if necessary. Effort is also taken to reduce the required computation used in the modelling process.

To meet the design goals, a non-iterative spatial-domain linear minimized mean-squared error (LMMSE) solution is used. A spatial-domain solution is highly desirable for video data, since it easily allows for arbitrary pixel locations and works well with locally varying statistical models. Transform domain methods (e.g., frequency domain) typically do not work as well with locally varying models, and are made difficult with the non-uniformly spaced data resulting from frame registration. For a spatial-domain processing approach, the LMMSE solution provides a good tradeoff between quality and simplicity. It provides an HR interpolation
based directly on the local pixels and, unlike alternative approaches, does not require any external training or multiple iterations. It uses a simple second-order correlation model, which provides content information and is not especially costly to compute. The distortion and noise models can be incorporated directly into the LMMSE approach, so a separate post-processing enhancement phase is not required. The proposed algorithm also includes some methods for reducing the effects of registration error. The size of the area being processed can vary, allowing some tradeoff between model accuracy and algorithm speed. These features of the proposed approach meet the three main design goals.

5.3.2 Principal Algorithm

The approach presented here is very similar to the method used by Hardie in [79], but was developed independently. The main advantage of the proposed approach over Hardie’s (as well as alternative methods like [97]) is that there are no requirements for specific limitations in the motion model, or in the distortion model. Restricting the registration model of video to purely translational motion essentially allows nothing more than glorified still-image super-resolution. The approach presented here allows reconstruction in the case of truly arbitrary motion, allowing advanced registration models to be employed for true video super-resolution.

Each HR frame of the sequence is estimated separately. The corresponding LR frame is divided completely into non-overlapping core blocks, the basic structure around which the enhancement takes place. All blocks are processed using the procedure below. To lower the required computation and emphasize local model variance, the size of the core block is kept small so that only a handful of pixels from the current LR frame fall within its boundaries. A larger number of HR pixels will fall within the core block, their number and positions being dependent on the desired resolution enhancement factor. Even with constant LR block size, the number of HR pixels can vary between adjacent blocks as a fractional scaling factor might necessitate sub-pixel shifts in the relative HR pixel locations. While the estimation determines only these HR pixels within the core block, obtaining an
accurate estimate requires use of additional LR information from the area outside this block. For this, two additional block types are introduced: the source information block, which is centered on and envelopes the core block, and the correlation block, which is centered on and envelopes both the two smaller blocks. A sample depiction of these blocks superimposed over LR and HR pixel locations is shown in Fig. 5.1.

![Figure 5.1: Illustrative case of core, source information, and correlation blocks superimposed over known pixels of LR frame. The HR pixels to be interpolated correspond here to a 2.5×2.5 scaling enhancement.](image)

All LR data contained within the source information block is used to determine the LMMSE estimate. The LR data within the correlation block contributes to this estimate indirectly: it is used to determine the local statistical model required by the LMMSE solution. A larger source information block is used so that all estimated HR pixels inside the core block are sufficiently supported by all their neighboring LR pixels, not just those LR pixels falling inside the core block. The large size of the correlation block is used so that sufficient neighboring data contributes to the estimated correlation trends of the local region. Obviously, since these blocks envelope the non-overlapping core blocks which form a complete set of the LR frame, adjacent source information and correlation blocks will overlap. This contributes to a desirable effect: since much of the same data is used to estimate adjacent HR blocks (those HR pixel locations falling within the core block), block-processing artifacts are significantly diminished in the final estimated frame. In fact, proper selection of relative block sizes completely eliminates any visible blocking artifacts in the estimated HR frame (results follow in Section 5.4).

In addition to information from the current LR frame, there are also portions of the neighboring frames which can be used in the enhancement, depending
on the inter-frame registration. Pixels from the current frame are referred to as the *base information*, while those registered from the neighboring frames are referred to as the *auxiliary information* (the terms *base frame* and *auxiliary frames* are also used). The relationship between the auxiliary information and the base frame is defined by the registration. While the base information is uniformly spaced as depicted in Fig. 5.1, the auxiliary information will generally be non-uniformly spaced. Pixels from both the base and auxiliary frames that fall within a given source information block are used to interpolate the HR estimate. The local correlation model, which is constructed to direct the interpolation, is computed only from the base information falling inside the correlation block. The auxiliary information is not used to compute the model since it is inherently less accurate and is also non-uniformly spaced. The correlation model estimation procedure, discussed in Section 3.3.2, uses a Fourier-based approach which can very efficiently produce a model from regularly spaced data. The introduction of non-uniform sampling to the Fourier transform becomes problematic and is best avoided.

In considering the auxiliary information, the existence of registration error cannot be ignored. This error can come from two sources. First, content which has a correct registration location might be misregistered onto an incorrect location. This is simply an effect of estimation error and, as with the similar problem of misalignment errors in the case of still-image registration, is unavoidable. The problem can also be complicated by the existence of local motion. Whereas in the still-image case content of the entire image can be used to find global alignment parameters, in the case of video the smaller portions of the frame affected by local motion provide less information for determining the true registration. The accuracy is dependent on the particular registration method used, but the reduction of information certainly has a tendency to increase the expected error. In addition, some registration techniques (e.g., block-based motion estimation) are simply incapable of fully accurate motion descriptions.

The second source of registration error is caused by temporal changes in the sequence, which is not a concern in the still-image case. As a consequence of temporal changes to the sequence, it is possible for there to be certain portions of
the sequence which cannot be registered onto the base frame. The ideal registration in this case is to simply remove these unregisteable portions from the source data. In practice, this cannot always be accomplished and is again very much dependent on the characteristics of the particular registration algorithm used.

Since registration error is generally unavoidable, the reconstruction algorithm attempts to mitigate the effects. Error can be easily accounted for as noise with the LMMSE recovery approach (details are discussed below), improving overall performance. Since pixels can potentially be significantly misregistered, a data audit is first performed to confirm the validity of the auxiliary information. This is done by comparing the individual auxiliary frame pixels with the base frame pixels they are registered near, with a goal of removing registered auxiliary data which is significantly different from that of the base frame. The approach in this section considers a relatively simple comparison between each individual registered auxiliary pixel and a value interpolated from the base frame (alternative methods of comparison can be valid and might produce superior results). The interpolation finds an estimate of the frame value at the registered position (typically located at some fractional index). Since this estimate comes directly from the base frame data, it is free from temporal inconsistencies. However, it is still a crude estimate (a cubic interpolation implemented with a MATLAB default function is used here) and designed merely to test the validity of the registration. A more accurate estimate is not needed for this portion of the algorithm and can be computationally inefficient. The final comparison examines the absolute difference between the estimate and registered pixel, and can either discard the registered pixel if this difference is greater than some pre-selected threshold, or consider a higher level of noise attributed to the pixel in the noise model. Implementation can also be based on a combination of observation discard and a noise buffering approach.

From here, the HR estimation algorithm can be described. Per block, the source data (pixels from the base frame and non-discarded auxiliary pixels) are collected as vector $\mathbf{x}$. The unknown HR pixels to be estimated are represented as vector $\mathbf{y}$. The proposed linear interpolation solution is given by

$$\mathbf{y} = \mathbf{A}^T \mathbf{x},$$

(5.13)
for which $A$ is determined according to the MMSE criterion as

$$A = R_{xx}^{-1}R_{xy},$$ (5.14)

where $R_{xx}$ is the autocorrelation matrix of the source data and matrix $R_{xy}$ indicates the cross-correlation between the source data and HR estimate vectors. In the practical case, these required correlation values are unknown and must be estimated. The correlation coefficients are determined directly from the neighboring pixels in the correlation block (Fig. 5.1), and enable the solution to be content adaptive. While other modelling components (specifically: registration, blur, and noise) are determined externally and assumed to be provided to this proposed algorithm, this correlation model for the local content statistics is determined, per-block, alongside the HR estimate.

The use of local correlation models is somewhat complicated by the non-uniformity of the source data, $x$, and the potentially arbitrary positioning of the desired HR pixel locations, $y$, within the core block. Both these vectors are composed of samples from a single continuous 2D function, and the entries of the required correlation matrices $R_{xx}$ and $R_{xy}$ are simply the appropriately spaced values of the autocorrelation of this single function. Per block, this autocorrelation function is modelled as locally wide-sense stationary and must be estimated from the known source data. As mentioned previously, only the uniformly spaced base frame data is used to allow for more efficient computation. However, since both $x$ and $y$ can contain arbitrarily located samples, this generally requires non-uniform sub-sampling of the local correlation function. Interpolating the correlation function at random sub-intervals can be both computationally costly and prone to error, violating the intended design goals.

The proposed approach seeks to circumvent these complications by finding an equivalent representation for $x$ which is composed only of uniformly spaced samples. This is done using a linear distortion model to translate between the known data in $x$ and a new representation, $\bar{x}$. The entries of $\bar{x}$ are composed of samples located on the HR estimation grid, that is, the same samples which compose $y$ in the current and neighboring blocks. Each entry $x_i \in x$ can be
expressed as a linear combination of entries $\bar{x}_j \in \bar{x}$, or

$$x_i = \sum_j h_{i,j} \bar{x}_j. \quad (5.15)$$

The weights $h_{i,j}$ are selected according to the distortion function assumed acting on the current block. The distortion function models the anti-aliasing present in the HR to LR down-scaling, as well as any additional blurring present in the source frames. Note that the form of (5.15) is very general, allowing arbitrary distortion. The actual coefficients used are dependent on the base distortion model as well as the motion model (which can be viewed as affecting which coordinates of the PSF function the HR pixels fall on). The fact that the reconstruction requires some non-insignificant level of distortion is not a problem in practice, since some non-trivial PSF is guaranteed to exist (in the very least as a result of anti-aliasing).

Note that, although the $h_{i,j}$ coefficients can be easily determined as a function of the distance between $x_i$ and $\bar{x}_j$, this must be considered as a distance subsequent to the motion process, i.e., in the motion-space of the relevant auxiliary frame. The location of the $\bar{x}_j$ values are uniformly distributed in the motion-space of the base frame, but are possibly non-uniformly distributed in the motion-space of the auxiliary frame. While the reconstruction requires a registration mapping for the auxiliary $x_i$ pixels onto the base frame, selection of $h_{i,j}$ coefficients requires a reverse registration mapping of $\bar{x}_j$ onto the auxiliary frames. Depending on the registration model used, this can require additional computation. The forward and reverse registrations between two frames should be consistent. Note that in simplified registration scenarios like purely translational motion, identical spacing exists between $x_i$ and $\bar{x}_j$ in all motion spaces, which is the same property required in alternative methods that allows the motion and PSF functions to commute.

The vector equivalent for (5.15) is

$$\mathbf{x} = \mathbf{H}\bar{\mathbf{x}}, \quad (5.16)$$

where the individual $h_{i,j}$ make up the entries of $\mathbf{H}$. That is, the distortion of each observed pixel corresponds to one row of $\mathbf{H}$. Unlike in [79], there is no requirement for a spatially-invariant distortion function. In fact, each pixel can have a unique
PSF applied to it. All the model requires is linear distortion coefficients to fill in
the entries of \( H \). With (5.16) in place, the correlation matrices are then expressed as

\[
R_{xx} = H R_{\bar{x}\bar{x}} H^T
\]

and

\[
R_{xy} = H R_{\bar{x}y}.
\]

Since \( \bar{x} \) is a resampling onto the HR grid and \( y \) is by definition composed of HR
pixels, the new correlation matrices \( R_{\bar{x}\bar{x}} \) and \( R_{\bar{x}y} \) provide statistical information for
a single uniformly spaced 2D function. These matrices still require estimation, but
this task has been simplified from a non-uniform sampling correlation estimation
problem to a uniformly sampled correlation scaling problem. The method used to
determine the HR correlation model from the LR data was presented in 3.3.2.

Introducing (5.17), (5.18), and an additive noise term to (5.14) produces
the LMMSE solution

\[
A = (H R_{\bar{x}\bar{x}} H^T + R_{nn})^{-1} H R_{\bar{x}y}.
\]

The matrix \( R_{nn} \) provides the autocorrelation of a noise vector \( n \) which is assumed
to additively corrupt observation vector \( x \). The vectors \( n \) and \( x \) are assumed to
be statistically independent. The individual entries of \( n \) are also typically inde-
pendent of one another, providing a diagonal \( R_{nn} \). A noise model also introduces
useful properties for guaranteeing invertibility in (5.19) and improving the stability
of a nearly-singular \( R_{xx} \). In the traditional sense, \( n \) represents the effects of acquisi-
tion noise and quantization error. In this application, it is also used to indicate
uncertainty of the source information, with a larger value corresponding to pixels
with less certain registration accuracy. In this way the effects of misregistration
can be mitigated without having to fully discard erroneous observations from the
reconstruction.

The final HR block estimate is determined by applying (5.19) to (5.13). Al-
though the motivation behind (5.17) and (5.18) was to reduce complications in the
correlation estimation, there is also a positive secondary result: the local distortion
is incorporated into (5.19). The inclusion of the distortion and noise terms in the
solution allows a single HR estimation step, in contrast to alternative approaches which require separate phases for data fusion and deblurring [97]. Perhaps most interestingly, by incorporating a potentially arbitrary linear distortion model directly within the defined correlation matrices, the proposed MMSE solution is able to accommodate the temporal motion typically found in real video sequences, rather than limiting the motion model to simple cases that can commute with spatially invariant distortion models. Additionally, the proposed solution is non-iterative, making it attractive from an implementation standpoint. A discussion of computational requirements is follows.

5.3.3 Computational Analysis

The current experimental implementations are performed using MATLAB script files which, while useful for testing, are significantly slower than an optimized software or hardware implementation. With this in mind, some discussion of computation issues is presented to provide insight into the relative computational load for the algorithm, and compare the change in computation associated with changing parameters of the proposed reconstruction algorithm (e.g., the different block sizes). The result provides insight into the computational characteristics of the proposed algorithm, which serves as a useful tool for design of a real-time super-resolution implementation.

There are several different variables affecting the average computation time per-pixel:

1. Core block size, $s_c$.
2. Source information block size, $s_s$.
3. Correlation block size, $s_{AC}$.
4. Scaling factor, $m$.
5. Presence and amount of auxiliary frame data.
6. Effective blurring function size, $s_B$. 
Without loss of generality, all sizes and scaling factor are assumed to be square, and the labels above refer to the single-dimension length or value. The sizes represent the number of LR pixels spanned. The first three of these items are simply the blocks shown in Fig. 5.1. Scaling factor is dictated by the application requirements. The fifth item, auxiliary frame data, indicates the amount of additional source information which must be processed. This is largely dependent on the amount of registration information supplied to the reconstruction algorithm. The final item in the list above, blurring size, determines the extent of non-zero $h_{i,j}$ in (5.15). Typically these coefficients are taken from some assumed blurring function, which is assumed to have negligible value between distant $x_i$ and $\bar{x}_j$ and can therefore be set to zero for most $(i, j)$ pairs. An increase in the number of non-negligible weights increases the overall computation time. In the current implementation method, an entire column of $H$ is found at once by relating a single $\bar{x}_j$ with the known vector $\mathbf{x}$. Since the source pixels near the perimeter of the source information block are corrupted by blur from HR pixels outside the block, the blurring function size $s_B$ indicates the size of a blurring block larger than the source information block (but still generally smaller than the correlation block). All HR pixels falling within this blurring block contribute to the construction of $H$.

The complete computation cycle can be divided into distinct parts:

1. FFT-based correlation model construction.
2. Inverse solution computation through (5.19).
3. Blurring function construction to determine values of $H$.
4. Indexing correlation function to determine $R_{\bar{x}\bar{x}}$ and $R_{\bar{x}y}$.
5. Organizing auxiliary data to place in data vector $\mathbf{x}$.
6. Additional overhead (setup, control, and minor processes).

As will be seen, the largest portion of computation is taken up by the fourth item, indexing. For every entry of both $R_{\bar{x}\bar{x}}$ and $R_{\bar{x}y}$, this step requires first
determining the spacing between two vector points, then translating this positional difference to a location on the estimated correlation function, which supplies the appropriate value to the correlation matrix. This is essentially a doubly-imbedded indexing for relatively large amounts of data (bear in mind that the correlation matrix is not for $x$, but for all the HR pixels blurring together to form all the pixels in $x$). Additionally, there is some question on the efficiency as implemented in MATLAB.

All computation tests below examine a block of a color image containing 3240 LR pixels in the base frame. The addition of auxiliary pixels will increase the count as noted in the final test. The computation time is unaffected by the image content. Noting the list of computations above, it is seen that the majority are concerned only with the spatial location of the pixel. Consequently, there is very little increase in processing required for color images over that required for grayscale images. In fact, it has been found that a correlation model can be determined for all three color spaces from the grayscale equivalent, so the only additional requirements for color processing are to determine the base frame’s grayscale equivalent, and two additional computations of (5.13). These additional steps are contained within the “additional overhead” item listed above.

Initial comparisons are provided for the case of no auxiliary frame data, i.e., each frame being enhanced separately. In this case, all the source information is uniformly spaced, reducing the required computation. However, only an initial data organizing step is removed, and the remaining portions of the algorithm are affected only in that the size of the data vector is relatively smaller\(^2\). In present implementation, this data organizing step is probably the least optimized, relying on a direct search to determine which auxiliary pixels fall within the source data block. A more efficient implementation of this task should be easily attainable for practical use.

The effects of correlation block size are examined first. Since this step

\(^2\)It is possible to construct a faster algorithm that will never process multi-frame data, but this simplification takes away from the intended scope of the problem and will not be examined here. Even greater reductions in computation time are possible if, additionally, only integer scaling factors are considered.
Table 5.1: Required computation time for correlation estimate, varying $s_{AC}$. The combined remaining calculations require approximately 15 additional seconds.

<table>
<thead>
<tr>
<th>$s_{AC}$</th>
<th>11</th>
<th>15</th>
<th>19</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>run time (sec)</td>
<td>2.95</td>
<td>3.5</td>
<td>4.0</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Table 5.2: Required computation times (in seconds) for different components of the restoration process with varying resolution enhancement factor.

<table>
<thead>
<tr>
<th>$m$</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>correlation estimation</td>
<td>2.72</td>
<td>3.37</td>
<td>4.0</td>
<td>3.7</td>
</tr>
<tr>
<td>inverse solution</td>
<td>2.15</td>
<td>4.62</td>
<td>9.87</td>
<td>18.6</td>
</tr>
<tr>
<td>blurring matrix</td>
<td>3.21</td>
<td>4.95</td>
<td>7.21</td>
<td>9.85</td>
</tr>
<tr>
<td>indexing</td>
<td>4.28</td>
<td>11.05</td>
<td>45.61</td>
<td>103.9</td>
</tr>
<tr>
<td>additional overhead</td>
<td>0.89</td>
<td>0.96</td>
<td>1.06</td>
<td>0.96</td>
</tr>
<tr>
<td>total time (sec)</td>
<td>13.24</td>
<td>24.95</td>
<td>67.75</td>
<td>103.9</td>
</tr>
</tbody>
</table>

uses only LR data, it is not affected by the presence or absence of auxiliary data and can be analyzed separately. The correlation estimate computation depends primarily on $s_{AC}$ and secondarily on $m$. Fixing the other parameters at $s_C = 1$, $s_S = 3$, $s_B = 5$, and $m = 2.5$, the correlation block size is varied. Outside of the correlation model computation, there is only negligible variance in the total run time. The other processes amount to approximately 15 seconds of run time, and the correlation computation time is shown in Table 5.1 for varying $s_{AC}$. A significant portion of the processing is taken up by the cross-boundary blurring performed prior to the FFT to avoid the estimation of artificial horizontal and vertical edges.

Adjusting the scaling parameter affects the result primarily in that it causes an increase in the number of HR pixels falling within the core and blurring blocks. This leads to an increase in the size of $\bar{x}$, increasing the required computation time for all stages of the algorithm. Adjusting $m$ while locking the other parameters to $s_C = 1$, $s_S = 5$, $s_B = 7$, and $s_{AC} = 15$ leads to the computation times shown in Table 5.2.

Next, the size of the core block is adjusted. Note that due to the specific method of implementation, the sizes of the other blocks were adjusted along with $s_C$, such that $s_S = 2 + s_C$, $s_B = 4 + s_C$, and $s_{AC} = 14 + s_C$. The scaling factor
Table 5.3: Computation times (in seconds) and number of block iterations for processing of 3240 LR pixels, varying core block size.

<table>
<thead>
<tr>
<th></th>
<th>s_C 1</th>
<th>s_C 2</th>
<th>s_C 3</th>
<th>s_C 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>correlation estimation</td>
<td>2.95</td>
<td>0.89</td>
<td>0.50</td>
<td>0.20</td>
</tr>
<tr>
<td>inverse solution</td>
<td>2.64</td>
<td>1.04</td>
<td>1.27</td>
<td>1.45</td>
</tr>
<tr>
<td>blurring matrix</td>
<td>1.63</td>
<td>1.02</td>
<td>0.81</td>
<td>0.91</td>
</tr>
<tr>
<td>indexing</td>
<td>9.66</td>
<td>5.47</td>
<td>5.77</td>
<td>6.41</td>
</tr>
<tr>
<td>additional overhead</td>
<td>1.10</td>
<td>0.30</td>
<td>0.25</td>
<td>0.13</td>
</tr>
<tr>
<td>total time (sec)</td>
<td>17.88</td>
<td>8.72</td>
<td>8.60</td>
<td>9.10</td>
</tr>
<tr>
<td>iterations for 3240 pixels</td>
<td>3240</td>
<td>810</td>
<td>360</td>
<td>225</td>
</tr>
</tbody>
</table>

remains constant at \( m = 2.5 \) for all cases. Since only 3240 LR pixels are being processed, the number of iterations required to process all data varies with \( s_C \), which serves to reduce total computation. At the same time, the data-block sizes must increase to provide sufficient representation of the region surrounding the core block, which increases the number of computations. If the block gets very large this increase will begin to significantly outweigh the reduction (additionally, the correlation measurements are less specific for larger blocks). Results, shown in Table 5.3, indicate fastest performance with \( m = 2 \) or \( m = 3 \). For \( m \geq 4 \) the increase in required indexing will begin to grow more significantly. The relative changes will also vary with the other block sizes.

Increases in the final two size parameters, \( s_S \) and \( s_B \), tend to contribute the greatest increase in computation time. The blurring block size, \( s_B \), determines the size of \( \bar{x} \), which leads to significant increases in the sizes of correlation matrices and computation time spent on indexing, as well as a more modest increase in the size of \( H \), which increases the computation time in the blurring and inverse solution steps. Since \( s_S < s_B \), an increase in the source data block contributes indirectly to an increase in indexing time. It also leads directly to an increase in \( x \) which also requires additional computation time for the blurring and inverse solution steps. The results of several different computational tests are provided in Table 5.4. Most notably, substantial increases are seen with slight increases to the core block size.

Finally, the effects of using auxiliary frame information are examined. The same uniform LR block of 3240 pixels is used to simulate the base frame. Addi-
Table 5.4: Computation times (in seconds) for several different sets of block sizes, $m = 2.5$ for all cases.

<table>
<thead>
<tr>
<th>$(s_C, s_S, s_B, s_{AC})$</th>
<th>$(1,5,7,15)$</th>
<th>$(1,7,9,19)$</th>
<th>$(1,3,7,19)$</th>
<th>$(1,5,9,19)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>correlation estimation</td>
<td>4.0</td>
<td>4.48</td>
<td>4.77</td>
<td>4.12</td>
</tr>
<tr>
<td>inverse solution</td>
<td>9.87</td>
<td>35.82</td>
<td>6.52</td>
<td>23.88</td>
</tr>
<tr>
<td>blurring matrix</td>
<td>7.21</td>
<td>22.1</td>
<td>2.74</td>
<td>11.35</td>
</tr>
<tr>
<td>indexing</td>
<td>45.61</td>
<td>138.8</td>
<td>45.63</td>
<td>144.26</td>
</tr>
<tr>
<td>additional overhead</td>
<td>1.06</td>
<td>2.2</td>
<td>1.04</td>
<td>1.13</td>
</tr>
<tr>
<td>total time (sec)</td>
<td>67.75</td>
<td>203.4</td>
<td>61.0</td>
<td>184.74</td>
</tr>
</tbody>
</table>

$(s_C, s_S, s_B, s_{AC})$ | $(1,3,9,19)$ | $(2,6,8,16)$ | $(3,7,9,21)$ |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>correlation estimation</td>
<td>4.5</td>
<td>0.8</td>
<td>0.46</td>
</tr>
<tr>
<td>inverse solution</td>
<td>16.33</td>
<td>4.48</td>
<td>4.67</td>
</tr>
<tr>
<td>blurring matrix</td>
<td>4.29</td>
<td>3.25</td>
<td>2.47</td>
</tr>
<tr>
<td>indexing</td>
<td>101.04</td>
<td>21.77</td>
<td>16.16</td>
</tr>
<tr>
<td>additional overhead</td>
<td>1.14</td>
<td>0.66</td>
<td>0.22</td>
</tr>
<tr>
<td>total time (sec)</td>
<td>167.3</td>
<td>30.76</td>
<td>23.98</td>
</tr>
</tbody>
</table>

Table 5.5: Computation times (in seconds) for reconstructions with and without auxiliary frame data.

<table>
<thead>
<tr>
<th></th>
<th>base frame</th>
<th>base and auxiliary frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>auxiliary data sorting</td>
<td>0</td>
<td>19.0</td>
</tr>
<tr>
<td>correlation estimation</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>inverse solution</td>
<td>9.87</td>
<td>25.0</td>
</tr>
<tr>
<td>blurring matrix</td>
<td>7.21</td>
<td>21.8</td>
</tr>
<tr>
<td>indexing</td>
<td>45.61</td>
<td>44.1</td>
</tr>
<tr>
<td>additional overhead</td>
<td>1.06</td>
<td>1.2</td>
</tr>
<tr>
<td>total time (sec)</td>
<td>67.75</td>
<td>115.1</td>
</tr>
</tbody>
</table>

tionally, 12033 non-uniformly spaced pixels were introduced throughout the block. With $m = 2.5$, the density of source information is approximately 0.753 samples per HR pixel. The reconstruction parameters were selected as $s_C = 1$, $s_S = 5$, $s_B = 7$, and $s_{AC} = 15$. A comparison is shown in Table 5.5. The addition of auxiliary data increases computation time for constructing the blurring matrix and calculating the inverse solution. Both these increases are associated with an increase in the size of $\mathbf{H}$ resulting from a large $\mathbf{x}$. The auxiliary data also required sorting to determine, for each iteration, which specific values belong in $\mathbf{x}$.  

5.4 Simulation Results

Initial simulations are performed on a portion of the mobile sequence, which contains various local motion components in combination with a global zoom-out. The original sequence was reduced from CIF resolution of $352 \times 288$ to QCIF resolution of $176 \times 144$. The LR sequence was then used to determine an estimate of the HR sequence, attempting to remove the effects of anti-aliasing and return to the original resolution. An $8 \times 8$ block-based inter-frame motion estimation was determined to map content from all the auxiliary frames onto the base frame. It should be noted that this provides a highly inaccurate registration, especially since the block size is a fairly significant portion of the LR frame size. Quarter-pixel resolution was used in the motion estimate, corresponding to half-pixel resolution of the original HR sequence, further preventing high levels of accuracy.

Reconstruction parameters were selected as $(s_C, s_S, s_B, s_{AC}) = (3, 5, 9, 21)$. A global Gaussian distortion model was assumed, with $\sigma = 0.9$. The threshold for auxiliary data discard was set at 50 (on a $[0, 255]$ scale), meaning each registered pixel was used only provided that $\phi$, the difference between its value and the value of its interpolated equivalent, was no less than 50 (averaged over all three color components). This is actually a very high threshold, and admits a large portion of auxiliary data, much of which is significantly misregistered. To counteract registration artifacts in the reconstruction, the noise term of each auxiliary pixel was made proportional to $\phi$. Referring to (5.19), this means the diagonal entries of $R_{nn}$ corresponding to highly misregistered pixels are very large, effectively diminishing their contribution to the reconstruction. Interestingly, modelling the poorly registered pixels as highly noisy produced noticeably superior results to dimply discarding them using a lower threshold, although the advantage was lost if the threshold grew too large.

A single frame was first enhanced using ten adjacent frames (the five immediately before and after). Some results are shown in Fig. 5.2, where (a) and (b) depict the some of the LR and the target HR frames. Obvious motion differences are present in the three LR figures shown in (a), but the content is otherwise sim-
ilar in each frame. The proposed reconstruction method was used to produce the HR frame estimates in (c), which used only the base frame, and in (d), which used all eleven frames. Using additional frames provided a sharper result with reduced ringing. The PSNR for these reconstructed frames are 20.97 dB and 21.73 dB, respectively. Admittedly, PSNR does not necessarily correspond to visual quality. In fact, it was possible to further increase the PSNR in the case of multiple frames by reducing the assumed level of noise, but this contributed to an increase in misregistration artifacts. PSNR results also vary for depending on $\sigma$, the level of assumed blur. A global value was used, and further improvements may be possible using different $\sigma$ depending on the type of local content (e.g., decreased at smooth locations near edges to reduce ringing artifacts). For comparison, a spatial reconstruction was performed using the MATLAB griddata() function which implements a Delaunay triangulation-based cubic interpolation. The results are shown in (e) using a single frame (PSNR of 20.95 dB) and (f) using all eleven frames (PSNR of 20.31 dB). The cubic interpolation does not provide the robustness of the proposed method and its reconstruction clearly contains significant registration artifacts.

Fifty frames of the mobile sequence were reconstructed to examine the presence of temporal inconsistencies. This sequence reconstruction used only four adjacent auxiliary frames registered onto each base frame, slightly reducing the reconstruction quality in comparison to the results of Fig. 5.2. PSNR as a function of frame index is shown in Fig. 5.3.

These results show the effectiveness of the proposed reconstruction method and demonstrate its robustness in the presence of low-quality registration. This high level of robustness makes the method attractive for practical scenarios, but does not demonstrate its full capabilities. To do so, a synthetic problem is constructed in which a very high resolution image is blurred and non-uniformly sampled to simulate perfect registration in a video scenario. The non-uniform sampling consists of two sub-samplings. First, a low-rate uniform sampling representing the uniformly spaced pixels of the base frame, and second, a set of random sampled representing the auxiliary data. Unlike the previous simulation, the correct positioning (or registration) of the auxiliary data is perfectly known, allowing for the
Figure 5.2: Frame super-resolution from eleven consecutive LR frames, three of which are shown in (a): frame 10 in the upper-left, frame 20 in the upper-right, and the base LR frame 15 in the bottom. The original HR frame 15 is shown in (b). The proposed method is applied to the base frame alone to produce (c) and with the ten auxiliary frames to produce (d). Comparison with the MATLAB Delaunay triangulation-based cubic interpolation is shown using the base frame alone for (e) and data from the ten auxiliary frames to obtain (f).
possibility of very high quality reconstructions. Maintaining a constant uniform sampling (the base frame), the average overall sampling density is increased by increasing the number of auxiliary samples. Although no motion is directly considered, the LSI PSF used here would be required to commute with the noise. Reconstruction is then performed and the HR estimate is compared to the HR original to produce the results of Fig. 5.4. The quality of the proposed method steadily increases with an increase in sampling density. For comparison, the MATLAB cubic interpolation is again used. Since this interpolation does not deblur the result, a separate deblurring phase is applied subsequent to the interpolation, providing an increase in PSNR and visual quality. The cubic interpolation results also increase with an increase in sampling density, but not as significantly as with the proposed method. The frames are increased in size by a factor of $2.5 \times 2.5$, and are shown in Fig. 5.5. At the lowest sampling density, there is little difference in quality between the proposed method (c) and the cubic interpolation (e). However, at the highest sampling density, the proposed reconstruction (d) is of significantly higher quality for all types of content (e.g., edge, texture, and writing) than the cubic interpolation (f). In fact, there is very little difference between the cubic interpolation reconstructions for both low (e) and high (f) sampling densities.

Finally, a series of simulations are presented to demonstrate the ability of this proposed method to be applied to arbitrary motion scenarios, rather than lim-
Figure 5.4: Reconstruction PSNR as a function of average sampling density, simulating the performance in the case of ideal registration.

Limited to purely translational cases. Comparisons between the proposed approach and the similar reconstruction approach of Hardie [79] are also provided. Exact registration information is assumed known in all cases. Three registration scenarios are explored: purely translational global motion, simulated zoom, and simulated optical flow (per-pixel motion). When possible, comparisons with the Hardie approach provide the same a priori information for both methods. For the purely translational case, the methods are identical except for the fact that local statistics are modified in the proposed approach using the LR-to-HR correlation conversion of 3.3.2, while the Hardie reconstruction scales the noise based on the local sample variance. In simulations presented in [79], Hardie quantized the assumed noise function to 20 levels. Simulations here did not do this, which introduced a small problem: as the noise in (5.12) was related inversely to the estimated local variance $\sigma_d^2$, image portions with little variance (i.e., nearly constant intensity) would often be reconstructed under an assumption that noise power was much greater than signal power, leading to severe reconstruction artifacts. Quantization of $\sigma_d^2$ avoids this problem, provided the minimum value is sufficiently large. In the absence of quantization, the implementation provided here assumes a floor value on the estimated $\sigma_d^2$, sufficiently large to avoid any noticeable reconstruction artifacts. Otherwise, the implementation of the standard Hardie approach is effectively the same.
Figure 5.5: Reconstruction results for simulated ideal registration. LR base frame (a) and desired HR equivalent (b). The proposed reconstruction at 0.16 samples/HR pixel (c) and 1.3267 samples/HR pixel (d). MATLAB cubic reconstruction with deblurring at 0.16 samples/HR pixel (e) and 1.3267 samples/HR pixel (f). Results (c) and (e) use only the uniformly spaced base frame pixels from (a).
The first motion simulation considered is the purely translational motion. The original Athens image from Fig. 5.6 is distorted using a Gaussian PSF with covariance matrix
\[
\begin{bmatrix}
1.5 & 0 \\
0 & 1.5
\end{bmatrix}.
\]
Three LR frames are taken from this distorted image with a (3, 3)-fold decimation, after introducing translational motion to two of the LR images. This simple motion model is directly commutable with the distortion. A small level of noise \(\sigma_n^2 = 0.15\) was assumed to exist in all observations. Super-resolution is performed using the Hardie reconstruction and the proposed method, in each case from both a single LR observation frame and all three LR observation frames. The proposed reconstruction uses the block size parameters: \(s_C = 1\), \(s_S = 5\), \(s_B = 7\), and \(s_{AC} = 15\). The Hardie reconstruction uses the equivalent of \(s_C = 1\) and \(s_S = 7\), and does not have a definition for \(s_B\) and \(s_{AC}\) (these sizes are not required for this different method). These reconstructions are shown in Fig. 5.7. The scenarios and their reconstruction performances (in PSNR) are:

(a) Hardie reconstruction from three LR frames, 28.89 dB.
(b) Hardie reconstruction from a single LR frame, 33.55 dB.
(c) Proposed reconstruction method from three LR frames, 29.06 dB.
(d) Proposed reconstruction method from a single LR frame, 33.24 dB.

From the recovered images in Fig. 5.7 and their PSNR measurements, the two reconstruction approaches are shown to have very comparable levels of performance. There are two main differences in the approaches. First, the Hardie approach used a large observation window. This was deemed necessary due to a regularly occurring reconstruction artifact which appeared in the reconstruction if the window were too small. This artifact was likely related to an insufficient number of observation samples (in [79], similarly large observation windows were used, even in cases where comparisons with alternative approaches relied on smaller windows). This is simply a difference in implementation, which can be adjusted for. The second key difference was in the method used for estimating the unknown
HR correlation information. The proposed approach used a model adjusted by the content, whereas Hardie’s approach uses a stationary model assumption (and adjusts the relative noise for each reconstruction block). In spite of this seeming advantage, the proposed approach does not offer a significant difference in PSNR performance. However, it may provide more significant qualitative improvements along edges, similarly to as discussed in [78].

The second reconstruction simulation uses a zoom model, providing the three LR frames shown in Fig. 5.8. Other than the motion model, all other degradation and reconstruction conditions are identical to the previous test. The Hardie reconstruction is inapplicable to the case of a scaling registration model, and is thus only applied to the single frame case. Reconstructions shown in Fig. 5.9 and their PSNR values are:

(a) Proposed reconstruction method from a single LR frame, 30.52 dB.
(b) Proposed reconstruction method from three LR frames, 32.84 dB.
(c) Hardie reconstruction from a single LR frame, 30.35 dB.

The original elephant image is provided in Fig. 5.9(d). The reconstruction shown in Fig. 5.9(b) has significant improvement to the central portion of the HR image, corresponding to the area falling under the span of multiple LR frames.
Figure 5.7: Reconstructions of the Athens image (Fig. 5.6) determined through: (a) Hardie reconstruction using three LR frames, (b) Hardie reconstruction using a single LR frame, (c) proposed reconstruction using three LR frames, and (d) proposed reconstruction using a single LR frame.
Figure 5.8: Three LR images simulating zoom of the original elephant image. A simulated zoom factor of 1.2 is present between (b) and (a), and between (c) and (b).

The final reconstruction simulation seeks to mimic an optical flow registration, in which each individual pixel of the second frame has a unique motion assigned to map it onto the first frame. Assuming the original image was indexed with \((0, 0)\) corresponding to the upper left corner and \((1, 1)\) corresponding to the lower right corner, the motion was determined by the equations

\[
\begin{align*}
x_2 &= 0.9x_1 + 0.1y_1, \\
y_2 &= 0.2x_1 + 0.8y_1.
\end{align*}
\]

Gaussian distortion was given for the first frame by the covariance matrix

\[
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\]

and for the second frame (subsequent to motion) by the covariance matrix

\[
\begin{bmatrix}
1 & 0.4 \\
0.4 & 1.1
\end{bmatrix}.
\]

Since the motion will not commute with LSI distortion it is unable to be considered with the Hardie approach and other alternative. A set of uniform samples were taken from each distorted image, providing the set of LR observations, each LR image is a quarter of the size of the original HR image. Reconstruction block size
Figure 5.9: Reconstructed images attempting to recover original elephant image: (a) proposed reconstruction using a single LR frame–Fig. 5.8(a), (b) proposed reconstruction using all three LR frames from Fig. 5.8, (c) Hardie reconstruction using a single LR frame, and (d) original HR image.
Table 5.6: Summary of reconstruction PSNR measurements in the case of ideal registration.

<table>
<thead>
<tr>
<th>Image (motion type)</th>
<th>frames</th>
<th>PSNR (dB), proposed</th>
<th>PSNR (dB), Hardie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Athens (translational)</td>
<td>1</td>
<td>29.06</td>
<td>28.89</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>33.24</td>
<td>33.55</td>
</tr>
<tr>
<td>Elephant (zoom)</td>
<td>1</td>
<td>30.53</td>
<td>30.35</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>32.21</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>32.84</td>
<td>N/A</td>
</tr>
<tr>
<td>Einstein (optical flow)</td>
<td>1</td>
<td>32.46</td>
<td>32.22</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>34.78</td>
<td>N/A</td>
</tr>
</tbody>
</table>

parameters are the same as considered in the previous two simulations. The two LR images are shown in Fig. 5.10. The original image is shown in Fig. 5.11(a), along with reconstructions:

(b) Hardie reconstruction from a single LR frame, 32.22 dB.

(c) Proposed reconstruction method from a single LR frame, 32.46 dB.

(d) Proposed reconstruction method from two LR frames, 34.78 dB.

Figure 5.10: Pair of distorted LR images used for the optical flow simulation.

A summary of the PSNR results comparing the proposed method to that of Hardie is provided in Table 5.6. Note that all these tests relied on the same reconstruction block sizes: $s_C = 1$, $s_S = 5$, $s_B = 7$, and $s_{AC} = 15$ for the proposed method, and $s_C = 1$ and $s_S = 7$ for the Hardie reconstruction.

Finally, a simulation is performed to examine the improvement of reconstruction relative to the size of the observation window, $s_S$. The rest of the block sizes are fixed at $s_C = 1$, $s_B = s_S + 2$, and $s_{AC} = 23$. $s_S$ is then increased in
Figure 5.11: Original Einstein image (a), along with reconstructions: (b) Hardie reconstruction using a single LR frame, (c) proposed reconstruction using a single LR frame, and (d) proposed reconstruction using both LR frames from Fig. 5.10.
Table 5.7: Reconstruction PSNR as size of observation window is increased.

<table>
<thead>
<tr>
<th>Source data block size, $s_S$</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB), (1 frame)</td>
<td>31.3945</td>
<td>32.4575</td>
<td>32.7117</td>
<td>32.7597</td>
</tr>
<tr>
<td>PSNR (dB), (2 frames)</td>
<td>33.3646</td>
<td>34.7020</td>
<td>34.9043</td>
<td>34.9784</td>
</tr>
</tbody>
</table>

size, which introduces more data to the block reconstruction without changing the size of the reconstructed portion. As discussed in Section 5.3.3, too large an increase in the observation window can significantly increase the required computation time. Thus, if the benefit of an increased reconstruction window size is minimal beyond a particular point, this suggests a good size for practical implementation. The results are summarized in Table 5.6, using the Einstein image with the simulated optical flow registration. PSNR results for both single- and dual-frame reconstructions are shown. Although the results are content dependent, the trend shows a very minimal increase from $s_S = 7$ to $s_S = 9$. The reconstructed images are provided in Fig. 5.12, where (a)-(d) show the single frame reconstruction with respective $s_S = 3, 5, 7, 9$ and (e)-(h) show the two frame reconstruction with respective $s_S = 3, 5, 7, 9$.

5.5 Acknowledgement

Portions of the text of this chapter are adapted from material that been submitted for publication as: R. S. Prendergast and T. Q. Nguyen, “Digital video super-resolution,” under review with IEEE Trans. Image Proc., 2008. The dissertation author was the primary researcher for this publication, and the co-author directed and supervised the research which forms the basis for this chapter.
Figure 5.12: Einstein reconstructions with varying observation window sizes. Single frame reconstructions are shown in (a)-(d) with respective $s_S = 3, 5, 7, 9$ and two-frame reconstructions are shown in (e)-(h) with respective $s_S = 3, 5, 7, 9$. 
6 Summary of Contributions

Within this dissertation, the super-resolution problem was considered using a two-step process, with separate phases contributing to modelling and reconstruction. Under this framework, linear minimum-mean squared error reconstruction solutions were considered. The selection of LMMSE reconstruction approaches in turn dictated specific forms for the content and the degradation models. Contributions of this dissertation were in the areas of content modelling and MMSE reconstruction approaches. With the increasing prevalence of both high-resolution display systems and competition for transmission resources, the contributions of the research presented in this dissertation should prove useful over a wide range of applications.

6.1 Contributions

Spectral modelling for images. Two classes of content modelling were contributed in Chapter 3: stationary spectral density models for the global case, and locally stationary correlation models for use with a localized reconstruction approach. For the stationary case, the common features observed in image spectra were determined, and two separate parametric models were developed to provide an efficient description for an image’s average features. Ultimately, the models were combined to take advantage of the unique features each described individually. For the purpose of resolution enhancement, the most important feature of the proposed models was an ability to estimate the higher frequencies required for an HR description, but lost to the LR ob-
servations. This was based on the observation that image spectra followed radial trends, and the development of an estimation process which enabled extension of the calculated low-frequency content description to the missing higher frequencies.

For the case of locally stationary models, a correlation description was required, but estimation was performed through a familiar frequency domain description. This solution combined observed low-frequency content with a smooth model for high-frequency content, providing a stationary spectral description for a localized region. The equivalent correlation domain model was then determined from the estimated spectral description. This approach essentially modified the previously developed stationary spectral modelling approach for use as a local content model. With the relative shortage of observed information present in the local scenario, the local model is more heavily based on assumptions than its global counterpart, but at the same time it provides a better description since the local model is based only on local observations.

**Still-image super-resolution reconstruction.** The contributions toward still-image super-resolution reconstruction primarily stemmed from the application of an MMSE solution to the undersampled filter-bank problem to the image scenario. The approach required use of very specific models, which largely restricts the applicability of the solution in certain cases. The method was found to be most useful for collectively undersampled sets of LR images. The approach used an efficient non-iterative frequency-domain solution, although the need to estimate specific content and degradation models cannot be discounted in the overall computation requirements of the solution. The main limitations of the approach were the requirements for an HR spectral model based only on LR observations, and for a degradation process that conformed to the structure of the analysis filter-bank. While this first limitation was solved through the development of new image spectral models, the second limitation cannot be directly avoided due to the inherent structure of the problem description. Specifically, the limitations for purely transla-
tional motion and strictly rational factors of resolution enhancement were most limiting. Although the possibility of resampling the observations to conform within the limits of the structure has been briefly investigated, this typically introduces unacceptably large error in most applications. Reconstruction results were shown to be comparable with alternative approaches in the case of high degradation modelling accuracy, however performance did decrease significantly in the presence of registration errors. The addition of simple secondary non-linear post-processing tools (e.g., median filtering) was demonstrated to be helpful for improving performance in such non-ideal scenarios.

**Video sequence super-resolution reconstruction.** The proposed video sequence reconstruction solution was designed to improve upon many of the deficiencies found in the still-image solution. First, the reliance on a global content model was eliminated and a local equivalent was employed through a correlation domain representation, with a spatial domain reconstruction implementation. Under a localized reconstruction approach, the degradation model can also change locally, allowing spatially variant distortion, non-stationary noise, and arbitrary sampling patterns (general motion). The proposed reconstruction approach was designed to maintain the temporal consistency of the observed LR sequence. This was done by cross-checking all observations from additional frames with the known data in the current frame, either eliminating additional frame data entirely or assigning it extra noise. Since registration is non-ideal in practice, this step ensures that severe misregistration is sufficiently guarded against. Finally, the proposed approach introduced a resampling to the correlation matrices based on the linear distortion, mapping the observation-aligned correlation matrices to matrices constructed from the estimated uniformly sampled correlation information. Further, by incorporating the distortion with the correlation matrices, the LMMSE solution does not require realignment of the observation pixels onto a uniform grid like some alternative approaches. Incorporating the distortion into the correlation matrices also makes the proposed approach
truly applicable to video super-resolution scenarios, rather than limited to very specific cases (i.e., purely translational motion models).

6.2 Importance of the Contributions and Extensions

In comparison with alternative contemporary approaches, the methods in this dissertation offer some distinct advantages. The stationary content models provide the most specific image spectra descriptions in parametric models. The approaches also demonstrated how unobserved HR spectral content can be estimated with relatively high accuracy allowing for improved reconstruction performance. Similar advantages were provided for the local correlation model, although the advantage is certainly less pronounced due to the low number of observation pixels. Reconstruction results for the still image case were of marginal importance. The method was shown to significantly outperform prior Fourier based techniques, which have been otherwise largely abandoned in the face of newer approaches. However, the alternative reconstruction approaches do not always suffer from the same limitations as the proposed approach, and are thus often better suited in many applications. In contrast, the proposed video sequence reconstruction approach offers many advantages over prior methods, and can be practically applied to real-world video scenarios.

Future work in these topics should primarily be concentrated in improving the video reconstruction. This is motivated by the novelty of the result and its increased applicability over prior approaches. The widespread presence of digital video systems in modern life provides significant interest for the subject. Extensions to the presented results in digital video super-resolution reconstruction should seek to improve the quality of reconstruction and the computational efficiency of the implementation. The reconstruction approach should also seek to better provide an accurate reconstruction in the presence of uncontrollable modelling errors, of which inaccurate registration represents the most significant bottleneck for further improvements to reconstruction quality.
A MMSE Solution to Generalized Undersampling Problem

Portions of this appendix are taken directly from [84].

The problem seeks a solution to a generalized sampling problem with sub-Nyquist sampling densities. This is an extension to the traditional Papoulis generalized sampling expansion [81], which demonstrated a signal could be recovered from the set of $C$ uniform samplings, each having been passed through a unique linear filter and sampled at $1/C$th the Nyquist rate. Papoulis provided a recovery solution to the problem and showed that the solution required the set of linear filters be a linearly independent set. Later, established results in multi-rate filter bank theory allowed an equivalent discrete-time solution to be formulated as a maximally decimated filter bank [20]. The discrete-time formulation enables practical implementation of the GSE, where the sampling process is modelled through a set of analysis filters, $[H_0(e^{j\omega}), \ldots, H_{C-1}(e^{j\omega})]$, and the reconstruction process is found through the set of synthesis filters, $[F_0(e^{j\omega}), \ldots, F_{C-1}(e^{j\omega})]$, which are designed based on the analysis filters to determine a perfect reconstruction (PR) solution.

The proposed generalization to the problem is also formulated through a filter bank model, given by the structure shown in Fig. A.1. Unlike the traditional GSE, the proposed approach considers the scenario where $C$ uniform samplings are each sampled at $1/D$th the Nyquist rate, which for $C < D$, leads to an under-
sampled signal. It can be shown through filter bank theory that no PR solution to this problem exists. The proposed solution therefore seeks to determine an MMSE solution under a stochastic signal assumption, rather than a PR solution under a deterministic signal assumption. This provides a solution for the optimal set of synthesis filters as a function of the second order statistics of the input signal and the transfer functions of the set of analysis filters.

Many prior works also investigated the filter bank solutions applied to stochastic signaling problems [19, 100–105]. The proposed approach will seek the optimal linear solution that minimizes the mean of $|e[n]|^2 = |x[n] - y[n]|^2$, providing a Wiener filtering solution for the generalized sampling problem. The remainder of this appendix is organized as follows. First, preliminary setup of the problem is provided in Section A.1. The optimal synthesis filter solution to the problem is given in Section A.2. Finally, discussion of some additional considerations and extensions to the basic problem are provided in Section A.3.

### A.1 Problem Setup

The definitions presented here are based on the analysis developed by Ohno and Sakai in [101] for the purpose of minimizing a filter bank’s time-averaged MSE (TAMSE) when a subband is lost. The approach considers a discrete-time cyclo-stationary input even though primary interest will be for wide-sense stationary (WSS) inputs. The reason for this is that even when the filter bank is provided a
stationary input, it will generally produce an output that is cyclostationary with period $D$ (corresponding to the order of decimation). Because of this, the error between the input and output signals is cyclostationary. For generalization purposes, the input is considered so as well (since stationarity is a subset of cyclostationarity). Specifically, $x[n]$ is wide-sense cyclostationary with period $D$, or WSCS(D), meaning its first- and second-order expectations are periodically stationary with period $D$. This is expressed as

$$E[x[n]] = E[x[n + kD]] \quad \text{and} \quad R_{xx}[n, m] = R_{xx}[n + kD, m + kD], \quad (A.1)$$

for every integer $k$, where $R_{xx}[n, m] = E[x[n]x^*[m]]$. Herein, a zero-mean process is considered.

The cyclic correlation function is defined as

$$R_{\alpha xx}[u] = \frac{1}{D} \sum_{k=0}^{D-1} R_{xx}[k + u, k]e^{-j2\pi \alpha k}, \quad (A.2)$$

for $\alpha = n/D$ and integer $n$ (the notation used in [101] is based on that of [106]). Since the exponential term is periodic with $\alpha$, (A.2) only needs to be determined for $0 \leq \alpha \leq (D - 1)/D$ (or, $n = 0, 1, \ldots, D - 1$) to be defined for all valid $\alpha$. The discrete-time Fourier transforms of the cyclic correlation functions produce the cyclic spectral densities (CSDs) through

$$S_{\alpha xx}(e^{j\omega}) = \sum_{u=-\infty}^{\infty} R_{\alpha xx}[u]e^{-j\omega u}. \quad (A.3)$$

Through (A.2) and (A.3), the time-averaged variance of a WSCS(D) process is defined as

$$\sigma_x^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} E[x[n]x^*[n]] = \frac{1}{D} \sum_{n=0}^{D-1} R_{xx}[n, n]$$

$$= R_{xx}^0[0] = \frac{1}{2\pi} \int_0^{2\pi} S_{xx}(e^{j\omega}) d\omega. \quad (A.4)$$

To determine a filter bank’s CSD input/output relationship, a matrix representation for the CSDs of (A.3) is defined. This $D \times D$ CSD matrix is expressed as $S_{xx}(e^{j\omega})$, and the $(p, q)$th element is defined as

$$[S_{xx}(e^{j\omega})]_{(p,q)} = S_{xx}^{(p-q)/D}(e^{j\omega}W^p) \quad \text{over} \ |\omega| \leq \pi/D, \quad (A.5)$$
for \( p, q = 0, \ldots, D - 1 \), using the standard definition of \( W = e^{-j(2\pi/D)} \). This representation is only valid for \( |\omega| \leq \pi/D \), however this is sufficient to completely represent all \( D \) CSDs of (A.3). Since \( R_{xx}[u] = 0 \) for all valid \( \alpha \neq 0 \) if and only if \( x[n] \) is WSS, it is easily verified that a diagonal \( S_{xx}(e^{j\omega}) \) is a necessary and sufficient condition for a WSS input.

The analysis and synthesis filters are represented using alias component (AC) matrices [20], defined as

\[
H_{AC}(e^{j\omega}) = \begin{bmatrix}
H_0(e^{j\omega}) & \cdots & H_{C-1}(e^{j\omega}) \\
\vdots & \ddots & \vdots \\
H_0(e^{j\omega}W^{(D-1)}) & \cdots & H_{C-1}(e^{j\omega}W^{(D-1)})
\end{bmatrix}
\]

for the analysis bank and an equivalent representation \( F_{AC}(e^{j\omega}) \) for the synthesis bank. Like the CSD matrix, these AC representations are also only considered over \( |\omega| \leq \pi/D \) but completely and uniquely describe the analysis and synthesis filters in the frequency domain. Defining the matrix product

\[
P(e^{j\omega}) = \frac{1}{D}F_{AC}(e^{j\omega})H_{AC}^T(e^{j\omega}),
\]

CSD matrices relating the input \( x[n] \) and output \( y[n] \) are defined as

\[
S_{yy}(e^{j\omega}) = P(e^{j\omega})S_{xx}(e^{j\omega})P(e^{j\omega}),
\]

\[
S_{yx}(e^{j\omega}) = P(e^{j\omega})S_{xy}(e^{j\omega}),
\]

\[
S_{xy}(e^{j\omega}) = S_y e^H(e^{j\omega})P(e^{j\omega}),
\]

The superscript \( H \) denotes the conjugate transpose. The CSD matrix corresponding to \( e[n] = x[n] - y[n] \), the reconstruction error of the filter bank, is found to be

\[
S_{ee}(e^{j\omega}) = S_{xx}(e^{j\omega}) - S_{yx}(e^{j\omega}) - S_{xy}(e^{j\omega}) + S_{yy}(e^{j\omega})
\]

\[
= S_{xx}(e^{j\omega}) - P(e^{j\omega})S_{xx}(e^{j\omega})S_{xx}^H(e^{j\omega})P(e^{j\omega})
\]

\[+P(e^{j\omega})S_{xx}(e^{j\omega})P^H(e^{j\omega}),\]

which represents the error as a function of the filter bank and the input signal statistics.
A.2 Synthesis Filter Optimization

This section determines the frequency domain representation for the synthesis bank for minimizing the system’s MSE where the input signal statistics and sampling process (modelled through the analysis bank) are assumed known. The approach is especially of interest for the case of $D > C$, which generally results in a loss of information, thereby negating the possibility of a PR solution.

The time-averaged mean-squared error (TAMSE) is determined through an equivalent to (A.4) as

\[ \sigma_e^2 = \frac{1}{2\pi} \int_0^{2\pi} S_{ee}(e^{j\omega}) d\omega = \frac{1}{2\pi} \sum_{k=0}^{D-1} \int_{-\pi/D}^{\pi/D} S_{ee}(e^{j\omega}W^k) d\omega \]  

(A.9)

\[ = \frac{1}{2\pi} \int_{-\pi/D}^{\pi/D} \text{tr}(S_{ee}(e^{j\omega})) d\omega. \]  

(A.10)

Thus, the TAMSE is a function of the trace of (A.8). Since $S_{ee}(e^{j\omega})$ is strictly non-negative, the minimization of $\sigma_e^2$ is equivalent to the minimization of $\text{tr}(S_{ee}(e^{j\omega}))$. Since (A.8) is only defined for $|\omega| \leq \pi/D$, the minimization of (A.10) with respect to the synthesis AC matrix can be stated as

\[ \arg\min_{F_{AC}(e^{j\omega})} \text{tr}(S_{ee}(e^{j\omega})) \text{ for all } |\omega| \leq \pi/D. \]  

(A.11)

The solution to (A.11) is found by setting the derivative of $\text{tr}(S_{ee}(e^{j\omega}))$ with respect to $F_{AC}(e^{j\omega})$ equal to zero (the relevant operations can be found in Appendix E of [107]). This produces

\[ F_{AC,\text{opt}}^*(e^{j\omega}) = DQ(e^{j\omega})R^{-1}(e^{j\omega}), \]  

(A.15)
provided the inverse of $H^{H}_{AC}(e^{j\omega})S_{xx}^{T}(e^{j\omega})H_{AC}(e^{j\omega})$ exists.

To ensure the derivative’s zero-point is a global minimum, the convexity of $\text{tr}(S_{ee}(e^{j\omega}))$ is examined. Ordering the entries of $F_{AC}(e^{j\omega})$ vectorially through the concatenation of its rows, the Levi matrix (or, complex Hessian) [108] of $\text{tr}(S_{ee}(e^{j\omega}))$ with respect to $F_{AC}(e^{j\omega})$ is calculated. The result is a block diagonal matrix with identical size $C \times C$ blocks equal to $R(e^{j\omega})/D^2$, a scaled version of (A.14). This structure means convexity is proven if and only if $R(e^{j\omega})$ is positive definite. Although a detailed examination of the necessary conditions for a positive definite $R(e^{j\omega})$ is not provided here, there are some easily determined conditions that produce a non-invertible $R(e^{j\omega})$, resulting in a semi-definite matrix for which a unique minimum cannot be found. These conditions are discussed below.

With the solution in place, an expression for the minimized TAMSE is found by applying (A.15) to (A.8), resulting in the optimal error CSD matrix,

$$S_{ee,\text{opt}}(e^{j\omega}) = S_{xx}(e^{j\omega}) - Q^{H}(e^{j\omega})R^{-H}(e^{j\omega})Q(e^{j\omega})$$

(A.16)

$$= S_{xx}(e^{j\omega}) - \frac{Q^{H}(e^{j\omega})F_{AC,\text{opt}}^{H}(e^{j\omega})}{D}.$$  

(A.17)

As through (A.10), the optimal TAMSE $\sigma_{e,\text{opt}}^2$ can be determined from this expression.

### A.3 Extensions and Additional Considerations

#### A.3.1 Optimality of Decimated Output Sequences

As shown above, the time-averaged MSE is minimized through (A.15). However, since this error is WSCS($D$), the expected squared error will generally vary periodically with the time index. This prompts the question: is this expected squared error minimized at all times? By considering the errors of all $D$-fold decimated output sequences it will be shown that yes, the solution (A.15) does not periodically sacrifice instantaneous performance to optimize average performance. The left portion of Fig. A.2 shows the block diagram producing one of
these decimated sequences, defined through

\[ e_k[n] = e[nD + k] = x[nD + k] - y[nD + k], \quad (A.18) \]

for \( k = 0, 1, \cdots, D - 1 \). Each of these sub-sequences is WSS with an MSE determined by

\[ \sigma^2_{e,k} = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} R_{ee}[nD + k, nD + k]. \quad (A.19) \]

Figure A.2: Equivalent reconstruction models used to determine the MSE of a decimated output sequence.

The system on the left of Fig. A.2 can be simplified to the equivalent system on the right, where \( E_{i,k}(e^{j\omega}) \) is the \( k \)th polyphase component of \( F_i(e^{j\omega}) \) based on the type-1 decomposition [20]

\[ F_i(e^{j\omega}) = \sum_{k=0}^{D-1} e^{-j\omega k} E_{i,k}(e^{j\omega D}). \quad (A.20) \]

The mean-squared variance (A.18) can therefore be minimized through selection of an optimal set of \( E_{i,k}(e^{j\omega}) \). As will be proven below, combining the optimal \( E_{i,k}(e^{j\omega}) \) through (A.20) determines the same optimal synthesis bank of (A.15), showing \( F_{AC,\text{opt}}(e^{j\omega}) \) minimizes the MSE for all time indices.

The \( k \)th set of polyphase components is represented through the vector \( E_k(e^{j\omega}) = [E_{0,k}(e^{j\omega}), \cdots, E_{C-1,k}(e^{j\omega})] \). Similarly to (A.15), the mean-squared variance of \( e_k[n] \) is minimized through selection of the optimal filters

\[ E_{k,\text{opt}}(e^{j\omega D}) = h_k^T(e^{j\omega})QH(e^{j\omega})R^{-1}(e^{j\omega}), \quad (A.21) \]

where \( h_k^T(e^{j\omega}) = e^{j\omega k}[1, W^k, \cdots, W^{k(D-1)}] \). As before, (A.21) is only valid for \(|\omega| \leq \pi/D\). The alias components of the polyphase representation (A.20) are
found to be
\[
F_i(e^{j\omega}W^p) = \sum_{k=0}^{D-1} e^{-j\omega k}W^{-pk}E_{i,k}(e^{j\omega D}W^pD) = \sum_{k=0}^{D-1} e^{-j\omega k}W^{-pk}E_{i,k}(e^{j\omega D}). \tag{A.22}
\]
These alias components make up the entries of \( F_{AC}(e^{j\omega}) \) which, applying (A.21) to (A.22), is written as
\[
F_{AC}(e^{j\omega}) = A(e^{j\omega})B(e^{j\omega})Q^H(e^{j\omega})R^{-1}(e^{j\omega}), \tag{A.23}
\]
where
\[
A(e^{j\omega}) = \begin{bmatrix}
1 & e^{-j\omega} & \cdots & e^{-j\omega(D-1)} \\
1 & e^{-j\omega}W^{-1} & \cdots & e^{-j\omega(D-1)}W^{-(D-1)} \\
\vdots & \vdots & \ddots & \vdots \\
1 & e^{-j\omega}W^{-(D-1)} & \cdots & e^{-j\omega(D-1)}W^{-(D-1)^2}
\end{bmatrix} \tag{A.24}
\]
and
\[
B(e^{j\omega}) = \begin{bmatrix}
h^T_0(e^{j\omega}) \\
\vdots \\
h^T_{(D-1)}(e^{j\omega})
\end{bmatrix}. \tag{A.25}
\]
Realizing \( A(e^{j\omega}) = DB^{-1}(e^{j\omega}) \), the result (A.23) simplifies to the original optimal solution (A.15), thereby showing equivalence of the solutions. Additionally, through (A.19) the decimated output MSE is given by
\[
\sigma^2_{e,k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{e,k}(e^{j\omega})d\omega = \frac{D}{2\pi} \int_{-\pi/D}^{\pi/D} S_{e,k}(e^{j\omega D})d\omega, \tag{A.26}
\]
for which the decimated error PSD is shown to be
\[
S_{e,k}(e^{j\omega D}) = \frac{1}{D}h^T_k(e^{j\omega}) (S_{xx}(e^{j\omega}) - Q^H(e^{j\omega})R^{-H}(e^{j\omega})Q(e^{j\omega})) h^*_k(e^{j\omega}). \tag{A.27}
\]
Solving (A.26) for \( 0 \leq k \leq D - 1 \) allows the expected squared error at any given time index to be found.

An important implication of this result is that, since all \( D \)-fold decimated sequences are individually optimal, the output does not need to be implemented at the model rate \( T_M \). In fact, since selection of \( D \) is only bounded by the minimum value necessary for model accuracy, an output signal can be produced at nearly any rate (it must be an integer ratio of \( 1/T_P \)). This enables lower-cost reconstruction implementations without reducing the modelling accuracy of \( x[n] \).
A.3.2 Additive Noise and Interference

The dual analysis bank model of Fig. A.3 is used to consider \( w[n] \), an additive noise or interference signal with known statistics. This input has the same requirements and assumptions as \( x[n] \). The separate set of analysis filters referred to as \( G_0(e^{j\omega}), \ldots, G_{C-1}(e^{j\omega}) \) allows \( w[n] \) to be considered through either the same sampling process as \( x[n] \) (in which case \( G_{AC}(e^{j\omega}) = H_{AC}(e^{j\omega}) \)) or a unique sampling process (e.g., to model multipath distortions of an interfering signal). One useful configuration is an advance chain, where \( G_1(e^{j\omega}) = e^{j\omega} \). This considers the subband cross-correlations using the WSCS(\( D \)) auto-correlation of \( w[n] \) through

\[
E[w[nD + p]w^*[mD + q]] = E[v_p[n]v_q^*[m]],
\]

(A.28)

for all \( n, m \) and \( p, q = 0, 1, \ldots, C - 1 \). For \( C \leq p, q \leq D - 1 \), (A.28) can be set to zero. This configuration models noise that is introduced subsequent to the sampling process.

![Figure A.3: Dual analysis bank model used to consider the presence of an interfering source. The system can be generalized to consist of \( K + 1 \) merging analysis banks to model \( x[n] \) in the presence of \( K \) distinct interfering sources.](image)

To further generalize, \( K \) distinct interfering sources \( (w_1[n], \ldots, w_K[n]) \) can be considered using \( K + 1 \) merging banks of analysis filters. This framework is useful for considering generalized MIMO problems, since a separate synthesis bank can be determined for each output. For this general problem, a separate alias component matrix is constructed for each analysis bank. The matrix corresponding to the \( k \)th
interfering source $w_k[n]$ is defined as through (A.6), and labelled $G_{k,AC}(e^{j\omega})$.

Using this generalization, the optimal reconstruction filter bank and error CSD matrix are still respectively determined through (A.15) and (A.17). However, the component matrices $Q(e^{j\omega})$ and $R(e^{j\omega})$ must be modified to account for the multiple source structure. The new definitions for multiple sources are

\[
Q(e^{j\omega}) = H_{AC}^T(e^{j\omega})S_{xx}(e^{j\omega}) + \sum_{k=1}^{K} G_{k,AC}^T(e^{j\omega})S_{xw_k}(e^{j\omega}), \tag{A.29}
\]

\[
R(e^{j\omega}) = H_{AC}^T(e^{j\omega})S_{xx}^H(e^{j\omega})H_{AC}^*(e^{j\omega}) + \sum_{l=1}^{K} H_{AC}^T(e^{j\omega})S_{xw_k}^H(e^{j\omega})G_{l,AC}^*(e^{j\omega})
+ \sum_{k=1}^{K} \sum_{l=1}^{K} G_{k,AC}^T(e^{j\omega})S_{w_kw_l}^H(e^{j\omega})G_{l,AC}^*(e^{j\omega}). \tag{A.30}
\]

It is worth noting that for a given set of analysis banks modelling a MIMO scenario, $R(e^{j\omega})$ does not change with respect to selection of the desired signal. This is useful for reducing the design complexity since each distinct synthesis bank is determined by applying a unique $Q(e^{j\omega})$ to a common $R^{-1}(e^{j\omega})$.

### A.3.3 Numerical Considerations

While an analytical solution has been considered so far, a feasible solution requires a numerical approach. For this, the solution (A.15) is determined for varying frequencies over $|\omega| \leq \pi/D$, providing sampled versions of the synthesis filters’ frequency responses through piecewise combinations of the alias components. The sampled frequency responses should be of sufficiently fine resolution from which accurate filter realizations can be determined. A high resolution model will generally also allow an accurate calculation of $\sigma_{e, opt}^2$ through numerical integration of (A.10).

For the solution (A.15) to exist at a particular $\omega$, the matrix $R(e^{j\omega})$ of (A.14) must be invertible. As mentioned above, $S_{xx}(e^{j\omega})$ is diagonal for a WSS
x[n], with real non-negative terms representing portions of the power spectrum. This means \( S_{xx}^H(e^{j\omega}) = S_{xx}(e^{j\omega}) \), allowing the \((p, q)\)th entry of \( R(e^{j\omega}) \) to be found through

\[
\left[ R(e^{j\omega}) \right]_{(p,q)} = \sum_{k=0}^{D-1} S_{xx}(e^{j\omega W_k}) H_p(e^{j\omega W_k}) H_q^*(e^{j\omega W_k}), \tag{A.31}
\]

where \((p, q) = 0, \ldots, C - 1\). To be invertible, \( R(e^{j\omega}) \) must be of full rank \( C \), requiring diagonal \( S_{xx}(e^{j\omega}) \) to have at least \( C \) nonzero entries. This is not true at frequencies for which there are fewer non-zero aliased components than samplers. Although \( R(e^{j\omega}) \) is not invertible in these cases, the reduced number of aliased components actually indicates over-sampled frequencies which can be made free of reconstruction error. To circumvent this problem, the zero-valued portions of \( S_{xx}(e^{j\omega}) \) can be replaced with a small positive value \( \epsilon \) (much smaller than the non-zero valued spectral components), thereby providing an invertible approximation of \( R(e^{j\omega}) \). The unaltered \( S_{xx}(e^{j\omega}) \) is used for \( Q(e^{j\omega}) \) (A.13), the non-inverted portion of the solution, ensuring zero-valued portions of the input spectrum remain so at the output. Invertibility of (A.31) also places some restrictions on \( H_{AC}(e^{j\omega}) \), generally requiring unique timing offsets or distortions for the samplers. In the event of duplicate sampling, the redundant branches of the model should be removed, corresponding to the removal of linearly dependent columns from \( H_{AC}(e^{j\omega}) \).

An alternate numerical solution was presented for this problem in the earlier work of Ohno and Sakai [101]. Their approach examined the error caused by missing subband of a biorthogonal filter bank in the correlation domain. Numerical optimization techniques were then used to find synthesis filters of predetermined length. Although their solution could not guarantee the globally optimal solution, it did have the advantage of allowing specific design constraints (e.g., linear phase synthesis filters).

### A.4 Acknowledgement

The text of the appendix, in full, is adapted from material that has been published as: R. S. Prendergast and T. Q. Nguyen, “Minimum mean-squared error...
reconstruction for generalized undersampling of cyclostationary processes,” IEEE Trans. Sig. Proc., 2006. The dissertation author was the primary researcher for this publication, and the co-author directed and supervised the research which forms the basis for this chapter.
Bibliography


