Bayesian super-resolution with application to radar target recognition

Richard Oliver Lane

A thesis submitted to
The University of London
for the degree of
Doctor of Engineering

Communications Engineering Doctorate Centre
Department of Electronic and Electrical Engineering
University College London

February 2008

Abstract

This thesis is concerned with methods to facilitate automatic target recognition using images generated from a group of associated radar systems. Target recognition algorithms require access to a database of previously recorded or synthesized radar images for the targets of interest, or a database of features based on those images. However, the resolution of a new image acquired under non-ideal conditions may not be as good as that of the images used to generate the database. Therefore it is proposed to use super-resolution techniques to match the resolution of new images with the resolution of database images.

A comprehensive review of the literature is given for super-resolution when used either on its own, or in conjunction with target recognition. A new super-resolution algorithm is developed that is based on numerical Markov chain Monte Carlo Bayesian statistics. This algorithm allows uncertainty in the super-resolved image to be taken into account in the target recognition process. It is shown that the Bayesian approach improves the probability of correct target classification over standard super-resolution techniques.

The new super-resolution algorithm is demonstrated using a simple synthetically generated data set and is compared to other similar algorithms. A variety of effects that degrade super-resolution performance, such as defocus, are analyzed and techniques to compensate for these are presented. Performance of the super-resolution algorithm is then tested as part of a Bayesian target recognition framework using measured radar data.

Key Phrases: — Automatic target recognition (ATR), Bayesian methods, classification, deconvolution, image restoration, Markov chain Monte Carlo (MCMC), point spread function (PSF), radar, statistics, superresolution.

Declaration

I, Richard Oliver Lane, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

Date:

Novel aspects of the work

The major original contributions from this thesis are:

- A Markov chain Monte Carlo (MCMC) Bayesian super-resolution algorithm that estimates the full probability distribution of the complex scattered field of a target has been developed and tested in Section 3.2. A similar algorithm for estimating the underlying radar cross-section of a target has been developed and tested in Section 3.3. Although a full-distribution Bayesian algorithm that effectively uses a parametric scattering centre model has previously been proposed in [2], the present author is not aware of any such algorithm applicable to a non-parametric radar model having been published. In [147] it is stated that to their knowledge, "no works have been done on applying MCMC on the super-resolution problem". However, the work from which this thesis is derived does, in fact, pre-date that work, as evidenced by the conference paper [80].

- The minimum mean-square error (MMSE) super-resolution algorithm has been used to estimate the full Bayesian distribution of the super-resolved complex scattered field. It is well known that the MMSE solution to the

super-resolution problem is equal to the *mean* of the Bayesian solution under assumptions of Gaussian statistics. However, an approximation to the *complete* Bayesian solution has been developed in Section 3.5 of this thesis, whereby an estimate of the posterior complex-field covariance matrix is derived from the MMSE solution. This method uses knowledge of the noise covariance matrix but does not require the specification of a prior covariance for the complex field, as is usual.

- A joint autofocus and super-resolution algorithm using the full probability distribution under Bayesian statistics has been developed and tested in Chapter 4.

- A Bayesian super-resolution target-recognition framework based on sound theoretical principles has been developed and tested in Chapter 5. Several other authors have proposed the combination of super-resolution and target recognition to improve system performance. However, the choice of algorithms has thus far been rather *ad hoc*. The development of a full-distribution Bayesian super-resolution algorithm in this thesis allows information about uncertainty introduced during the super-resolution process to be taken into account in the target recognition process.

- An approximate analytic description of the radar point spread function has been derived in Appendix C for situations where the signal is defocused by cross-track acceleration.

Publications

The following conference papers were derived from the work presented in this thesis:

- R. O. Lane, "Estimating radar cross section using Bayesian image restoration", *Proceedings of the London Communications Symposium*, pages 1–4, September 2003. [77]

- R. O. Lane, K. D. Copsey, A. R. Webb, "A Bayesian approach to simultaneous autofocus and super-resolution", *Proceedings of SPIE*, vol. 5427, pages 133–142, April 2004. [80]

- K. D. Copsey, R. O. Lane, A. R. Webb, "Designing NCTR algorithms when operating sensor conditions differ from training conditions", *International conference on radar systems (Radar 2004)*, Toulouse, France, October 2004. [31]

- K. D. Copsey, R. O. Lane, S. Manchanda, A. R. Webb, "Bayesian approach to exploiting prior targeting information within a weapon seeker", *NATO RTO SET Symposium SET-080, Target Identification and Recognition using RF Systems*, Oslo, Norway, October 2004. [29]

- K. D. Copsey, R. O. Lane, S. Manchanda, A. R. Webb, "Bayesian approach to recognising relocatable targets", *NATO RTO SET Symposium SET-080, Target Identification and Recognition using RF Systems*, Oslo, Norway, October 2004. [30]

- R. O. Lane, K. D. Copsey, A. R. Webb, "Assessment of a Bayesian approach to recognising relocatable targets", *NATO RTO SET-096 specialists' meeting on the millimeterwave advanced target recognition and identification experiment (MATRIX 2005)*, Oberammergau, Germany, May 2005. [81]

- R. O. Lane, "Super-resolution and the radar point spread function", *Proceedings of the London Communications Symposium*, pages 5–8, September 2005. [78]

- R. O. Lane, "The effects of Doppler and pulse eclipsing on sidelobe reduction techniques", *IEEE National Radar Conference*, Verona, NY, April 2006. [79]

To my parents

Acknowledgements

The accomplishment of this thesis was made possible with the help and support of many people. First of all, I would like to thank my academic supervisor, Prof. Chris Baker of University College London, for making useful suggestions in which to direct my studies, providing an independent opinion of my work, and making helpful comments on early drafts of this thesis. I would also like to thank my industrial supervisor, Dr. Andrew Webb of QinetiQ Malvern, for suggesting many of the research avenues pursued here and for continuously monitoring progress of my research.

The work presented here was carried out whilst I was employed at QinetiQ Malvern, where I have had many useful discussions with several colleagues. Dr. Sunil Manchanda helped me think laterally by discussing ideas outside the mainstream and also made me think how work at QinetiQ could be used towards more academic study. Dr. Keith Copsey provided a lot of support in project work and helped code the Monte Carlo algorithm presented later in this thesis. Dr. Adrian Britton injected more realism in discussions surrounding the work. Many other colleagues also provided support, guidance and assistance in a variety of ways.

I am indebted to my parents and grandparents for their love, and support: both emotional and financial. I am also grateful to my partner, Eleanor, and son, Thomas, for providing enjoyable distractions from work and being understanding when I had to spend time away from them.

This work was funded in part by the EPSRC Engineering Doctorate (EngD) scheme and the UK Ministry of Defence Corporate Research Programme. The thesis was typeset using LaTeX2e and the WinEdt text editor.

Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Appendix B

Appendix C

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Appendix C

List of Abbreviations

AIC	Akaike Information Criterion
APES	Amplitude and Phase Estimation of a Sinusoid
AR	Auto Regressive
ATR	Automatic Target Recognition
CDF	Cumulative Distribution Function
CFAR	Constant False Alarm Rate
CID	Constrained Iterative Deconvolution
EPSRC	Engineering and Physical Sciences Research Council
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Techniques
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FM	Frequency Modulated
GTD	Geometric Theory of Diffraction
HDVI	High-Definition Vector Imaging
IFFT	Inverse Fast Fourier Transform
IMP	Incremental Multi-Parameter
IMU	Inertial Measurement Unit
ISAR	Inverse Synthetic-Aperture Radar
LS	Least Squares
MAP	Maximum *a posteriori*
MCMC	Markov Chain Monte Carlo
ME	Maximum Entropy
M-H	Metropolis-Hastings
ML	Maximum Likelihood
MMSE	Minimum Mean-Square Error
MRF	Markov Random Field

MSTAR	Moving and Stationary Target Acquisition and Recognition
MUSIC	Multiple Signal Classification
MVDR	Minimum-Variance Distortionless Response
NCTR	Non-Cooperative Target Recognition
NN	Nearest Neighbour
PDF	Probability Density Function
PFA	Probability of False Alarm
PGA	Phase Gradient Autofocus
PSF	Point Spread Function
PWF	Polarimetric Whitening Filter
RCS	Radar Cross Section
RF	Radio Frequency
RJMCMC	Reversible Jump Markov Chain Monte Carlo
RVDR	Reduced-Variance Distortionless Response
SAR	Synthetic-Aperture Radar
SI	International System of Units
SNR	Signal-to-Noise Ratio
SVA	Spatially Variant Apodization
SVD	Singular Value Decomposition
TAM	Toeplitz Approximation Method
TLS	Total Least Squares

Chapter 1

Introduction

1.1 Historical Context

Radar is a sensor technology that has been in existence in one form or another for over one hundred years. In 1904 Christian Hülsmeyer filed a patent that described how the detection of reflected radio waves could be used as an early warning detection system for ships [63]. Hülsmeyer built a working prototype but there was little interest in his system as it had a short range of only one mile [136]. Systems with improved ranges were made by workers at the US Naval Research Laboratory: first in the 1920s to measure the height of the ionosphere with a pulsed radar, and then in the early 1930s to detect the presence of aircraft using continuous-wave radar. From 1935 there was rapid development in the run-up to the second world war, with Watson-Watt's team in the UK demonstrating a pulsed technique to measure the range of aircraft. By the end of the war, radar systems with the ability to show range and angle information on the same display had been developed by several countries [136]. In the 1950s synthetic-aperture radar (SAR) was invented: a system for obtaining high resolution in both the range and cross-range directions [44]. The first spaceborne SAR, SeaSat, was launched in 1978 and provided images of the Earth's surface. Although many radar systems in use today still have resolutions in the order of tens of metres, an increasing number of systems have resolutions of one metre or down to even ten centimetres [44].

With high-resolution radar it becomes possible not only to detect a target's location but also to recognize the general class of target or type within a class.

The recognition process can be carried out by a human operator who looks at a signal or image and uses a combination of his skill, knowledge and reference manuals to perform the task. However, the operator is prone to fatigue if working over extended periods of time and has a limit to the number of signals or images that can be recognized per hour. If computers could reliably perform this task automatically then potentially thousands of images per hour could be processed. The subject of automatic target recognition (ATR) has been developing significantly since the 1960s [27]. Research programmes have been accelerated with the advent of high-speed low-cost computers and an increasing military need for reliable ATR. ATR systems generally rely on techniques such as statistical pattern recognition [155] or artificial neural networks [6]. Tait [142] provides a thorough review of the radar ATR process and divides it into a number of steps: radar measurements, generating a target database, target signature models, recognition algorithms and data processing functions.

When ATR systems employ multiple sensors, consideration must be given to how data from different sources is exploited. Either several radar sensors or a combination of radar and other types of sensor, such as infra-red, could be in use. The data from each sensor might provide information such as target position, target features, or an estimate of the target type. If possible, it would be advantageous to directly compare images from the various sensors as images provide more information than higher level features alone. However, because the images from different sensors have different attributes this could prove to be difficult. If the sensors are limited to different types of radar then a direct comparison is more likely to be possible. One of the main ways in which the images from different radars differ is that of resolution – the minimum separation required for closely spaced scattering centres on a target to be distinguished. This is a fundamental limit based on the bandwidth of the transmitted waveform and the physical size of the radar system compared to the transmitted radio wavelength. One way of making the images comparable is to use super-resolution for the sensors with a poor resolution. Super-resolution is the use of signal processing techniques to increase the resolution beyond physical limits by using knowledge of a system's point-target response, and by making assumptions about the scene of interest. Once images from all the sensors have been processed to the same resolution using super-resolution, standard ATR techniques

can be applied and the data from the various sensors compared directly.

A good way of combining data from different sources is through the use of Bayesian statistics [82]. The main motivation behind a Bayesian approach lies in the unique ability of Bayesian statistics to handle limited and possibly conflicting pieces of information in a fully consistent manner. In particular, Bayesian theory provides a consistent mechanism for manipulating probabilities assigned to data. Further advantages in the use of Bayesian techniques include the ability to cope with additional prior information and the production of confidence intervals and other statistics of interest for the parameters estimated. Jaynes gives arguments in favour of Bayesian methods in [68] and a history of the subject in [70].

1.2 Literature Survey

1.2.1 Introduction to Super-resolution

Super-resolution may be defined as the use of signal processing techniques to improve a system's resolution beyond the classical Rayleigh resolution limit. An image with Rayleigh resolution is the result of a convolution between a point spread function and a high-resolution representation of the scene of interest. Deconvolution removes the effect of the point spread function and reveals the high-resolution scene. Therefore all deconvolution algorithms are super-resolution algorithms. In radar or communications systems using an array of antennas, weights can be applied to signals received at each antenna to precisely steer nulls in the direction of a nearby strong signal. This allows the direction and power of a second signal to be determined without power leaking from the nulled signal. Since the two targets can be closer together than the Rayleigh limit, this form of processing – sometimes referred to as direction finding – can also be considered to be super-resolution. Radar data is often obtained in the frequency domain and an inverse Fourier transform is applied to obtain data in the spatial domain, which is more useful than frequency data for detection, recognition and interpretation purposes. The Fourier transform of an ideal point target is a complex sinusoid with a real-part frequency that depends on the position of the target. Therefore any technique that is able to estimate the parameters of multiple sinusoids closely spaced in frequency

is equivalent to a super-resolution technique. Accordingly this literature survey covers the areas of super-resolution, deconvolution, direction finding, and estimation of sinusoids, considering them to be equivalent problems. In the following discussion of techniques, the terms point source, scattering centre, target, and sinusoid will be used interchangeably depending on the terminology used originally to describe an algorithm and the context in which it is being discussed. Whole target objects, such as a military vehicles discussed in later chapters, are considered to be composed of a number of these scattering centres.

There are two major classes of super-resolution model. The first class describes the scene as a finite number of point scatterers that can take any position in the scene, and the received continuous radar signal is sampled for digital processing. This is sometimes known as a high-level model [140], the scattering centre model [90], or the parametric model [96]. The advantage of this model is that major isolated scatterers in the scene, which can often account for the majority of received energy at the radar, are well modelled because of their positional accuracy. The disadvantage of this model is that the number of scatterers in the scene must be estimated as well as their position and amplitude. Without a regularization procedure, algorithms based on this model create a large number of scatterers with small amplitude, which are generally related to noise rather than true structure in the scene. Another disadvantage of this model is that a linear superposition of perfect point scatters may not completely describe a complex target.

The second class of model assumes the scene is a continuously varying high-resolution function and during digital processing we consider samples of this function on a regular grid. It is sometimes known as a low-level model [140], the continuum scattering model [90], or the non-parametric model [96]. This model has the advantage that extended targets are better modelled and the number of "scatterers" is fixed according to the sample spacing and size of the scene, which means this parameter does not have to be estimated. The disadvantage is that isolated strong scatterers are less well modelled if they are not positioned at a scene sample point. In a similar manner to the first class of model, when noise is present, algorithms will tend to give non-zero amplitude to areas of the scene with no valid target, unless a regularization procedure is used.

A review of the super-resolution literature is now given. A large number of super-resolution algorithms have been proposed so it is not possible to mention every single one. However, this review discusses most of the more commonly used super-resolution algorithms and at the end of Section 1.2.3 references to other super-resolution reviews are made available.

1.2.2 Point Source Super-resolution

Perhaps the most studied point-source super-resolution algorithm, for data measured from an array of receivers, is the MUSIC algorithm, introduced by Schmidt [129]. The MUSIC algorithm models the observed data as the combination of a signal subspace and an orthogonal noise subspace. The number of point sources in a scene is determined by analyzing the eigenvalue spectrum of the measured covariance matrix. The direction of the sources is determined by the position of the strongest peaks in a power function. Parameters of the point sources can then be calculated.

The ESPRIT algorithm has lower computation and storage requirements than MUSIC and results in a solution that is more robust to errors in the positioning of array elements [124]. This is achieved by putting a translational geometric constraint on the array element positions. However, the constraint is a mild one, which admits the commonly used uniform linear array, for example, and it has been shown that the least squares (LS) version of ESPRIT is statistically equivalent to the Toeplitz approximation method (TAM) for uniform linear arrays [119]. Asymptotically for large signal-to-noise ratios (SNRs) total least squares (TLS) ESPRIT has the same mean-square error as LS ESPRIT [119]. Other advantages of ESPRIT are that knowledge of the array manifold and source correlation is not required, and the bias of target position estimates generated by MUSIC in low SNR conditions along with the omission of some targets is not present when ESPRIT processing is used [124].

An image processing algorithm that has gained widespread popularity in the astronomical community [32, 59] but has also been used in the processing of microwave measurements appropriate to radar is CLEAN [149]. In this algorithm the peak of the image is selected and this is assumed to relate to the strongest point source in the scene. The system response to a source of this strength at that position is subtracted from the image to leave a residual. The

process is then repeated using the the residual from the previous iteration as the input image to the peak-picking step until either the residual is reduced to an acceptable level relating the assumed noise power or the desired number of sources have been detected. In the original algorithm, point sources within a beam-width become combined into a single source when convolved with a "CLEAN beam", which is usually the central portion of the point spread function without sidelobes. This last step means the algorithm can no longer be said to have super-resolution properties as it is impossible to resolve two or more targets in the main beam. However, by omitting the last step or using a beam narrower than the original main beam, CLEAN may be considered a crude super-resolution algorithm [132].

One disadvantage of CLEAN is that once the position and strength of the strongest target has been estimated this is fixed for the rest of the algorithm. In practice, knowledge that other targets are present in the scene means that the estimate should be modified. Without modification all the CLEAN algorithm can do is create spurious low-power targets near the main target to compensate for errors. This was probably the motivation behind the CLEAN beam as described above. The RELAX algorithm [83] was proposed specifically as an improvement to CLEAN, whereby at each iteration when a new target is discovered the parameters of all the previously discovered targets are re-estimated in an iterative loop until the change in a cost function is below a specified threshold. Thus target parameters are more accurately estimated. In addition, the algorithm makes use of a generalized Akaike information criterion (AIC) [1] to determine the number of targets automatically, rather than using an arbitrary fixed number. An improvement that reduces computation time of the the algorithm avoids the zero-padding of the fast Fourier transform (FFT) by using a zoom-FFT approach [86].

The incremental multi-parameter (IMP) algorithm [94] is another algorithm that seeks to estimate target parameters one at a time, while refining previous estimates as new targets are discovered. This algorithm was developed more than five years before RELAX but to date there appears to have been no attempt in the literature to link the two algorithms, with IMP attracting much less attention than RELAX. In IMP, rather than subtracting the effect of a dominant scatterer from the measured data, at each iteration the data is projected onto a

subspace orthogonal to that spanned by the calibration vectors corresponding to current estimates of the dominant scatterer positions. In a similar manner to RELAX, previous estimates of scatterer parameters are refined at each major iteration until the parameters change by less than a specified amount before a new scatterer is scanned for. The algorithm is terminated when the residual power is equal to that expected for noise and clutter, or noise alone. The subspace formulation of IMP is reminiscent of MUSIC but IMP is able to operate at SNRs more than 10 dB lower than those required by MUSIC and correlated signals are automatically accommodated without the need for pre-whitening [94].

An efficient algorithm that uses several matrix algebra techniques is the matrix pencil method [127]. This method has the advantage that, in addition to ideal point sources, damped and undamped frequency-dependent behaviour is modelled, which allows a better parameter estimation for point sources that follow this model. In the matrix pencil method, a Hankel data matrix containing shifted versions of the noisy measured data is formed. The Hankel matrix is decomposed using singular value decomposition (SVD) and the number of scatterers in the scene is determined by analyzing the singular values. Two filtered matrices are formed using the singular vectors for overlapping partitions of the data, and are then combined. An eigenvalue decomposition of the combined matrix is performed with the eigenvalues giving the position and frequency-dependent properties of scatterers. The complex amplitudes of the scatterers can then be determined using a standard least-squares pseudoinverse approach. Certain versions of the ESPRIT algorithm are classed as matrix pencil methods [61].

According to the theory of analytic continuation, if the spectrum of an image is known within a limited bandwidth then, for an object of finite size, the image spectrum may be found throughout the whole frequency domain [57]. In practice, finite sampling and noise prevent an exact application of the theory. However, a technique known as bandwidth extrapolation is based on the analytic continuation principle. Bandwidth extrapolation models the data in the frequency domain as an auto-regressive (AR) process [71], also known as a linear prediction model [99] or an all-pole model [150]. Several techniques, such as those due to Pisarenko and Prony are available to estimate the AR

coefficients but Burg's technique [15] appears to offer advantages over the others in terms of stability using measured data [99]. The position of scatterers can be determined from solutions to a characteristic equation involving the AR coefficients. Alternatively, the bandwidth can reliably be extrapolated by up to a factor of four using the coefficients, and then using a standard windowed FFT to estimate the scene. It has been shown that, for Gaussian random processes with known autocorrelation lags, the spectral estimate using an AR model is identical in analytical form with that produced using the maximum entropy (ME) principle [15]. An alternative ME method is covered in more detail in the distributed-source super-resolution literature survey.

A recent important advance in high-resolution target modeling has been the use of attributed scattering centres [115]. This theory uses a parametric scattering model for point targets based on the geometric theory of diffraction (GTD). The advantage of this model over others is that the frequency and angle dependence of scatterers is taken into account, which allows the accurate modeling of a wider variety of scatterers than the ideal point scatterer. The additional types of scatterer specifically modelled are flat plates, dihedrals, single- and double-curved surfaces, straight edges, curved-edge diffraction and corner diffraction. The fact that scatterers are more accurately modelled means that a higher accuracy in resolution is achievable for scatterer locations. Although the parameters may be estimated using a maximum likelihood method, this is a computationally intensive task. It has been shown that parameters of reduced complexity models, such as the damped or undamped exponentials previously mentioned, can be related directly to GTD parameters. Use of the simpler exponential models may be appropriate in many circumstances as the bias introduced by them is often small compared to GTD-based parameter variances for realistic radar specifications [115]. The number of scatterers under any of these models may be determined using an order-selecting maximum likelihood technique.

The number of scatterers in a scene may also be determined using Bayesian analysis to compare the probability of various hypotheses that certain numbers of scatterers exist [64]. However, the examples given in [64] have quite restrictive assumptions on prior knowledge, such as requiring the scatterer amplitudes to be known. Using less restrictive assumptions, it is possible to es-

timate both the amplitude and position of a known number of sources using a Markov chain Monte Carlo (MCMC) method based on Gibbs sampling [19] and the Metropolis-Hastings algorithm [24]. This approach has been used for cases when the noise variance is both known [4] and unknown [46]. The basic method has also been extended to exhaustively test models with different numbers and types of source such as chirp, decay, and periodic signals in correlated noise [47]. However, if there are a large number of possible models this procedure can become computationally intensive. A more efficient way to estimate the model order is the reversible jump Markov chain Monte Carlo (RJMCMC) method [39, 52]. RJMCMC has been used to estimate the number and parameters of: sinusoids in additive noise [2], point sources in both multiplicative Gaussian noise [54] and Poisson noise [140], and general frequency modulated signals [28]. It has also been applied to the related problem of polynomial-phase signal parameter estimation [144–146]. However, while the polynomial-phase model could be extended to include multiple targets, the particular case studied appears to be limited to a single manoeuvering target and the power of the reversible jump technique is used to estimate the polynomial order rather than the number of targets.

Further analysis of point source models is not included here as the unknown number of scatterers results in a variable number of dimensions in the Bayesian framework and algorithm introduced in later chapters. Explanation of variable-dimension parameter space would require a more detailed RJMCMC Bayesian analysis, which is beyond the scope of this thesis. The interested reader is referred to [39, 52].

1.2.3 Distributed Source Super-resolution

Classical analysis of spectral information is through the use of the Fourier transform. A non-coherent super-resolution technique based on Fourier-domain transformations is constrained iterative deconvolution (CID) [122]. The algorithm proceeds by taking the inverse of the point spread function in the Fourier domain and decomposing it into the sum of a geometric series. At each iteration an inverse fast Fourier transform (IFFT) is applied to the current estimate of the scene in the frequency domain. A positivity constraint is then applied to the image and it is converted back to the frequency domain using an FFT. The next

term in the geometric series is then added before the next iteration. While the algorithm does achieve super-resolution, it is applicable to non-coherent images only, which limits its application.

Another super-resolution technique based on alternating FFTs, but with application to complex data, is super-SVA [139]. Spatially-variant apodization (SVA) is a powerful technique used to eliminate finite-aperture induced sidelobes [138]. The elimination of sidelobes increases the bandwidth of the signal and the super-SVA algorithm uses this property to improve the system resolution. In the first step of the algorithm, an FFT is applied to the data, SVA is used to remove the sidelobes and the data is transformed back to its original form using an IFFT. An inverse weighting corresponding to the IFFT of the SVA-preserved main lobe is applied to the data, which is then truncated such that the bandwidth has been extended by 50%. The central portion of the spectrum is replaced by the original data and the whole process is repeated until the bandwidth is extrapolated to the desired amount – usually a factor of two. Both CID and super-SVA are able to increase the bandwidth of the original signal because they apply a non-linear function (positivity for CID and SVA for super-SVA), which results in spectral growth. However, the theoretical performance of these algorithms is not well understood as neither is defined in terms of an optimality criterion.

Capon's maximum-likelihood (ML) method was an early attempt at high-resolution spectrum estimation based on an optimality criterion [17]. It is also known as the minimum-variance distortionless response (MVDR) and the reduced-variance distortionless response (RVDR). With this technique the resolution is improved by minimizing the energy contributed by interferers while keeping unit gain on the location of interest. Several variations on the basic technique are possible that mitigate the reduced rank of the covariance matrix used in the method. They are usually based on quadratic constraints or subspace constraints [5]. Other methods for estimating the covariance matrix are given in [16].

The amplitude and phase estimation of a sinusoid (APES) algorithm, like Capon's method, is an adaptive finite impulse response (FIR) filtering algorithm [84]. However, unlike Capon, APES is a matched filter because it estimates both noise and interference in the covariance matrix. Although it is an approximate

ML algorithm, APES tends to produce more accurate spectral estimates than Capon because it reduces the amount of noise that leaks through the filter.

One of the most basic approaches to deconvolution that has a theoretical justification is the least squares (LS) method. This minimizes the squared difference between the measured image data and the image generated from an estimate of the scene. When using a matrix formulation of the problem the solution is given by the Moore-Penrose pseudoinverse, which has been used in [126]. However, the least squares approach is unstable with respect to perturbations in the data: the method effectively over-fits noise, and a small change in the measured data results in a very large change in the solution.

The most popular approach to stabilize the problem is Tikhonov regularization. In comparison to LS, this uses an additional quadratic term that penalizes solutions with large amplitudes and mitigates the effect of noise amplification. The approach is sometimes also known as diagonal loading [5]. Alternatively, the penalty function could penalize roughness to achieve a smooth solution. Tikhonov regularization leads to a solution that is linear in the measured data, which has advantages for ease of computation. Comparisons of Tikhonov regularization with other techniques are given in [25, 26, 53].

A more general approach than Tikhonov regularization uses a generalized penalization term that results in non-linear solutions. Special cases of this approach are the LS and Tikhonov regularization solutions mentioned above and also the maximum entropy solution mentioned below. Other penalty functions, such as those used in feature-enhanced imaging [20, 21], have been devised to emphasize specific attributes of a SAR image.

A similar approach to least squares uses the minimum mean-square error (MMSE) criterion. The solution is equal to the mean super-resolved scene under a Bayesian formulation and Gaussian statistics. The MMSE approach is popular (see [7, 10, 11, 22, 41, 53, 87, 90, 121, 133] for example), but there are a variety of implementations that often use an arbitrary threshold and generally use an iterative process to converge to a solution because the prior covariance matrix for the scene is often estimated as part of the procedure. Having said that, the fundamental approach is theoretically well-founded and performs well with simulated noisy data. The basic form of MMSE super-resolution where the scene and noise prior covariance matrices are known in advance is known

as Wiener deconvolution.

In singular value decomposition (SVD) the point spread function (PSF) matrix is decomposed into three matrices: two orthogonal matrices and a diagonal matrix with singular values on the diagonal [116]. Any singular value below a certain threshold is set to zero and the effective inverse of the PSF matrix is calculated using special properties of the SVD matrices while setting the inverse of zeroed singular values to zero. If all the singular values retain their original values then this is equivalent to the Moore-Penrose pseudoinverse. However, removal of small singular values, which are responsible for the amplification of noise, results in a better performance than the pseudoinverse [7]. The SVD method is equivalent to the MMSE method for high SNRs [87, 90, 121].

High-definition vector imaging (HDVI) introduces a new concept for SAR imaging [5]. HDVI uses standard techniques, such as Capon's maximum likelihood method or MUSIC, to generate a set of super-resolution images, with each image matched to a different model of target properties, such as an ideal point source or a broadside flash associated with target-ground interaction. In the final "image" each pixel is actually a vector of values describing the degree of agreement between the data and each model. The extra information provided is useful for image analysis and can help target recognition systems.

Prior knowledge about the problem or scene is best incorporated via the use of the maximum entropy principle, which allows the maximally non-committal inclusion of prior knowledge [69]. This means no additional spurious information is introduced to the model, such as leading zeros used to pad a Fourier transform. The only information used is that for which there is material evidence. There are different ways of using the maximum entropy principle, which are not necessarily equivalent [102]. "Classical" or "historic" maximum entropy [55, 56] considers the unknown scene as a probability distribution and selects the scene with the maximum Shannon entropy subject to constraints using Lagrange multiplier methods. This method is popular – it was first used for optical image reconstruction problems [49] and has later been extended for use in astronomy [14, 135], ultrasound [137], mass spectrometry [98], multispectral imagery [95] and radar [37]. "Maximum entropy in mean" is a second form of the ME principle. This considers the scene as mean values of a probability density function (PDF) and the data as linear constraints. The algorithm proceeds

by selecting the PDF with the maximum entropy subject to the constraints and finding the mean of the PDF [102]. This form of maximum entropy has been applied to ISAR data in [12], for example. A disadvantage of the maximum entropy method in these forms is that it fails to take measurement noise into account. According to Jaynes [69] only a Bayesian solution is adequate to deal with this problem.

Distributed-source super-resolution has been considered from a Bayesian viewpoint by several authors. One approach by Luttrell [88] is to find the maximum of the target cross-section posterior probability density function using an expectation maximization algorithm. In the particular problem studied there, the possibility that the image under consideration is defocused is also taken into account. This provides some level of robustness with respect to the fact that the point spread function is known only to a certain accuracy. A parametric defocusing model is used and the focus parameter is estimated simultaneously with the target cross-section under a Bayesian framework. In [102] maximum entropy is used to choose the prior probability distributions in a Bayesian model and the conjugate gradient method is used to maximize the posterior probability. In [66] it is stated that standard image restoration using a Tikhonov prior produces overly-smooth solutions, therefore an edge-preserving function and hyper-parameters are introduced in the form of prior information. Under this formulation, a maximum likelihood algorithm based on Gibbs sampling simultaneously estimates the hyper-parameters and the restored image. With the use of auxiliary variables, pixels are simultaneously processed in frequency space, which allows long-distance interactions and makes the algorithm faster than simple Gibbs sampling. This approach is found to be better that Wiener filtering. In [58] a numerical MCMC image restoration algorithm is applied to emission computed tomography data. The model follows poisson statistics and uses a neighbourhood function, which takes into account correlations between adjoining pixels. As with [66], the algorithm simultaneously estimates the hyper-parameters and the restored image. The advantage of this approach over other Bayesian methods is that it includes uncertainty in estimates of the hyper-parameter values. In [143] the Metropolis-Hastings algorithm is applied to a non-linear problem in remote sensing of the atmosphere. An adaptive proposal distribution is used to speed up convergence to the posterior distribu-

tion and the results are compared to a linearized least-squares solution. The Bayesian approach results in a lower error than that of the least-squares solution. In [161] variational methods are used to approximate the true Bayesian posterior distribution in medical and industrial deconvolution problems. This procedure converts an intractable analytic solution to a tractable one. MCMC methods are also studied in [161], but these are used for separating a mixture of Gaussian components rather than determining the pixel values themselves.

Recently, there has been a surge of interest in multi-frame image super-resolution in which a group of low-resolution images of a scene are combined to produce a high-resolution image of that scene [109]. Many multi-frame super-resolution algorithms are based on Bayesian statistics. In [147], for example, a Gaussian Markov random field is used as a prior density for the high-resolution image to incorporate knowledge that adjacent pixels in the image are correlated. The maximum *a posteriori* (MAP) solution is then found using an MCMC approach based on the Gibbs sampler. The approach is extended in [148], where outlier-sensitive bilateral filtering is used as a post-processing method to suppress image artefacts introduced during super-resolution. In [114] the low-resolution image registration parameters are considered as nuisance parameters and are marginalized to leave a function that can be optimized with respect to the high-resolution image. This procedure outperforms methods where the registration parameters are estimated via a MAP method are are fixed before the high-resolution scene is estimated. Although multi-frame image super-resolution algorithms are powerful, the radar target recognition application considered in this thesis has available only a single image, with a high dynamic range, with which to perform super-resolution. Therefore the advantages of these algorithms, such as the use of multiple frames and a limited dynamic range, are not applicable to this work.

The above super-resolution algorithms include most of the popular approaches in the literature. However, there are many other algorithms, most of which are based on similar principles to at least one of these. Kay and Marple [71] give an excellent tutorial on resolution and super-resolution concepts in the guise of spectrum analysis. A history of the most popular techniques including the FFT, autoregressive, moving average, Burg, Pisarenko, Prony, and Capon methods using common notation allows an easy compari-

son of the advantages and disadvantages between these various techniques. An extensive discussion of super-resolution imaging techniques, including some not mentioned here, is also given by DeGraaf [40] and Mehra *et al.* [96] with comparisons of performance using both simulated and measured SAR imagery. Pastina *et al.* [110] give a short review of various algorithms with several useful references. A discussion of the general super-resolution problem from the point view of degrees of freedom and the space-bandwidth product is given by Dickey *et al.* [43].

1.2.4 Target Recognition Using Super-resolution

Automatic target recognition has been studied in detail over the last fifty years but the idea of using super-resolution to aid radar target recognition has received serious attention for only about ten years. Super-resolution-aided target recognition has been proposed as a concept several times but both aspects of the problem are often not simultaneously tested either on real or simulated data: see [16, 42, 90, 110, 125, 134], for example.

Botha *et al.* [13] use a 2D version of the MUSIC algorithm to generate ISAR images of scale model aircraft. They compare image templates and high-level features, based on target shape and geometrical moments, using both a neural network and nearest-neighbour classification. The neural network and high-level feature combination is found to have the best generalization properties. However, a comparison between the MUSIC images and those produced by standard FFT processing is not made.

Zhang *et al.* [162] use the damped exponential Prony model and SVD to estimate scatterer positions. Features based on a wavelet transform of the super-resolved range profiles are extracted and used in a neural network for target recognition. The approach is compared to FFT processing and found to improve recognition performance.

Another approach using super-resolution range profiles where the target is modelled as a set of discrete scatterers is that of Liao and Bao [85]. The amplitude and position of the scatterers are determined using the RELAX method for both database range profiles and test profiles, and a graph-matching technique is used to determine the similarity between images. A beneficial side effect of this type of super-resolution is that the variation of target signature

with respect to aspect angle is lower in the super-resolved data than standard range profiles. This eases the task of database generation and storage. High classification rates are reported for measured data of three scale-model targets but the performance is not compared to that of any other technique so it is not clear from this paper alone whether or not there is an advantage in using the technique.

The HDVI super-resolution technique has been applied to both one dimensional range profile data [101] and two-dimensional SAR data [105] of ground vehicles in the moving and stationary target acquisition and recognition (MSTAR) measurement data set. Using standard FFT processing it was shown that target recognition performance increases with higher resolutions. When HDVI processing was used the target recognition performance was better than standard processing and was equivalent to a resolution improvement factor of two. The signature matching was based on the mean-square error metric for target images.

A point-enhanced SAR super-resolution technique based on generalized Tikhonov regularization has also been applied to the MSTAR data set [20]. The target recognition algorithm was based on selecting the strongest peaks of the super-resolved images for both test and training data and selecting the image that minimizes the total distance between matched peaks. Conventional FFT processing was also used with the peak picking process. It was shown that point enhancement provided a large performance improvement compared to FFT processing for a high signal-to-noise ratio. The same point-enhancement technique is used in [33] to process range profiles of ships. The Euclidean distance metric and a position-specific matrix matching algorithm based on quantized range profile amplitudes are compared using both standard FFT processing and the point-enhancement technique. Three sets of features were individually used with the Euclidean distance: range profile amplitudes, radar cross section, and ship length. In all cases apart from using the ship length feature, the point-enhancement technique improved classification performance. The position-specific matrix matching algorithm gave better results than the Euclidean distance. The point-enhancement technique has also been compared directly to Capon's ML method and MUSIC in an ATR framework using ISAR turntable data of tanks [62]. The features are range-profile magnitudes derived

from the ISAR images and the classifier is Bayesian with an assumed multivariate Gaussian feature distribution. The point-enhancement technique improved classification performance over standard processing and the ML and MUSIC algorithms were worse than standard processing.

In other tests performed on a different data set and using different algorithms, classification performance is shown by Mehra *et al.* to be improved when using super-resolution for both training and test data [96,97]. This is even the case for targets hiding in a forest when using a fully polarimetric foliage-penetrating radar sensor. The recognition algorithm was based on Fisher linear discriminant analysis for feature reduction and a Euclidean distance for signature matching. However, it was not stated which super-resolution method was used. In related work on foliage penetration it is shown that using either fully polarimetric data or super-resolution on their own improves target recognition performance over a single standard-resolution polarization channel [120]. However, the combination of polarimetric data and super-resolution offers no advantage over using either technique on its own.

A study by Kim *et al.* used the MUSIC algorithm to produce super-resolved range profiles of scale-model aircraft [73]. Feature vectors were based on normalized moments of range profiles followed by feature reduction using principal components analysis. A Bayesian classifier based on Gaussian statistics was used to recognize the targets. The MUSIC approach had a better recognition performance than using standard FFT processing and, for both types of processing, performance improved with resolution. This study was extended in [72] by the same authors to compare the AR bandwidth extrapolation method with MUSIC and FFT processing, within the same target recognition framework. For moderate signal-to-noise ratios the AR method outperformed both MUSIC and FFT processing. It was also shown that there was no additional improvement in performance if the bandwidth was extrapolated by more than a factor of three.

Other approaches to super-resolution and target recognition have been proposed and tested on their own but not compared directly to other techniques. Zhang *et al.* [163] use the matrix pencil method to estimate scatterer locations. Features are based on scatterer polarimetric properties and a neural net is used to perform the classification. Zwart *et al.* [165] use maximum likelihood and

expectation maximization both to obtain scatterer locations of simulated aircraft and to match the test data with the database data, while considering the matching of point locations as an assignment problem. Radoi *et al.* [117, 118] use the 2D MUSIC algorithm to process ISAR data of scale-model aircraft and a neural network for target recognition. Cui *et al.* [34] use 2D MUSIC and AR bandwidth extrapolation to obtain a high-resolution image of ground targets and their shadows as recorded in the MSTAR data set. However, the features used for the target recognition stage consist only of low-frequency components of the high-resolution images and do not make full use of the super-resolution process.

1.3 Motivation for Proposed Algorithms

The work outlined in the literature survey shows that super-resolution can indeed improve target recognition performance over standard FFT processing. However, in all the work cited above, algorithms have been selected in a rather *ad hoc* manner with each research group using their favourite combination of super-resolution technique, feature set, and pattern recognition algorithm. There is no sense of optimality or theoretical justification for the overall target recognition process or even certain individual steps. Also, none of the combinations of super-resolution and target recognition take into account uncertainty in the super-resolved images. This could be critical for target recognition systems because any spurious information potentially introduced by a super-resolution algorithm might result in an incorrect association between two unrelated targets.

The main objective of the work presented in this thesis is to improve the overall target recognition performance of radar systems where different images have differing resolutions. It is proposed to use the full probability distribution of super-resolved images as an input to the target recognition process. The distribution will allow, for example, the determination of whether a spike in an image is stable in amplitude and likely to be related to a specific scattering event on a target, or whether it is more likely to be part of a wider distribution of noise values introduced by the super-resolution algorithm. This process will be carried out using a numerical Markov-chain Monte-Carlo Bayes-

ian super-resolution algorithm, and an automatic target recognition model that incorporates super-resolution uncertainty. A measure of the approach's success is the achievable gain in the average probability of correct target classification for a given scenario over a basic super-resolution and target recognition processing. In addition to this, the Bayesian approach is theoretically justified and explicitly states all assumptions made in the data model.

A variety of effects that degrade super-resolution performance are analyzed in more detail than has currently been done in the literature, and techniques to compensate for these are presented. The usefulness of these techniques is based on whether the loss in performance can be recovered to any extent.

1.4 Thesis Layout

Chapter 2 gives a brief introduction to the radar, super-resolution and target recognition theory useful for understanding this thesis. Chapter 3 introduces the concept of Bayesian super-resolution. A new Markov chain Monte Carlo Bayesian super-resolution algorithm is presented with results of the algorithm using a simulated data set. Algorithm complexity is analyzed and an approximate Bayesian solution is proposed. Chapter 4 suggests a variety of factors that could degrade the baseline super-resolution performance seen in Chapter 3. A short review of techniques that attempt to restore baseline performance is given, followed by a new technique integrated with the Bayesian super-resolution algorithm. Chapter 5 further integrates the work of the previous chapters into a single Bayesian ATR framework and considers performance of the framework using both simulated and measured data. Finally, conclusions and recommendations for further work are given in Chapter 6.

Chapter 2

Background Theory

2.1 Radar

2.1.1 Introduction

This section gives a brief overview of radar theory useful to the reading of this thesis. A comprehensive description of radar systems is available in textbooks such as [136, 142].

Radar is a system for the detection and location of objects using the radio portion of the electromagnetic spectrum. A pulse of radio waves is sent from the transmitter into a region of interest. Energy reflected in the direction of the receiver by targets in the illuminated scene is detected in the receiver and indicates the presence of targets. The distance to a target may be calculated from the round trip delay of the pulse and the speed of radio wave propagation. Each pulse generates a range profile, which is a measure of the received energy as a function of distance from the radar.

2.1.2 Signal-to-Noise Ratio

In the simplest instance, the ability to detect a target is governed by the signal-to-noise ratio (SNR) of a radar receiver system. During the detection process a threshold is set and if the voltage in the receiver exceeds this threshold then a detection is declared. However, in all electronic circuits there is a certain level of thermal noise caused by electrons moving in random directions. Due to the random nature of thermal noise it is possible that the detection threshold may

be exceeded even though no target is present. This is known as a false alarm. If the average noise power is known then the threshold may be set to give a certain acceptably low probability of false alarm. Target detection is covered in more detail in Section 2.4.2. The issue of system noise will be seen to be critical for the successful operation of super-resolution algorithms.

One form of the radar range equation gives the signal-to-noise ratio as:

$$\frac{S}{N} = \frac{PG^2\lambda^2\sigma}{(4\pi)^3 kTB_nF_nR^4} \tag{2.1}$$

where:

S is the signal power received by the radar;

N is the power of the thermal noise generated in the receiver;

P is the power of the transmitter;

G is the gain of the antenna;

λ is the wavelength of radiation;

σ is the radar cross section of the target;

k is Boltzmann's constant;

T is the receiver temperature;

B_n is the receiver's noise-equivalent bandwidth;

F_n is the noise factor; and

R is the range from the radar to the target [136].

2.1.3 Resolution

It is required that radar systems are able to detect more than one target at a time. If two targets are at the same bearing from a radar but different distances then they may be detected individually as long as their separation Δr is greater than the range resolution of the system. This is illustrated in Figure 2.1 for a rectangular pulse. The spatial length of a radar pulse is $c\Delta\tau$, where c is the speed of radio wave propagation and $\Delta\tau$ is the length of the pulse in time. To

(a) Transmitted pulse (b) Reflected pulses

Figure 2.1: Resolution for a rectangular pulse. (a) Single pulse transmitted towards two targets. (b) Pulse reflected from two targets.

resolve two targets they must be separated by at least half the length of the pulse, so $\Delta r = c\Delta\tau/2$. The Fourier transform of a rectangular pulse with unit amplitude and width Δt is the scaled sinc function shown in Figure 2.2 [60]. The bandwidth, B, of the pulse is equal to $1/\Delta\tau$, which gives the down-range resolution of a radar in terms of the bandwidth as $\Delta r = c/(2B)$ [106]. Therefore, the higher the bandwidth the better the resolving power of the radar.

High bandwidth can be achieved with a very short pulse. This requires a high peak power for a given amount of energy in a pulse. Pulse compression is a technique used in radar systems to achieve both a high transmit energy and good range resolution by frequency or phase modulation of a long pulse, thus reducing the peak power requirement. When pulse compression is used it is the bandwidth of the system that gives rise to a resolution of $c/(2B)$, regardless of the length of the uncompressed pulse. The optimal receiver filter, in terms of the peak-signal-to-mean-noise power ratio, is the matched filter with an impulse response that is a time-reversed copy of the transmitted waveform [136]. A consequence of the filtering process is the production of range sidelobes, which mask weak targets in the presence of stronger ones. Although amplitude weighting may be used in the receive filter to reduce sidelobes, the weighting process increases the width of the main lobe and hence resolution is degraded. Therefore further techniques are required to achieve an improved detection performance while maintaining resolution.

An example showing the effect of resolution under ideal noiseless conditions

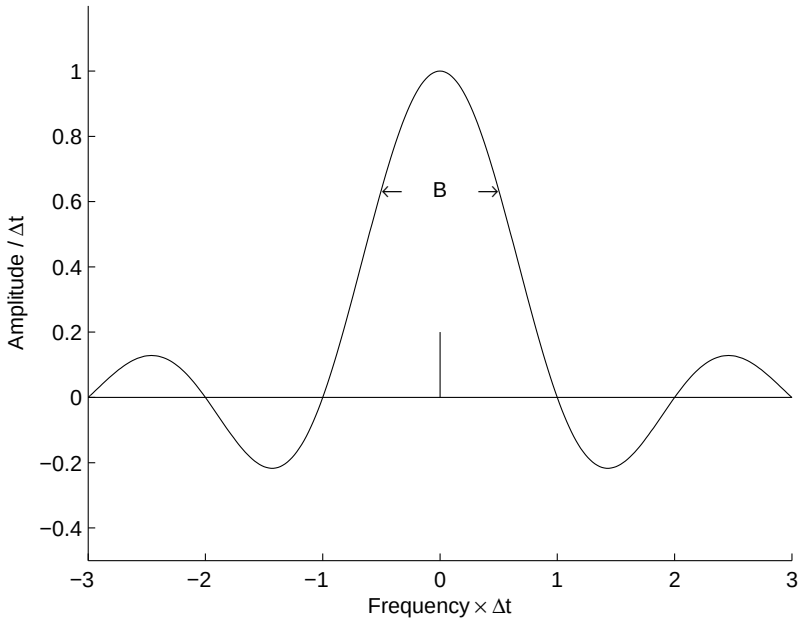


Figure 2.2: A scaled sinc function – the Fourier transform of a unit amplitude rectangular pulse of width Δt.

for two closely spaced targets is shown in Figures 2.3 and 2.4. In Figure 2.3 the targets, whose positions are marked by the vertical lines, are separated by a distance equal to the system resolution. The system response, indicated by the solid line, clearly shows the presence of two targets. However, note that interference between the sidelobes of one target and the main lobe of the other means the main peaks in the response are slightly shifted from the true target positions. In Figure 2.4 the targets are separated by a distance equal to half the system resolution. It is not possible to determine whether there are two targets or a single stronger target in the middle, without advanced processing.

The cross-range resolution of a non-coherent radar is approximately $R\lambda/D$, where R is the distance from the radar to the target, λ is the radio wavelength, and D is the antenna aperture dimension [106]. Thus, the cross-range resolution of a non-coherent radar is limited by the physical size of the antenna: the larger the antenna, the better the resolving power. A similar process to the matched filtering used in obtaining high down-range resolution can improve the cross-range resolution of the system. This is covered in the coherent imaging section below.

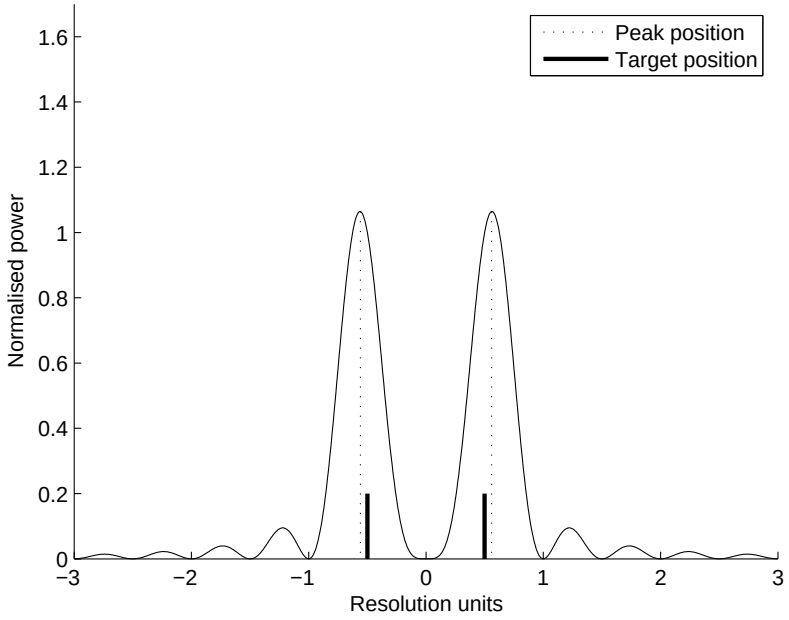


Figure 2.3: Example of two resolved targets

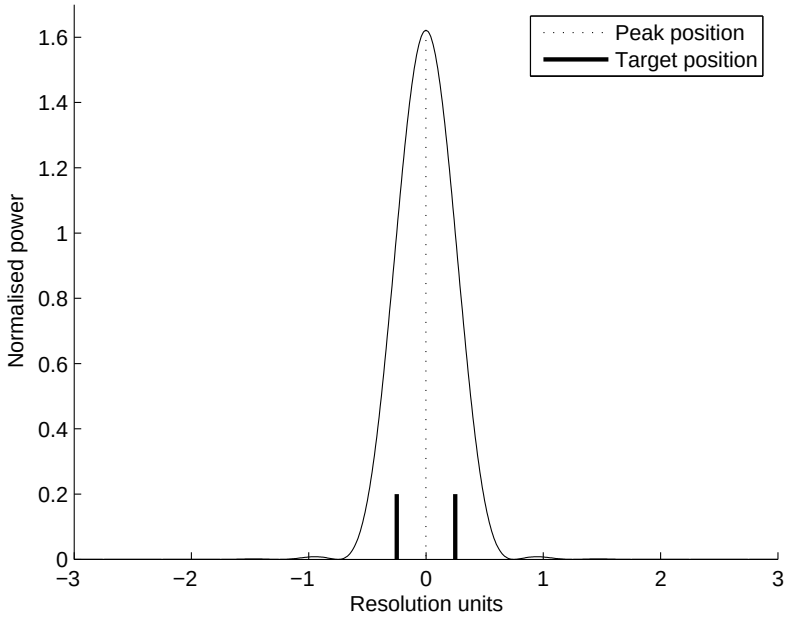


Figure 2.4: Example of two unresolved targets

2.1.4 Ambiguity Function

The radar system response to a single ideal point target as a function of distance or time is known as the point spread function (PSF). It will later be seen that deconvolution-based super-resolution techniques require knowledge of the PSF. For a matched filter, the PSF of a stationary target is obtained as the auto-correlation of the transmitted waveform. Due to the Doppler effect, moving targets alter the received signal and the correlation equation must be modified to include this. The function describing the PSF for various Doppler shifts is

$$\chi(T_R, f_d) = \int_{-\infty}^{\infty} u(t)u^*(t + T_R)e^{j2\pi f_d t}\, dt \tag{2.2}$$

where T_R is the time delay to target relative to a reference delay, f_d is the target Doppler shift and is positive for an incoming target, $u(t)$ is the complex transmitted signal, and $u^*(t)$ represents its complex conjugate [136]. The squared magnitude $|\chi(T_R, f_d)|^2$ is called the ambiguity function and was first introduced by Woodward [158]. The ambiguity function is an important description of radar performance – it determines the range resolution and sidelobe levels of a pulse as well as the Doppler resolution and range-Doppler coupling.

Figure 2.5 shows the ambiguity function of an unweighted, linear frequency modulated (FM) pulse. Several features are apparent in this diagram. The PSF for a stationary target is formed by taking a horizontal cut of the diagram at zero Doppler. From this it is possible to determine the range resolution of the system. By taking a vertical cut at zero range, the Doppler response is formed, from which the Doppler resolution can be determined. The Doppler resolution is the ability to detect two targets at the same range but traveling with different speeds. The ambiguity function shows a coupling between range and Doppler for linear FM pulses – if a target is moving then it will appear at a different range to its true position with the offset proportional to its velocity. In practice, this offset is small but a moving target does alter the point spread function enough to affect super-resolution techniques [79].

In reality the radar system has a two-dimensional PSF relating to both range and cross-range. The cross-range variation is not usually considered in terms of the ambiguity function and is normally displayed or described in conjunction with the down-range resolution at zero Doppler in a 2D plot or equation. Example 2D PSFs are shown in 2.6 to demonstrate the concept. Figure 2.6a shows

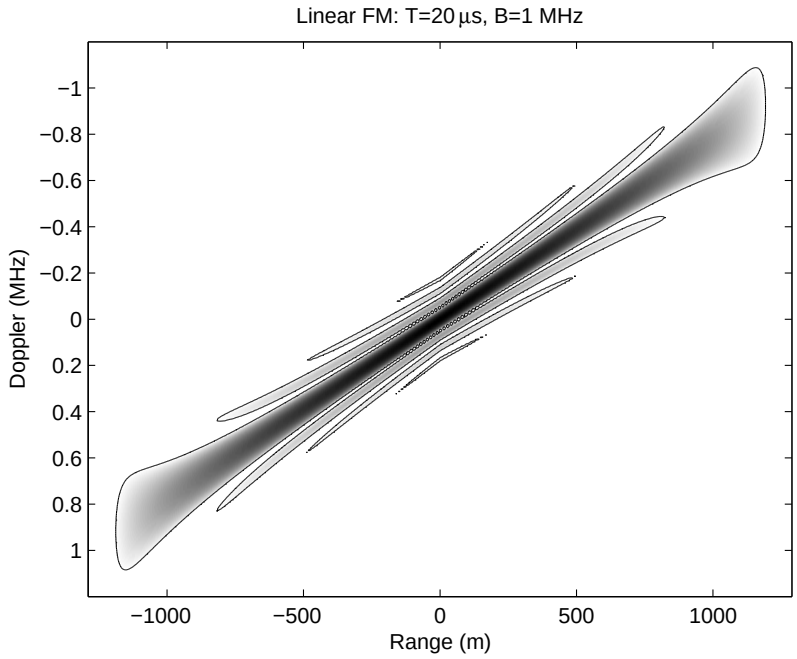
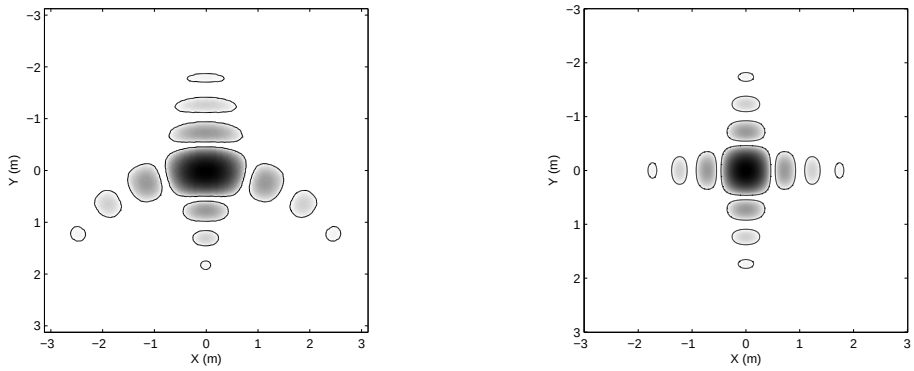


Figure 2.5: Ambiguity function of a linear FM pulse. Pulse length 20 μs, bandwidth 1 MHz. Detail below a -22 dB threshold is not shown.

the PSF of a non-coherent radar. The curved shape of the PSF is due to the fact that the response is a function of range and angle. The down-range resolution of the radar is constant over the whole image but, because the angular resolution is constant, the cross-range resolution is worse at longer distances from the radar. Figure 2.6b shows the PSF of a coherent radar. Here the down-range and cross-range resolutions are both constant for the whole image, to a first approximation. The 2D PSF is required to be known for the super-resolution of 2D images.

2.1.5 Coherent Imaging

In airborne synthetic-aperture radar (SAR) the radar antenna is mounted on a platform, flown along a specified path, and successive pulses are coherently combined. For strip-map SAR, the radar beam points at a constant angle relative to the motion of the aircraft. The imaging geometry of this is shown in Figure 2.7. Strip-map SAR systems with a synthetic-aperture length L_{SA} have a cross-range resolution of $R_0\lambda/(2L_{SA})$, which is an improvement over non-coherent radar.



(a) Non-coherent PSF (b) Coherent PSF

Figure 2.6: Example 2D point spread functions. (a) Non-coherent PSF with a range resolution of 1m and an angular resolution of 35^o. (b) Coherent PSF with a range resolution of 1 m and a cross-range resolution of 1 m. Detail below a -22 dB threshold is not shown.

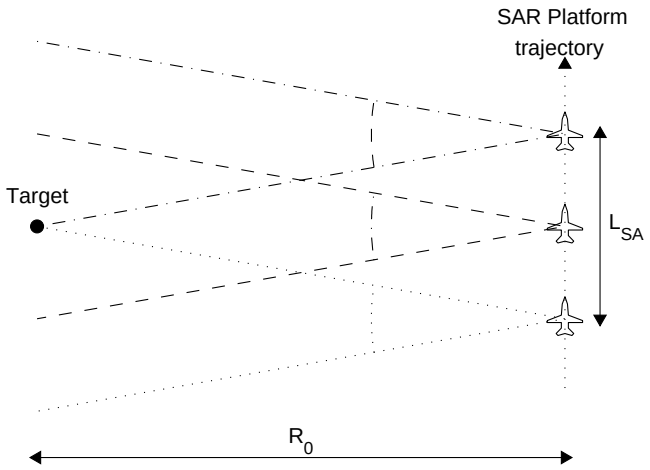


Figure 2.7: Imaging geometry for strip-map SAR

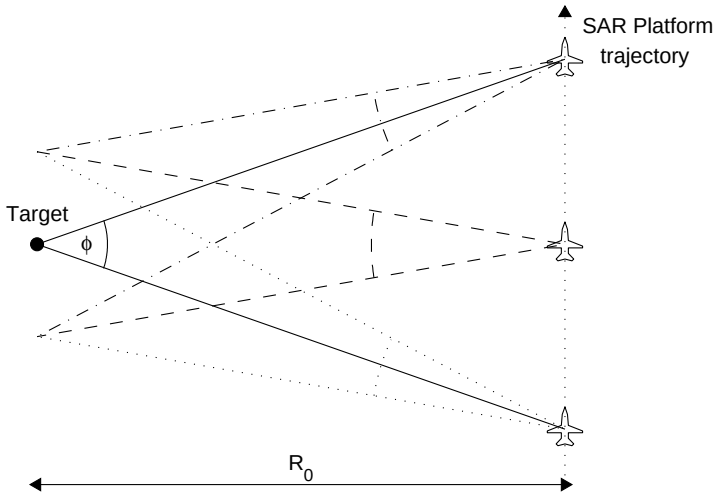


Figure 2.8: Imaging geometry for spotlight SAR

In spotlight SAR the radar beam points to a constant position on the ground; therefore relative to the moving radar platform the beam sweeps in a backward direction. The imaging geometry for spotlight SAR is shown in Figure 2.8. The cross-range resolution of this SAR mode is approximately $\lambda/(4\sin(\frac{\phi}{2}))$, where ϕ is the angle swept by the beam [18]. When ϕ is small and is measured in radians this reduces to $\lambda/(2\phi)$. Spotlight SAR achieves a better resolution than that of both strip-map SAR and non-coherent radar.

Another form of coherent imaging is inverse synthetic-aperture radar (ISAR). In ISAR imaging, the radar remains in a static position and the beam tracks a moving or rotating object. The principles are the same as spotlight SAR and the resolution of an ISAR system is also $\lambda/(4\sin(\frac{\phi}{2}))$, where ϕ is the angle through which the target rotates relative to the radar. A commonly used technique for gathering radar data representative of spotlight SAR is to place a target on a rotating turntable and record ISAR data using a fixed radar. This is usually a less expensive means of obtaining data than flying an aircraft.

All forms of coherent imaging require the phase of a signal as well as its magnitude. This makes coherent imaging systems more susceptible to degradation in a variety of non-ideal situations, when compared to non-coherent radar.

The major sources of degradation for airborne SAR are non-linear motion of the radar platform and moving targets. The manifestation of the degradation is defocusing or blurring of the radar image. Oliver [106] gives a good review of defocus effects.

Most airborne systems carry an inertial measurement unit (IMU) that measures vibrations and deviations of the aircraft from a straight-line trajectory. These measurements are used to make phase corrections to the data before standard processing. However, even with the use of an IMU residual errors may remain, which result in an image with a non-ideal point spread function. The change in PSF affects the performance of super-resolution techniques, which is one of the motivations behind the Bayesian super-resolution algorithm introduced in this thesis: it can take into account uncertainty in the PSF.

Moving targets are more difficult to deal with than non-linear motion of the radar platform. It is possible to adjust standard processing to take a single target's motion into account. However, this has the effect of blurring the surrounding image and if multiple targets are moving at different speeds they cannot all be focused at the same time. Vibrations in moving or stationary targets also defocus the image and it is difficult to compensate for these effects. These problems are not dealt with in the Bayesian algorithm introduced in this thesis as they are difficult to parameterize. Advanced autofocus techniques must therefore be used to restore image quality – see [48, 74, 91, 112, 156] for example.

2.2 Super-resolution

2.2.1 Introduction

It was seen in the previous section that there is a fundamental physical limit to both the cross-range and down-range resolution of a radar system. The non-ideality of the radar system is represented by the point spread function (PSF). Algorithms that attempt to remove the effect of the system PSF are known as deconvolution or super-resolution algorithms. Application of the algorithms allows the separation into multiple peaks of two or more closely-spaced targets that form a single peak at the nominal system resolution. Generally the algorithms work by using knowledge of the PSF and making assumptions about the

scene being imaged. In the literature survey we saw there were two types of scene model: the parametric scattering centre model and the non-parametric continuum scattering model. Here we concentrate on the second type of model.

Each target is assumed to have a high-resolution two-dimensional plan-view back-scattering coefficient or distributed radar cross section (RCS), which is a function of imaging geometry. When imaged by a coherent radar the RCS gives rise to a complex scattered field value. The RCS-to-field scattering model is described in Section 3.1.2. The image formation process is equivalent to the convolution of a 2D PSF and the high-resolution 2D complex representation of the target. This effectively places an amplitude-scaled copy of the PSF at each location on a sampled grid. A bulk phase is present in the complex image, which is due to the distance between the radar and the target. In addition to this, the various scattering mechanisms on the target have a phase relative to each other dependent on their precise position and construction, and thus coherently sum to produce a single complex number at each sampling point. The imaging process is described mathematically by

$$\mathbf{g} = \mathbf{T}\mathbf{f} + \mathbf{n}, \tag{2.3}$$

where $\mathbf{f}$ is a complex vector denoting the raster-scanned high-resolution 2D target representation, $\mathbf{T}$ is an appropriately formatted Toeplitz convolution matrix that applies the effect of the PSF (see Appendix A), and $\mathbf{g}$ is a complex vector representing the resultant low-resolution image. Surrounding clutter is implicitly included in the target representation through the use of non-zero complex values at positions in the vector $\mathbf{f}$ corresponding to spatial positions near a target but not actually within the target boundary. Thermal noise in the radar receiver is modelled by $\mathbf{n}$, which is a zero-mean circularly complex Gaussian with a diagonal covariance matrix $\mathbf{N}$.

It should be noted the above model assumes that target images are a linear superposition of functions of the target representation $\mathbf{f}$. In reality there are a number of effects that this model does not encapsulate. Multiple reflections between different parts of the target will create spurious detail in an image at ranges greater than the position of the initial reflection back towards the radar. Effects such as edge diffraction, point diffraction and obscuration are also not taken into account. The radar hardware may have some non-linear components, which would distort the image produced during processing. Distortions

could also be introduced by features on a target with a frequency-dependent response. Naturally, since these effects are not included in the model, any super-resolution technique based on this model cannot take them into account. However, although these effects are present, it is assumed that they do not have a substantial impact on the quality of the processed image or overall system performance.

The input information-theoretic signal-to-noise ratio (SNR) of the image is defined as

$$SNR_{in} = \frac{||\mathbf{Tf}||^2}{\sum diag(\mathbf{N})}, \tag{2.4}$$

where $diag(\mathbf{N})$ is a vector containing the main-diagonal elements of $\mathbf{N}$.

Estimation of the scene $\mathbf{f}$, given the measurements $\mathbf{g}$, is the super-resolution problem we are concerned with. If the estimated scene for a particular algorithm is $\hat{\mathbf{f}}$ then the output SNR is defined as

$$SNR_{out} = \frac{||\mathbf{f}||^2}{||\hat{\mathbf{f}} - \mathbf{f}||^2}. \tag{2.5}$$

It is only possible to calculate this metric when the true scene is known.

Note that the information-theoretic definition of SNR used for input and output SNRs is different from that used for target detection [88]. The SNR for target detection is defined as the peak-signal-to-mean-noise power ratio. This SNR definition is appropriate for target detection because it is only required to measure the strongest return from a target to enable it to be detected. However, in target recognition, information about other parts of the target is required and it is important to measure the power of the whole target in relation to the noise. The information-theoretic SNR is effectively the mean-signal-to-mean-noise power ratio. This measure of SNR also allows a direct comparison between input and output SNRs that would not be possible with the peak-only version. If the point spread function is a sharp peak with no sidelobes then for an ideal point target the information-theoretic SNR is smaller than the detection SNR by a factor equal to the number of samples in the calculation. For example a point target in a 1000-element range profile with an information-theoretic SNR of 15 dB would have a detection SNR of 45 dB. However, an extended target, of the same average power per range cell as the previous example, that fills the entire range profile would have both an information-theoretic and detection

SNR of 45 dB. In practice a finite-length target and non-point-like PSF result in a situation somewhere between these two extremes.

It is in fact difficult to define a single metric with which to measure super-resolution performance. In addition to the two metrics just mentioned several others have been proposed. Blacknell [7] proposes a heuristic performance metric based on the position and amplitude of recovered scatterers. Dickey *et al.* [43] prefer to use *super-resolution gain* – a metric that determines the degree to which a super-resolved image is similar to the true high-resolution scene as opposed to the low-resolution image. One problem with super-resolution performance metrics in general is that they each assume a different model. In the case where the metric is based on determining the locations of discrete scatterers, for example, it would be impossible to assess the performance of an algorithm that uses the continuum scattering model. A more fundamental problem with most metrics is that they require knowledge of the true scene. In simulated data the true scene is known because it is required to construct the simulation. However, when using measured data the true scene cannot perfectly be known. Nevertheless it is possible to use measured data to test super-resolution algorithms by starting with a high-resolution measurement and assuming this is the true scene. The data can have its resolution artificially degraded followed by the addition of noise. The result of super-resolution algorithms applied to the degraded image can then be compared with the original high-resolution image. Both the simulated-data and measured-data approaches are used in this thesis.

An overview of the theory behind several algorithms for solving the super-resolution problem under the above model is now given. At the end of this section a comparison of these algorithms is made using a simulation of a simple scene containing point targets.

2.2.2 Matrix Inverse

The benchmark solution against which other super-resolution algorithms should be compared is the Moore-Penrose matrix pseudoinverse. This is a well-known least-squares procedure for solving systems of linear equations and is the basis for super-resolution algorithms such as [126]. The approach minimizes, with respect to the recovered scene $\mathbf{f}$, a cost function

$$J = ||\mathbf{g} - \mathbf{T}\mathbf{f}||^2, \tag{2.6}$$

which is the sum of squared differences between the measured image $\mathbf{g}$ and the image $\mathbf{Tf}$ that would be produced given the solution $\mathbf{f}$. Differentiation of the cost function with respect to $\mathbf{f}$ and setting the result equal to zero gives the estimate for $\mathbf{f}$ as

$$\hat{\mathbf{f}}_{inv} = (\mathbf{T}^H\mathbf{T})^{-1}\mathbf{T}^H\mathbf{g}. \tag{2.7}$$

The H superscript denotes the Hermitian (complex conjugate) transpose of a matrix. The least-squares solution gives good results for general well-conditioned algebraic problems of the form $\mathbf{g} = \mathbf{Tf}$, where small changes due to noise in the measured data $\mathbf{g}$ result in small changes in the solution $\mathbf{f}$. This occurs in situations where the generalized condition number of the matrix $\mathbf{T}$, $k(\mathbf{T}) = ||(\mathbf{T}^H\mathbf{T})^{-1}\mathbf{T}^H||.||\mathbf{T}||$, is close to unity. The condition number is effectively a noise multiplier. However, the specific super-resolution problem we are interested in is ill-posed as stated due to the Toeplitz structure of the PSF convolution matrix. The condition number of $\mathbf{T}$ in this scenario is high, and small changes in $\mathbf{g}$ result in large changes in $\mathbf{f}$. Super-resolution performance using the pseudo-inverse algorithm is therefore usually poor due to the ill-conditioned nature of the problem – a large number of spurious scatterers are often placed at positions where there is no valid target. A general discussion of ill-posed and inverse problems is given in [128].

2.2.3 Minimum Mean-Square Error

A similar approach to least squares is the minimum mean-square error (MMSE) technique. The object of the MMSE approach is to choose the linear operator $\mathbf{R}$, such that the super-resolution solution given by $\hat{\mathbf{f}}_{mmse} = \mathbf{Rg}$ minimizes the expected norm J of the reconstruction error:

$$J = < ||\hat{\mathbf{f}}_{mmse} - \mathbf{f}||^2 > . \tag{2.8}$$

If the covariance matrix of the *a priori* statistical distribution of $\mathbf{f}$ is $\mathbf{W}$ then the solution is given by [87] as

$$\hat{\mathbf{f}}_{mmse} = \mathbf{WT}^H(\mathbf{TWT}^H + \mathbf{N})^{-1}\mathbf{g}. \tag{2.9}$$

In this basic form when $\mathbf{W}$ is fixed in advance, the MMSE solution is also known as a Wiener filter. However, in practice $\mathbf{W}$ is not known in advance because

prior knowledge of the scene is not well defined. Therefore an iterative scheme that estimates the $\mathbf{W}$ matrix from measured data must be used [41,90]. In one scheme, $\mathbf{W}$ is set to the identity matrix for the first iteration and in subsequent iterations it is estimated from the $\hat{\mathbf{f}}$ of the previous iteration. Any element of $\hat{\mathbf{f}}$ whose power is below a threshold is assumed to be homogeneous clutter and the combined variance of all these elements is calculated and entered into the appropriate diagonal elements of $\mathbf{W}$. Any element i above the threshold has its variance set to $|\hat{f}_i|^2$ and entered into the appropriate diagonal element of $\mathbf{W}$. Therefore $\mathbf{W}$ is always diagonal and has a high variance where there is a valid target or a large scatterer. The procedure is repeated by alternately estimating $\mathbf{W}$ and $\mathbf{f}$ until a termination criterion is met when $\mathbf{f}$ changes by less than a specified amount between iterations. An alternative algorithm is also possible where all diagonal elements of $\mathbf{W}$ are set to $|\hat{f}_i|^2$ between iterations. The two algorithms are referred to here as thresholded minimum mean-square error (MMSE-T) and all-pixel minimum mean-square error (MMSE-A), respectively. Another algorithm that works in a similar manner to MMSE-T, by partitioning the data into target and clutter pixels, is semi-sparse MMSE [133]. However, that algorithm requires prior knowledge of target and clutter statistics. The dual-model super-resolution technique of [16] also divides the data into target and clutter regions.

If $\mathbf{W}$ is fixed to be the identity matrix and the noise level is set to zero then a single iteration of the MMSE algorithm is identical to the matrix inverse. This demonstrates a failing of the matrix inverse algorithm – it effectively assumes no noise is present in the measured data, over-fits the data, and results in a very noisy output. In the MMSE algorithm, the noise covariance matrix $\mathbf{N}$ acts as a regularization parameter, allowing there to be a difference between the measured data and modelled data, commensurate with the noise level. A disadvantage of the basic MMSE algorithm is the requirement to know the noise power in advance but there are a number of ways to estimate it. One of these is to make a measurement of an area in the scene where there are known to be no targets or clutter present. This could be done by selecting a shadowed area of an image and calculating the variance of the pixels in that area. However, it may be difficult to automatically segment the image into shadow and non-shadow areas. A different way of obtaining noise-only data would be to have

the radar point at an area of the sky where there are no targets. Alternatively, rather than measuring it, the noise power could be calculated directly from the formula kTB_nF_n, where k is Boltzmann's constant, T is the temperature of the radar receiver, B_n is the noise-equivalent bandwidth and F_n is the noise factor. It would be advantageous if the noise power could be estimated without image segmentation or the need to make independent measurements because these operations introduce extra complexity into the system. One procedure for automatic noise power estimation integrated with the MMSE technique is given in [79].

2.2.4 Singular Value Decomposition

In singular value decomposition (SVD) super-resolution the PSF matrix $\mathbf{T}$ is decomposed into three matrices, $\mathbf{T} = \mathbf{USV}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and $\mathbf{S}$ is a diagonal matrix whose diagonal elements comprise the singular values s_i. Any singular value below a certain threshold is set to zero and the inverse of the new matrix is calculated as $\mathbf{T}_{svd}^{-1} = \mathbf{VS}^{-1}\mathbf{U}^H$ with the inverse of the zero values also set to zero [116]. Thus the SVD solution is given by

$$\hat{\mathbf{f}}_{svd} = \mathbf{T}_{svd}^{-1}\mathbf{g}. \tag{2.10}$$

If all singular values are retained then this is equivalent to the matrix inverse algorithm. Removal of several small singular values, which are responsible for the amplification of noise, results in estimated scenes with a higher SNR than the standard matrix inverse. In fact, the condition number of a matrix is equal to the ratio of the largest and smallest singular values so it would be expected that the algorithm should show an improvement over the matrix inverse. However, care must be taken not to remove too many singular values as each column of $\mathbf{U}$ or $\mathbf{V}$ corresponds to a specific structure in $\mathbf{T}$. When a singular value is set to zero the output matrix $\mathbf{T}_{svd}$ misses that structure from the scene estimate and this results in a biased estimate. The threshold level for singular values is not defined in the basic algorithm as it depends on the specific application. However, during testing of the algorithm using measured radar data it was found that removing singular values whose square s_i^2 was below a threshold of s_{max}^2/SNR provided a good compromise between too much noise and too much bias – see Appendix B.

2.2.5 MUSIC

The standard MUSIC algorithm consists of calculating the eigenvalues of the received data matrix, using the eigenvalue spectrum to decide the number of scatterers D, calculating signal power as a function of position, and setting scatterer positions according to the D peaks of this function [129]. A modified version of the algorithm is used here where, in common with the other algorithms in this section, the scene is effectively assumed to contain a pre-determined number of scatterers that exist on a discrete grid of sampling points. This modified version thus only uses the power-as-a-function-of-position step of the full MUSIC algorithm and does not perform as well as the original algorithm using the model for which it was designed. The modified version is included here to allow a simple comparison with other super-resolution techniques. The estimated intensity of $\mathbf{f}$ is given by

$$|\hat{\mathbf{f}}|^2_{music} = (\mathbf{T}^H\mathbf{T})^{-1}\mathbf{T}^H(\mathbf{g}\mathbf{g}^H - \mathbf{N})\mathbf{T}(\mathbf{T}^H\mathbf{T})^{-1}. \tag{2.11}$$

In the MUSIC algorithm $\mathbf{N}$ is usually estimated from the eigenvalues of $\mathbf{g}\mathbf{g}^H$. For zero-mean identically and independently distributed noise $\mathbf{N} = \lambda_{min}I$, where λ_{min} is the minimum eigenvalue of $\mathbf{g}\mathbf{g}^H$ and $\mathbf{I}$ is the identity matrix. If $\lambda_{min} = 0$ then the estimated scene is identical to that produced by the matrix inverse. This is another result that shows the standard matrix inverse algorithm assumes no noise is present in the image.

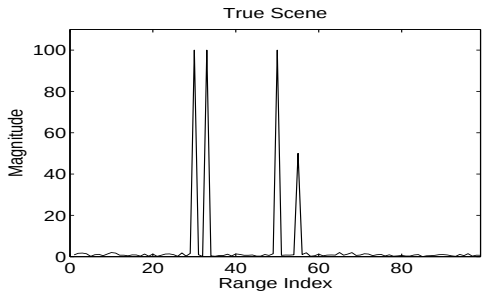
2.2.6 Comparison of Standard Algorithms

An example of how the above super-resolution algorithms perform is now presented. We consider a one-dimensional simulated scene $\mathbf{f}$ containing four point targets embedded in clutter. The scene is illustrated in Figure 2.9a. The first two targets are equal in amplitude and separated by approximately one quarter of the system resolution. The third and fourth targets are separated by approximately half the system resolution with the amplitude of the fourth target half that of the third. The targets are in-phase with each other; *i.e.* they are separated by integer multiples of $\lambda/2$. The clutter is uncorrelated with a random-phase and has a peak-to-mean signal-to-clutter ratio of 20 dB for the strongest targets. The point spread function of a signal with a bandwidth of B and no weighting is $\sin(\pi Bt)/(\pi t)$, which can be written in terms of the sinc

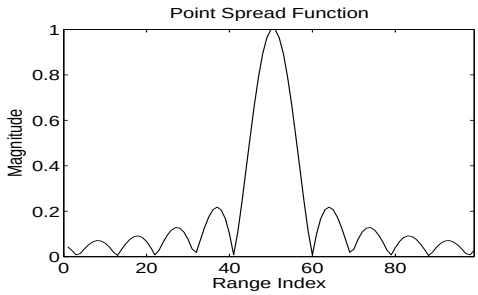
function as $B\text{sinc}(Bt)$ [60]. The magnitude of this is shown in Figure 2.9b. The one-dimensional noiseless "image" of the scene $\mathbf{Tf}$, which is the spatial-domain convolution of the true scene $\mathbf{f}$ with the PSF, is shown in Figure 2.9c along with a noisy version of the image ($\mathbf{g}$ in equation 2.3), where noise has been added at an SNR of 50 dB. This SNR has been chosen to illustrate operation of the algorithms in a relatively benign situation. At this high SNR the noiseless and noisy images are almost indistinguishable. It can be seen from the diagram that the pairs of closely-separated targets have each coalesced into a single peak and it is impossible to tell from the image alone how many targets are present. A number of significant sidelobes are also present in the image. It should be noted here that both the scene and image are complex-valued but the figures display only the magnitude of these quantities. The sinc function is real-valued but has alternating positive and negative sidelobes as shown in Figure 2.2.

Results of the matrix inverse, SVD, MUSIC, MMSE-A and MMSE-T super-resolution algorithms as applied to the noisy image are shown in Figures 2.9d to 2.9h, respectively. All algorithms apart from SVD have successfully resolved the four targets – SVD has resolved the second pair but not the more closely spaced first pair of targets. The matrix inverse and MUSIC algorithms have very similar results. This is because the input level of noise is small and the two algorithms behave similarly under these conditions. It is interesting to note the behaviour of the SVD algorithm as compared to the matrix inverse. The output noise level of the SVD algorithm is much lower than that of the matrix inverse, as expected. However, SVD seems to have spread the energy of the targets over a wider area, which has resulted in the unresolved first pair of targets. This suggests that too many singular values have been set to zero – if more were retained, then the scene estimate would be closer to that of the matrix inverse but with lower noise levels. The final two algorithms, MMSE-A and MMSE-T, give very similar results and are better estimators of the scene than the other algorithms – all four targets have been resolved and the output signal-to-noise ratio is much higher. Both algorithms have made reasonable estimates of the clutter but at a lower magnitude than in the true scene. MMSE-A has a few more clutter spikes unrelated to the true clutter than MMSE-T, which suggest that overall the MMSE-T algorithm has performed the best for this example.

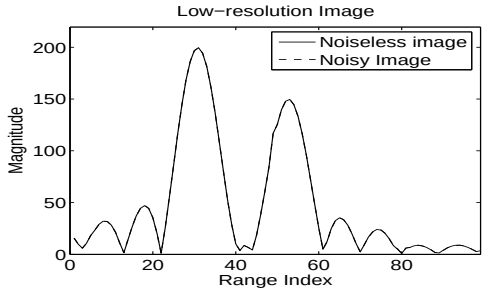
A second example is now given of the same scene but at a lower input



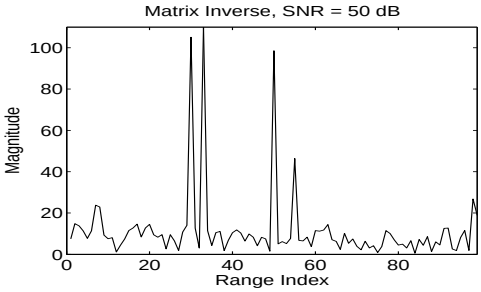
(a) True Scene



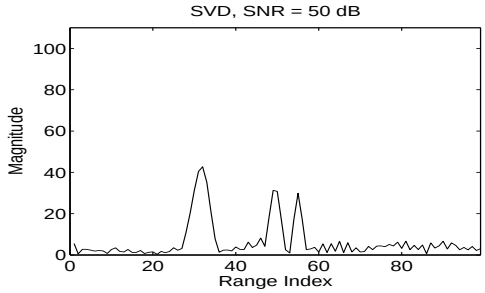
(b) Point Spread Function



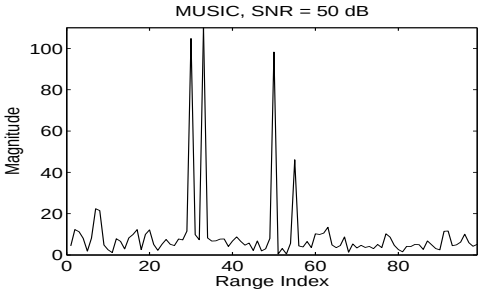
(c) Low-Resolution Image



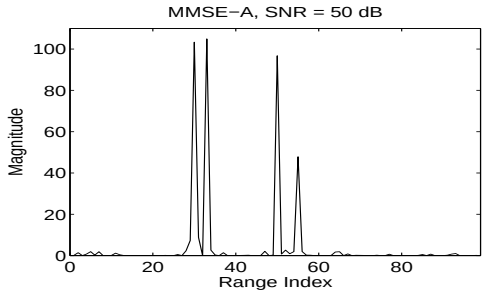
(d) Matrix Inverse algorithm



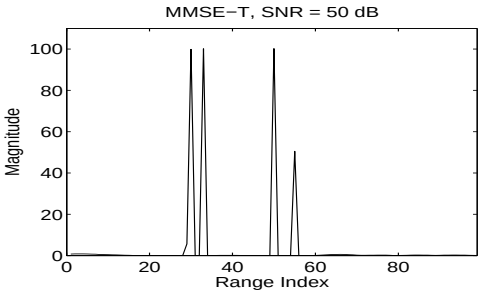
(e) SVD algorithm



(f) MUSIC algorithm

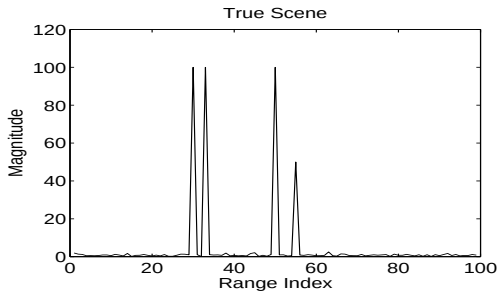


(g) MMSE-A algorithm

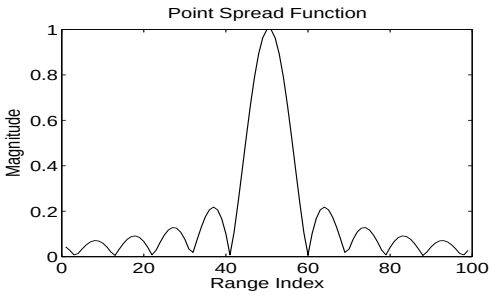


(h) MMSE-T algorithm

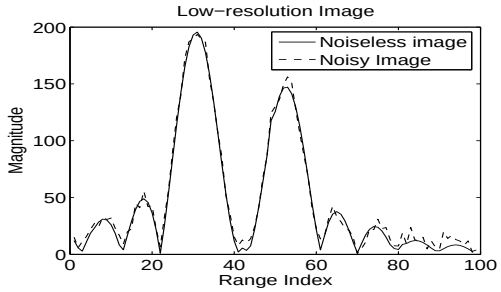
Figure 2.9: Simulation of a scene containing four targets and the results of five super-resolution algorithms applied to a low-resolution image of the scene. The signal-to-noise ratio of the low-resolution image is 50 dB.



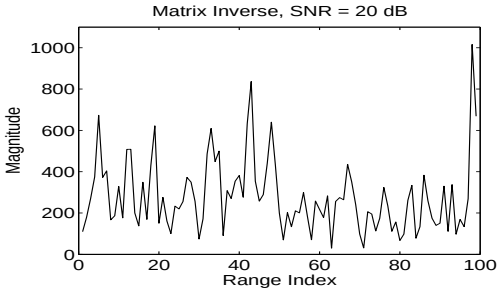
(a) True Scene



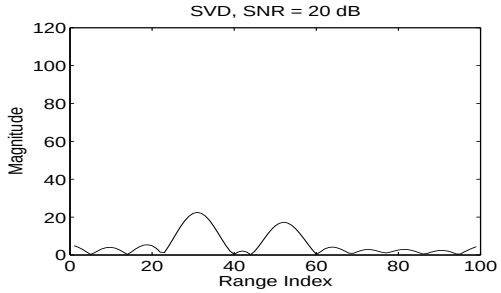
(b) Point Spread Function



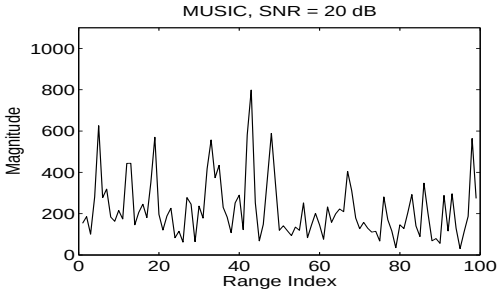
(c) Low-Resolution Image



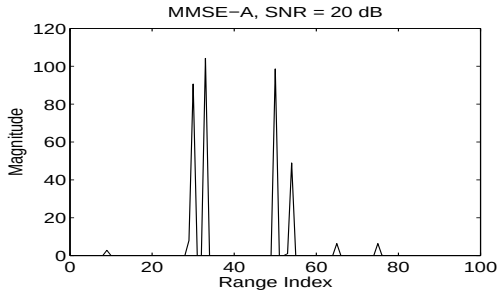
(d) Matrix Inverse algorithm



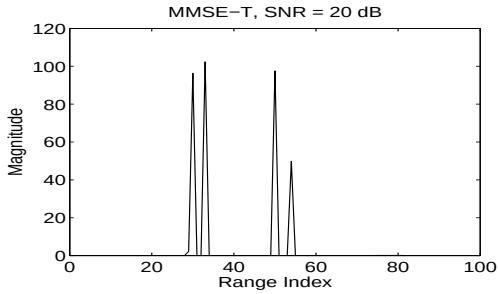
(e) SVD algorithm



(f) MUSIC algorithm



(g) MMSE-A algorithm



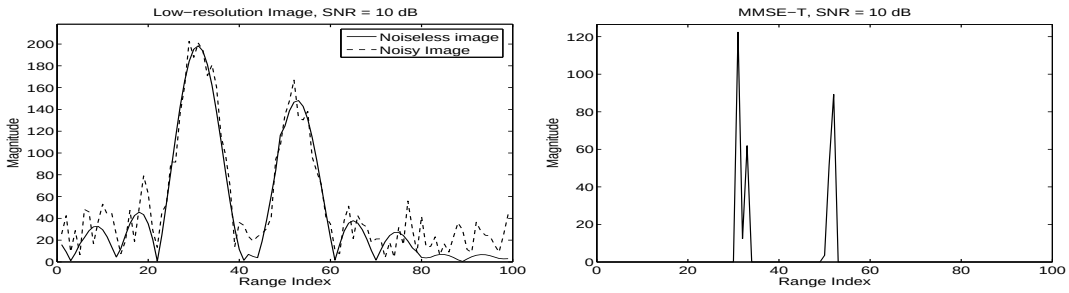
(h) MMSE-T algorithm

Figure 2.10: Simulation of a scene containing four targets and the results of five super-resolution algorithms applied to a low-resolution image of the scene. The signal-to-noise ratio of the low-resolution image is 20 dB.

signal-to-noise ratio of 20 dB. The scene, point spread function, and image are shown in Figures 2.10a to 2.10c. The difference between the noiseless and noisy images is now much more apparent than in the 50 dB SNR example. Results of the matrix inverse, SVD, MUSIC, MMSE-A and MMSE-T super-resolution algorithms as applied to the noisy image are shown in Figures 2.10d to 2.10h, respectively. The results at this lower SNR are much more variable than at 50 dB. Both the matrix inverse and MUSIC algorithms have completely failed to estimate the scene properly. The magnitude of the scene is generally wrong by a factor of about ten, none of the targets are visible, and the scene appears to consist solely of noise. This has happened because there is simply too much noise in the input image and the algorithms have over-fit the data. Note that the level of input noise would be easily low enough to allow target detection in the low-resolution image if each of the targets were separated by a distance more than the system resolution. Super-resolution is inherently a more difficult task than target detection and requires a much higher SNR for useful operation.

Performance of the SVD algorithm has degraded less severely than either the MUSIC or matrix inverse algorithms. For SVD, each pair of targets has merged into a single peak and overall the recovered scene is essentially a scaled version of the image so it could be considered to be a reasonable estimate for the scene given the high levels of noise. Therefore, although the SVD algorithm produces biased results, it is more robust with respect to noise than the matrix inverse algorithm. The MMSE-A and MMSE-T algorithms performed much better than the other algorithms – they have both resolved all four targets and there is very little noise in the output. Because the clutter level is the same as the noise level in this example, the algorithms have suppressed clutter as it is assumed to be noise. However, the MMSE-A algorithm has retained three low-amplitude spikes, which should be considered as output noise. There is not much difference between the MMSE-A and MMSE-T algorithm estimates but overall, considering more extensive simulations not reported here, the MMSE-T algorithm appears to have a slightly better performance than MMSE-A.

A final example is now given for an input SNR of 10 dB – a noise level where most algorithms would be expected to fail. The scene and point spread function are the same as in Figures 2.10a and 2.10b. The noisy image is shown in Figure 2.11a, where it can be seen that noise dominates the sidelobes and



(a) Low-Resolution Image (b) MMSE-T algorithm

Figure 2.11: Results of the MMSE-T algorithm applied to a low-resolution image of the scene. The signal-to-noise ratio of the low-resolution image is 10 dB.

even the two main peaks each contain several small noise spikes. This situation poses a formidable challenge to any super-resolution algorithm. The results for the matrix inverse, SVD and MUSIC algorithms are similar to those obtained for an SNR of 20 dB – either two small humps, in the case of SVD, or large noisy values, in the case of MUSIC and the matrix inverse. These results are not shown here to save space. The MMSE-A algorithm gives results similar to those for MMSE-T, which are shown in Figure 2.11b. Here we see MMSE-T has successfully resolved the pair of targets on the left but not the pair on the right. The algorithm has also overestimated the amplitude of the first target and the second target is approximately half the amplitude of the first, when it should be the same. These differences in amplitude could be critical when passed on to target recognition algorithms, especially ones such as the correlation classifier where the amplitude of a target at each pixel is the feature set used to discriminate between target classes. If it is known that the low SNR has caused imprecise amplitude estimates then this information could be used by the target recognition process to mitigate any variation. However, none of the algorithms presented thus far have a mechanism to do this.

The solution proposed by this thesis is to use a Bayesian super-resolution algorithm in conjunction with a Bayesian target recognition framework. In that manner uncertainty in the estimated scene and other relevant variables of the model can fully be taken into account [82]. It should be noted, however, that the Bayesian approach assumes the system model is an exact representation. If

the model is incorrect, there is additional uncertainty associated with this error. An overview of Bayesian statistics is now presented.

2.3 Bayesian Statistics

2.3.1 Introduction

This section sets out the essential statistical theory required to understand both the target recognition process introduced in the next section and the Bayesian super-resolution algorithm presented in the subsequent chapter. An emphasis is placed on Bayesian statistics – a particular interpretation of general statistical theory. For more detail on Bayesian data analysis see the textbooks [50] or [82].

Bayesian statistics is a way of representing the degree of belief in a statement or hypothesis in terms of probabilities, which may be manipulated using Bayes' theorem. As an example, consider the statement by a particular observer that "there is a 90% chance a tank is hiding behind those trees". Clearly either the tank is or is not hiding behind the trees. However, the observer has limited information on the situation and has had to make a statement based on what information is available to him. This could be knowledge of how often tanks hide in trees, whether or not it is known that tanks are in the surrounding area, or other information such as a recently-gathered radar image of the scene. The calculated probability depends on a mathematical model of the situation, assumptions about tanks and any measured data. Bayes' theorem provides a way to update the state of knowledge when new information is made available and is written as

$$P(H_1|D) = \frac{P(D|H_1)P(H_1)}{P(D)}, \tag{2.12}$$

where $P(H_1|D)$ is the posterior probability of the hypothesis H_1 "a tank is hiding behind those trees" after the data D is made available, $P(H_1)$ is the prior probability of the hypothesis before the data arrived, $P(D|H_1)$ is the likelihood of the data given hypothesis H_1, and $P(D)$ is the model evidence. If there are a number of mutually exclusive hypotheses H_i, then $P(D) = \sum_i P(D|H_i)P(H_i)$. The mathematical model is encompassed in the function $P(D|H_i)$, which can be constructed to any desired level of complexity. Assumptions about tanks would affect the values of $P(H_i)$.

This approach to probability differs from the classical frequentist view that probabilities represent limiting ratios of frequencies of occurrences of events [95]. The frequentist viewpoint only allows the assignment of probabilities to truly random variables – variables occurring in experiments that can be repeated, such as the rolling of a six-sided die. This is a severe limitation, as we require a system for determining our state of knowledge about a given situation and we are not able to repeat the experiment. For that reason the Bayesian view of probability is preferred.

Often our state of knowledge is not in the form of discrete hypotheses. We may for example want to know the value of a continuous variable such as the distance x of a target from a radar. In this case our state of knowledge is encapsulated in the form of a probability density function $p(x)$. The probability that the distance is in the range $(x, x + \Delta x)$ is $p(x)\Delta x$ in the limit as $\Delta x \to 0$. An example probability density function (PDF) is shown in Figure 2.12, where there are two regions of high probability.

We may also be interested in the probability that the distance is less than a certain value. This is known as the cumulative distribution function (CDF) and is calculated from the integral

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} p(X)\ dX. \tag{2.13}$$

The CDF of the distribution introduced in Figure 2.12 is shown in Figure 2.13. Note that all CDFs are monotonic increasing functions and

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1. \tag{2.14}$$

2.3.2 Mean, Median, Mode, and Confidence Intervals

It is usually inconvenient when conveying information about a variable in human-to-human interaction to use the complete probability density function. The analytic PDF may be mathematically complex or, when using a numerical PDF, many numbers are required to describe the distribution. If that is the case it is common to quote a single statistic to represent the location of the entire distribution. If the statistic is to be representative of the distribution then it should have some sort of central tendency. The three statistics most commonly

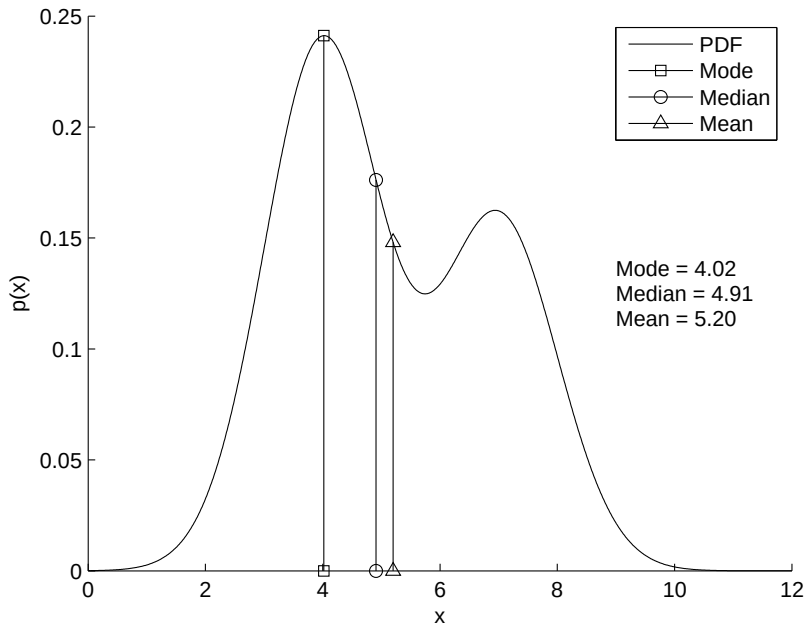


Figure 2.12: An example probability density function.

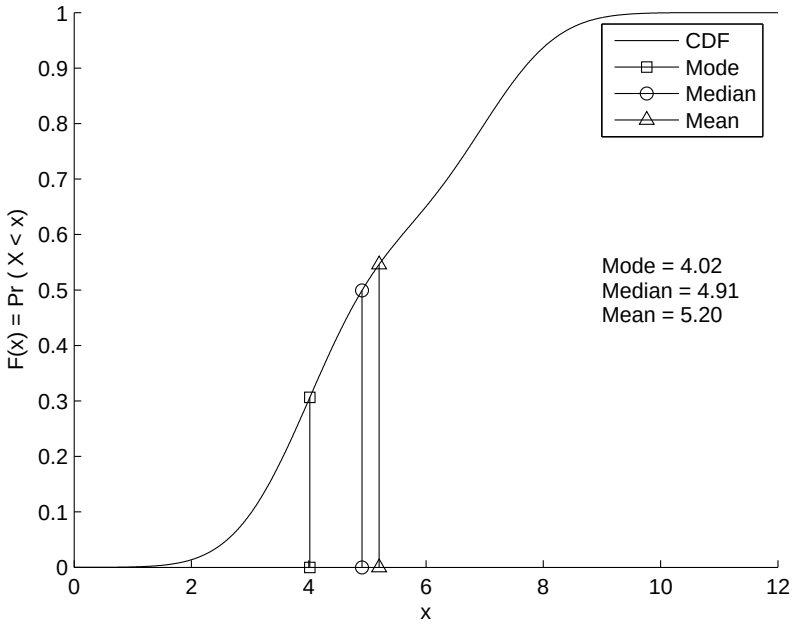


Figure 2.13: The cumulative distribution function of the PDF in Figure 2.12.

used to represent a distribution location are the mean, median and mode. The mean is defined as

$$\bar{x} = \int_{-\infty}^{\infty} xp(x)\ dx. \tag{2.15}$$

The median is the value that divides the PDF into two halves of equal probability – it is the 50th percentile. Thus the median is given by the solution for $\tilde{x}$ in the equation

$$\int_{-\infty}^{\tilde{x}} p(x)\ dx = 0.5. \tag{2.16}$$

Alternatively the median is more simply described as the value at which the CDF is equal to one half:

$$F(\tilde{x}) = 0.5. \tag{2.17}$$

The mode of a distribution $\check{x}$ is the value of x that maximizes value of the PDF. Thus

$$\check{x} = \arg\max_x p(x). \tag{2.18}$$

Note that this equation defines a global mode. If a distribution has local peaks in the density function then it is often described at multimodal even if the local peaks are lower than the global peak. If the distribution has been calculated after application of Bayes' theorem then the global mode is also known as the maximum *a posteriori* (MAP) value.

For a symmetric unimodal distribution the mean, median and mode coincide, when they are all defined. The mean, median and mode values of an example multimodal asymmetric distribution are shown in Figures 2.12 and 2.13 in relation to the PDF and CDF respectively. The figures demonstrate the advantages and disadvantages of the three statistics. The mode describes the region of high probability density well, as seen in Figure 2.12. However, a random variable drawn from the distribution has approximately twice the probability of being above the mode than below. Thus the mode is not necessarily a good indicator of the centre of the distribution. The mean is near the centre of the distribution, and the median is defined as the centre. However, these two statistics are in or near the dip of the density function – a region of lower probability than that surrounding the global mode. Therefore no single statistic captures all the useful information in describing the location of an asymmetric distribution.

In addition to the location of a distribution it is desirable to known its spread of values. The most common descriptor of this spread is the standard deviation σ_x defined as

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 p(x)\ dx. \tag{2.19}$$

The Gaussian distribution is completely defined by its mean and standard deviation and these two statistics are often the only ones quoted regardless of the underlying distribution. However, these two statistics on their own give no indication as to the asymmetry of a distribution. Therefore another often-used set of statistics are the mean and a confidence interval. A confidence interval is the set values that contain $\beta = 100\% - 2\alpha$ of the distribution. The choice of α is arbitrary but values of 5% or 2.5% are commonly used. If each end of the interval is a different distance from the mean this indicates an asymmetrical distribution. The $\beta = 90\%$ confidence interval for the distribution shown in Figure 2.13 is approximately (2.61,8.15). The fact that the top end of this interval is further from the mean than the bottom indicates the distribution is positively skewed. However, even this set of statistics does not reveal the multimodal nature of the distribution or its detailed shape. We could go on quoting higher order central moments, defined as

$$\mu_n = \int_{-\infty}^{\infty} (x - \bar{x})^n p(x)\ dx \tag{2.20}$$

for integer values of $n > 2$, but this rapidly becomes difficult for humans to interpret. Computers are better placed to process these types of statistic but high order moments are sensitive to a few extreme values when analyzing measured data so moments are not necessarily the best representation of a distribution.

The precise shape of the distribution function could be important when making decisions based on that shape. This is especially so in the target recognition application when a decision could result in the launch of a weapon. Ideally, the analytic form of the distribution should be used but in practice this is not always available. When a set of simple statistics such as the mean, median, mode and confidence intervals is an insufficient representation of a distribution, samples can be used to represent the PDF. The sampling representation of a PDF is explained in Section 2.3.7.

2.3.3 Multiple Variables

In our super-resolution and target recognition application we are interested in making inference on more than one variable simultaneously. That requires us to make use of joint probability distributions of the form

$$p(x_1, x_2, ..., x_n) = p(\mathbf{x}), \tag{2.21}$$

where the vector $\mathbf{x}$ contains all the variables of interest. Similarly to the one-dimensional case, the multivariate CDF is defined as

$$P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} ... \int_{-\infty}^{x_n} p(X_1, X_2, ..., X_n)\ dX_1\ dX_2\ ...\ dX_n. \tag{2.22}$$

The multivariate mean is defined by

$$\bar{\mathbf{x}} = \int_{-\infty}^{\infty} \mathbf{x} p(\mathbf{x})\ d\mathbf{x}, \tag{2.23}$$

and the covariance matrix is

$$\Sigma_{\mathbf{x}} = \int_{-\infty}^{\infty} (\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^H p(\mathbf{x})\ d\mathbf{x}. \tag{2.24}$$

The covariance matrix is positive semi-definite ($\mathbf{x}^H \Sigma_{\mathbf{x}} \mathbf{x} \geq 0$ for all $\mathbf{x}$) and Hermitian symmetric ($\Sigma_{\mathbf{x}} = \Sigma_{\mathbf{x}}^H$). For real random variables this means the entire matrix is real-valued and the diagonal elements are non-negative. Off-diagonal elements are positive or negative depending on the relation between variables. If the random variables are complex then the diagonal elements are still real and non-negative but the off-diagonal elements are complex in general.

In the very general case, second-order properties of complex random vectors are not completely described by the covariance matrix alone. If there is a correlation between real and imaginary parts of the vector elements then the complementary covariance defined by

$$\tilde{\Sigma}_{\mathbf{x}} = \int_{-\infty}^{\infty} (\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T p(\mathbf{x})\ d\mathbf{x} \tag{2.25}$$

must be used in addition to the covariance [130]. In the literature this matrix is also referred to as the *relation matrix* [113] or the *pseudo-covariance* [100]. A complex random variable whose complementary covariance is zero is said to be

proper [100] or to have second-order circularity [113]. Proper complex random variables are used in the vast majority of complex statistical analysis and are implicitly assumed in the widely-used standard form of the Normal distribution. Analysis of improper complex random variables is beyond the scope of this thesis and it is assumed throughout that all complex random variables are proper. The interested reader is referred to [100, 113, 130] for further information on complementary covariance matrices.

2.3.4 Marginalization

An important advantage of the Bayesian framework for statistics is the use of marginalization. Marginalization allows us to deal with "nuisance" parameters – parameters that we have no interest in but are required to be present in our model of the system. For example, when estimating a scene $\mathbf{x}$ from radar data blurred by a point spread function with an unknown parameter θ we are usually interested in $\mathbf{x}$ but not θ. Non-Bayesian methods might first estimate θ then $\mathbf{x}$, or even attempt to estimate them simultaneously. However, there may be an interaction between the two variables that is not taken into account. The function that contains all information about the variables is the joint density

$$p(\mathbf{x}, \theta) = p(\mathbf{x}|\theta)p(\theta). \tag{2.26}$$

Marginalization is the process of integrating over the nuisance parameters to leave only variables of interest. This gives

$$p(\mathbf{x}) = \int_{-\infty}^{\infty} p(\mathbf{x}|\theta)p(\theta)\ d\theta. \tag{2.27}$$

The left hand of equation (2.27) now contains only the data of interest, $\mathbf{x}$, and has removed the nuisance parameter θ.

Marginalization is thus able to reduce the dimensionality of a problem. If it is possible to do this analytically there is a large saving on computation time. If an analytic solution is unavailable then marginalization can be carried out numerically as long as it is possible to derive an expression for the marginalization integral. In that case an intractable problem has been converted into a tractable one, possibly at the expense of more computation depending on how the numerical integration is carried out.

2.3.5 Choosing the Prior Density

One problem associated with Bayesian statistics is how to choose the prior density for a variable. If prior knowledge can be represented by an analytic form, then this poses no problems as the appropriate expression can be inserted into Bayes theorem. However, knowledge about a variable could be encoded in other ways, such as minimum and maximum limits, or known moments of the prior density. In many cases there is no specific information available and we would like to use some sort of reference prior or noninformative prior, several of which have been proposed.

It is desirable to seek a reference parameterization of a variable with a uniform prior from which the prior of any other parameterization can be produced by variable transformation. One solution is to choose the reference parameterization such that the likelihood is in *data translated* form. This means that different values of data give rise to the same functional form for the likelihood except for a shift in location. One problem with this method is that it is not always possible to express the likelihood function in this form [82].

Jeffreys' rule states that the prior should be chosen as

$$p(\theta) \propto \sqrt{I(\theta|x)}, \tag{2.28}$$

where $I(\theta|x)$ is the Fisher information for θ defined as

$$I(\theta|x) = -\mathrm{E}\left[\frac{\partial^2 \log p(x|\theta)}{\partial \theta^2}\right], \tag{2.29}$$

and $\mathrm{E}[\ \cdot\ ]$ denotes taking the expectation with respect to the distribution $p(x|\theta)$ while holding θ constant. This rule has the property that the prior is invariant regardless of any transformation that may be performed on θ. Although this invariance property is desirable, there are certain situations where Jeffreys' prior cannot be applied [82].

The above rules in general give different prior distributions apart from cases where the variable is either a location or scale parameter. If the variable is a location parameter then the uniform prior $p(\theta) \propto 1$ can be used. If the variable is a scale parameter then $p(\theta) \propto 1/\theta$ is appropriate. Note that these are improper priors because they do not integrate to unity. This is not a problem as long as the posterior distribution is proper, although all models with an improper prior do not necessarily result in a proper posterior. The noninformative scale

prior $p(\theta) \propto 1/\theta$ should not be used in hierarchical models because the posterior distribution in that case is improper [50].

It is often the case that certain constraints are put on the prior distribution. The maximum entropy principle [69] states that if these constraints are all that is known about a distribution then we should maximize the Shannon entropy

$$H = -\int p(\theta) \log[p(\theta)] \; d\theta \tag{2.30}$$

with respect to $p(\theta)$ to obtain the prior distribution. The distribution $p(\theta)$ is then said to be "maximally noncommittal" with respect to all information except the specific data given [67]. If the constraints are of the form

$$F_k = \int p(\theta) f_k(\theta) \; d\theta, \quad k = 1, 2, ..., m, \tag{2.31}$$

then the prior distribution is given by

$$p(\theta) = Z^{-1} \exp\left[-\sum_{k=1}^{m} \lambda_k f_k(\theta) \right], \tag{2.32}$$

where Z is a normalizing constant and λ_k are Lagrange multipliers determined by the constraints [69]. This principle leads to a uniform distribution when all that is known about a distribution is the maximum and minimum value of a variable. If the distribution is zero mean with a known variance then this results in the Gaussian distribution [158]. The disadvantage of maximum entropy priors is that suitable constraints may not be available and the priors are not necessarily invariant to variable transformations [67].

When attempting to perform analytical calculations it is convenient if the posterior density has the same functional form as the prior density but with different parameters. If this is the case then the prior distribution is known as a *conjugate prior*. A conjugate prior may not exactly match the true prior distribution but could be close enough not to make a significant difference in the ensuing analysis. In fact, many conjugate priors contain the distribution derived using Jeffreys' rule, in which case there is a strong argument for using that particular prior [82]. The disadvantage of conjugate priors is that there are cases where they do not reasonably approximate true prior knowledge. However, with numerical methods it is unnecessary to restrict ourselves to conjugate priors and that is one of the advantages of the numerical MCMC super-resolution algorithm proposed in Chapter 3.

2.3.6 The Curse of Dimensionality

The curse of dimensionality is a phrase used to represent problems that occur when analyzing high-dimensional data [45]. This is especially relevant to radar data analysis, where images of scenes often contain in excess of one million pixels. In the super-resolution problem we are interested in estimating the joint probability density function of pixels values of a target. When representing a probability distribution for one or two dimensions it is possible to define values of the PDF on a discrete grid of points. For the purposes of illustration let us assume it is sufficient to use 100 points in the PDF representation for a single dimension. Thus a two-dimensional PDF would require 10,000 points, which is a small enough number to allow processing with reasonable computer resources. However, as the number of dimensions d increases, the number of points required to represent the PDF grows exponentially as 10^{2d} and rapidly it becomes impossible to store or process this PDF representation. If each dimension is independent then only $100d$ points are required and the problem becomes tractable. However, pixels in a radar image are correlated through action of both the main lobe of the system point spread function and sidelobes, which means the full dependance between pixels must be analyzed.

Another observation is that data is very sparse in high dimensions – it is unlikely that any particular data point is close to another in terms of the Euclidean distance. That is because with more dimensions there is a higher chance that in at least one dimension the two data points will be widely separated, resulting in a large overall distance. This sparseness manifests itself in slightly counter-intuitive ways. For example, a uniform probability distribution in high dimensions has almost all is probability concentrated in a thin shell at the boundaries of the distribution [35]. Also, the ratio of the volume of a unit hyper-sphere to a unit hyper-cube is very small for high dimensions [38]. Neither of these phenomena are apparent in the one-, two- or three-dimensional cases we are used to visualizing.

One way of mitigating the curse of dimensionality is through the use of samples to represent a distribution. This is more efficient than the grid-based method because a single sample represents information in all the dimensions. Therefore, fewer samples than grid points are required. However, care should be taken in viewing samples as a complete solution to the curse of dimensionality

because many samples are still required to represent high-dimensional data and in certain situations other problems still remain [38].

2.3.7 Sampling Representation of a Distribution

In the preceding sections we saw that it would be advantageous to represent a distribution using samples rather than an analytic or grid-based form. Here we examine a two-dimensional Gaussian mixture distribution to highlight differences between the three types of representation.

The analytic form of our example Gaussian mixture PDF is

$$p(\mathbf{x}) = \frac{1}{3}p_1(\mathbf{x}) + \frac{1}{3}p_2(\mathbf{x}) + \frac{1}{3}p_3(\mathbf{x}), \tag{2.33}$$

where

$$p_j(\mathbf{x}) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{j})^T {\boldsymbol{\Sigma}_\mathbf{j}}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_\mathbf{j})\right]}{\left[\det(2\pi\boldsymbol{\Sigma}_\mathbf{j})\right]^{1/2}}, \tag{2.34}$$

and

$$\bar{\mathbf{x}}_\mathbf{1} = [3, 2]^T, \quad \boldsymbol{\Sigma}_\mathbf{1} = \begin{bmatrix} 0.3^2 & 0 \\ 0 & 0.2^2 \end{bmatrix}, \tag{2.35}$$

$$\bar{\mathbf{x}}_\mathbf{2} = [4, 3]^T, \quad \boldsymbol{\Sigma}_\mathbf{2} = \begin{bmatrix} 0.6^2 & 0 \\ 0 & 0.3^2 \end{bmatrix}, \tag{2.36}$$

$$\bar{\mathbf{x}}_\mathbf{3} = [6, 2.5]^T, \quad \boldsymbol{\Sigma}_\mathbf{3} = \begin{bmatrix} 0.4^2 & 0 \\ 0 & 0.1^2 \end{bmatrix}. \tag{2.37}$$

Thus the mixture has three components and is completely described by 21 numbers – six for the mixture means, twelve for the mixture covariances and three for the mixture weights. A graphical representation of the grid-based form for this distribution is shown in Figure 2.14. The darker regions represent areas of high density, and contours of constant density are indicated on the diagram. Figure 2.15 shows the distribution using a sampling representation.

Clearly the analytic form of the distribution is the most accurate as it defines the distribution. However, in practical data analysis it is often impossible to derive a distribution in terms of simple functions like the exponential used in equation (2.34). Visually the two numeric representations provide similar information about the distribution. However, the low-density areas are better

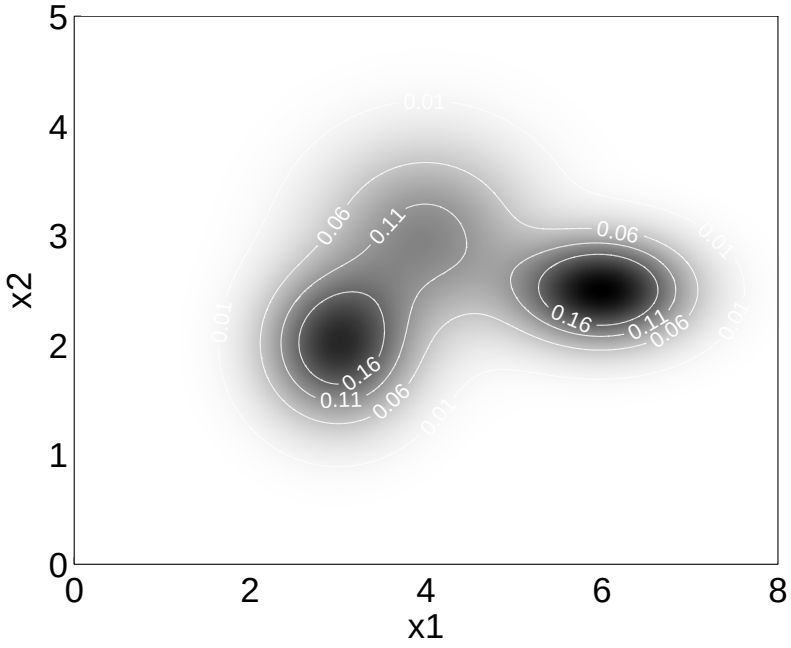


Figure 2.14: An example 2D probability density function.

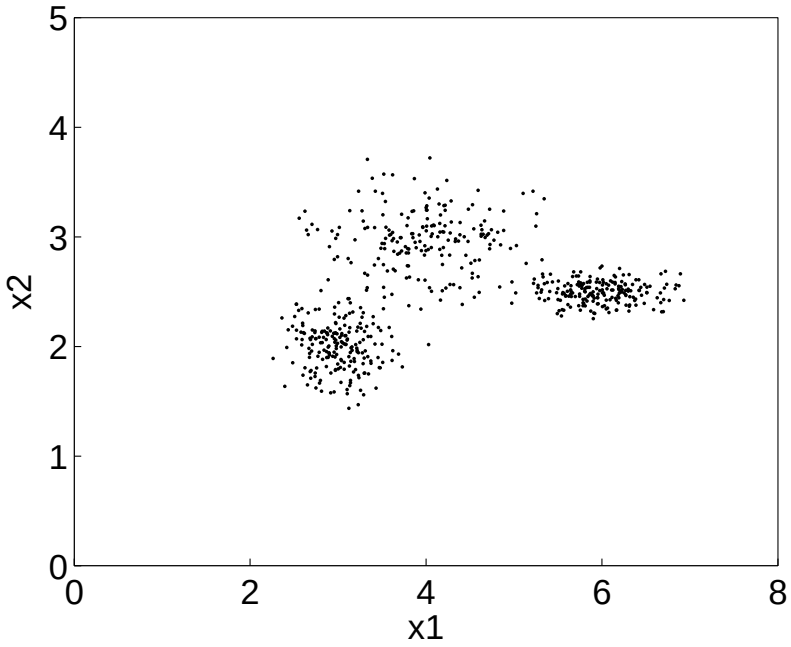


Figure 2.15: Sampling representation of the 2D PDF shown in Figure 2.14.

represented in the grid-based form of Figure 2.14 because the density is calculated over the whole region of interest whereas the sample representation in Figure 2.15 has no samples in areas of low density.

In terms of calculating statistics of interest the sample representation requires less processing power than the grid representation. For example, an estimate of the mean calculated from N sample values is

$$\hat{\bar{\mathbf{x}}}_s = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x_n}, \tag{2.38}$$

and an estimate of the mean calculated from the grid values is

$$\hat{\bar{\mathbf{x}}}_g = \sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2} x_{m_1,m_2}p(x_{m_1,m_2})\Delta x_1\Delta x_2, \tag{2.39}$$

where M_1, M_2 are the number of grid points used in each dimension and Δx_1, Δx_2 are the spacing between grid points. Equation (2.38) is a sum over N terms and equation (2.39) is a sum over M_1M_2 terms. Since $N < M_1M_2$ and the grid form requires multiplications it is quicker to calculate $\hat{\bar{\mathbf{x}}}_s$ than $\hat{\bar{\mathbf{x}}}_g$. In higher dimensions, although the number of samples required is higher, the number of grid points scales exponentially and the calculation saving is greater. In general, the Monte Carlo integration error is inversely proportional to the square root of the number of samples used [93]. Similar expressions can be derived for estimates of the covariance matrix or other statistics of interest.

Another advantage of the sample representation of a probability distribution is the ease of marginalization calculations. To remove a nuisance parameter, that dimension of the variable vector is simply ignored. In the two dimensional example above suppose the nuisance parameter is x_2 and we would like to know the unconditional PDF of x_1. This is represented by the x_1 values of the samples without any further processing. Contrast this to the grid-based representation where the calculation is

$$p(x_{m_1}) = \sum_{m_2=1}^{M_2} p(x_{m_1,m_2})\Delta x_2, \tag{2.40}$$

which clearly requires more processing. In higher dimensions a sum over an extra dimension is introduced for each nuisance parameter that is required to be marginalized, which adds a significant processing burden.

Given the advantages of a sampling representation for probability distributions in terms of avoiding intractable analytic calculations and mitigating against the curse of dimensionality we will use samples when formulating the Bayesian super-resolution algorithm introduced in Section 3.

2.4 Automatic Target Recognition

2.4.1 Introduction

Automatic target recognition (ATR) is the process of obtaining measured data of a scene potentially containing targets of interest, and determining by some method whether targets are present and, if so, what are the types of target. We are concerned here specifically with radar ATR – see [142] for an in-depth treatment of the subject. An overview of the ATR process is shown in Figure 2.16 and is summarized here. Details of the individual steps are described in sections 2.4.2 to 2.4.4.

The first step is to detect whether or not a target is actually present in the data. At this stage it may be determined that no targets of interest are present and it is unnecessary to carry out further processing. If a target is present then its signature must be extracted from the background. This could be in the form of range profiles or images centred on the target, or higher-level features such as target length and width. Once the important target features have been extracted they must be compared with a database of target signatures using a pattern recognition algorithm. Ideally, the algorithm calculates a probability that the target under test belongs to each of the target classes in the database. An "unknown" class may also be included. Non-probabilistic methods are also possible but are not discussed here – see [155] for further information on some of these techniques. If a recognition algorithm is used on its own, the test target is considered to be the database target with the highest probability. If used in combination with other algorithms the vector of probabilities is passed on to a higher-level algorithm that takes probabilities from a variety of sources to make a final decision. A more detailed explanation of these steps is now given individually.

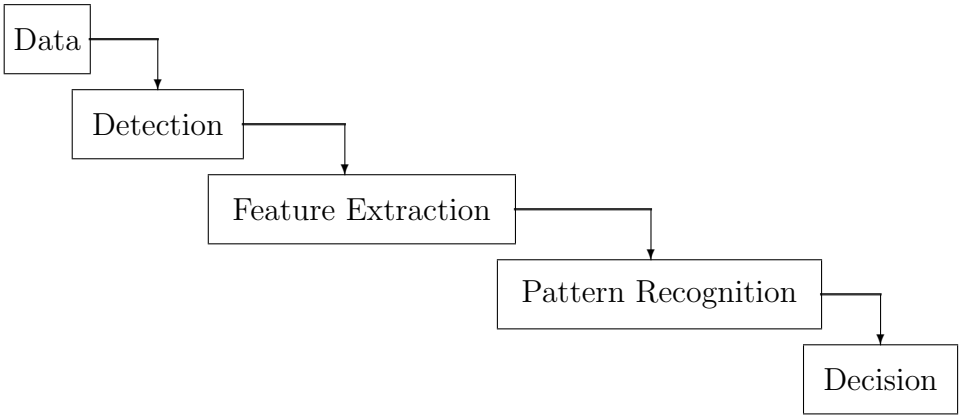


Figure 2.16: Overview of the target recognition process.

2.4.2 Target Detection

Target detection is the best understood element of the whole radar target recognition process as this was the original purpose of radar systems. In radar data the presence of a target is indicated by a response that is stronger than that of the background. If the target is a long way from the radar or exists in a non-cluttered environment, such as that of air targets, then the detection process is limited by thermal noise in the radar receiver. The issue of noise was introduced in Section 2.1.2. The probability density function of the noise magnitude is a Rayleigh function $p(x) = \frac{2x}{\sigma^2}\exp(-\frac{x^2}{\sigma^2})$, where σ^2 is the average noise power [136]. Target detection proceeds by first selecting an acceptably low probability of false alarm (PFA). This is the probability a target is declared when none is present, and is usually in the range 10^{-6} to 10^{-9}. Combination of the PDF and the PFA gives an amplitude threshold of $\sqrt{-2\sigma^2\log(PFA)}$. If the signal exceeds that threshold at a particular position in the the signal or image then a target is declared present at that point.

An illustration of noise-limited target detection using simulated range profile data is shown in Figure 2.17, where a single target is present at range index 120. The resolution of the system is such that the target length is smaller than a resolution cell, so the target appears as a single spike. The rest of the scene contains thermal noise based on Gaussian statistics. The threshold has been set such that the PFA is 10^{-6}. The target exceeds this threshold and is

therefore successfully detected. However, a noise spike at range index 45 has also exceeded the threshold and would be declared as a target – this is a false alarm. In practice, the average of several range profiles is used before target detection, which reduces the probability of false alarm for a given detection threshold.

When measuring certain types of target, the signal or image is said to be cluttered if there is detail in the signal that is related to neither the signal of interest nor thermal noise. With ground targets, for example, returns from the surrounding ground, bushes, trees, buildings or other features of the landscape appear in the signal. When this is the case the probability of a false alarm from one of these features is much higher than that due to thermal noise alone. In this situation the detection threshold should be increased to take into account the distribution of clutter power. A variety of models have been proposed for this with the K distribution being popular.

The K distribution arises from a compound model for fine and coarse scale clutter variations. Under this model, the clutter c is written as $c = x\sqrt{\tau}$, where x is a Gaussian circular complex random variable known as the speckle, and τ is a positive real random variable known as the texture. If τ follows the gamma distribution then $|c|$ follows the K distribution. The speckle is assumed to be spatially uncorrelated but the texture takes correlations between nearby locations into account. In measured data, especially high-resolution data, the distribution of clutter tends to be spikier than that due to Gaussian noise alone, and the K distribution provides a good fit to this data [152].

Constant false alarm rate (CFAR) detectors vary the detection threshold for different parts of the scene to allow locally bright clutter to be rejected while maintaining sensitivity in regions where the clutter power is low [136]. Alternatively, detections due to clutter could be allowed to pass on to the next stage of processing, where more advanced features can be used to reject these initial false alarms [75].

2.4.3 Feature Extraction

Feature extraction is the process of obtaining a useful subset of information about a target from an entire measured signal or image of a scene. This usually first proceeds by determining which part of the signal or image relates to the

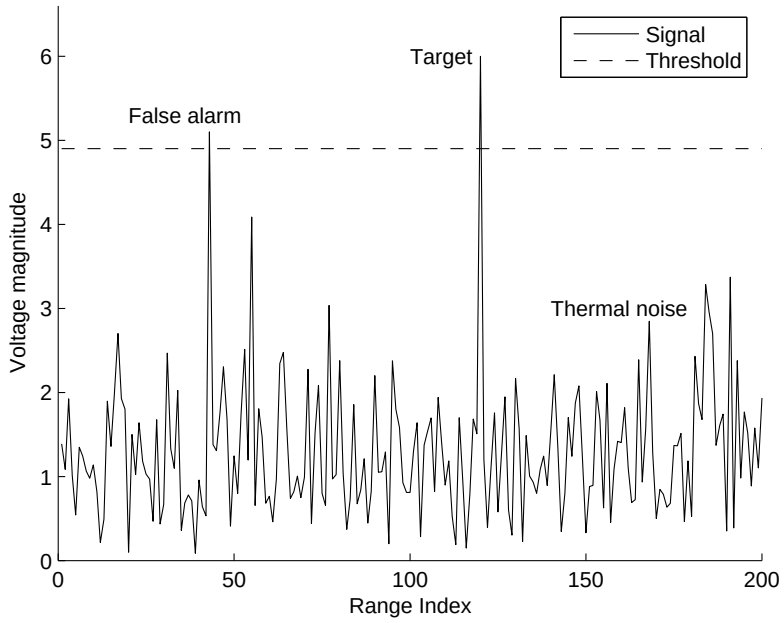


Figure 2.17: Target detection and false alarms.

target and what is part of the background. A variety of algorithms for this purpose are possible. These could place a certain size of window over the detected target position. A more adaptive approach would alter the size and shape of the window according to statistics of the target and background present in the particular scene being processed [23,111]. An example of this adaptive approach is shown in Figure 2.18 for measured millimetre-wave data. Pixels relevant to the target are contained within the black contour, everything outside the contour is related to background only. Windowed target data on its own is sometimes referred to as a target "chip". Once it has been determined which part of the signal relates to the target of interest, features are measured based the target chip.

There are two major classes of features – templates and high-level features. Templates consist of the measured data values for each pixel of the target chip. These could be either the power of each pixel, or for coherent radar, the complex value. Templates are a simple target representation and are easy to interpret for humans because when displayed on a screen they look the same as the measured data. An example template for the target introduced in Figure 2.18 is shown in Figure 2.19a, where the background pixels have been set to zero.

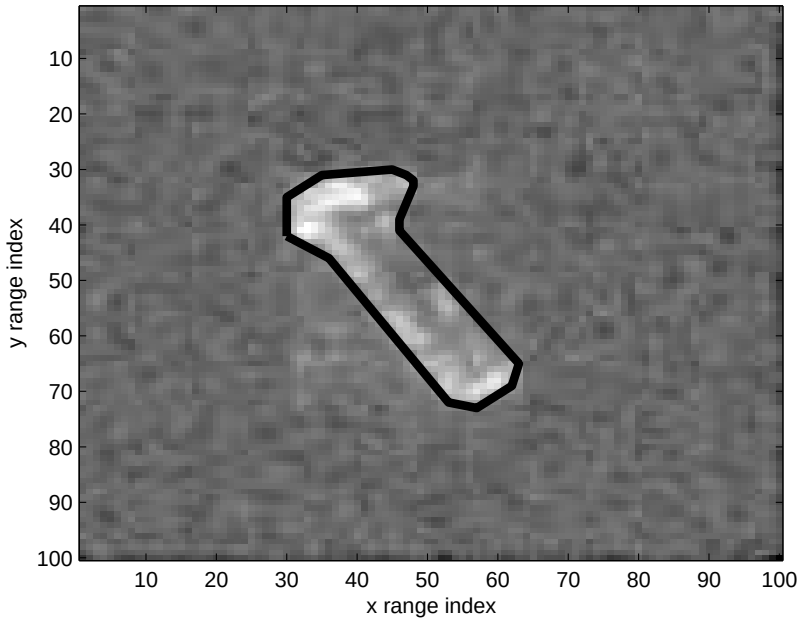
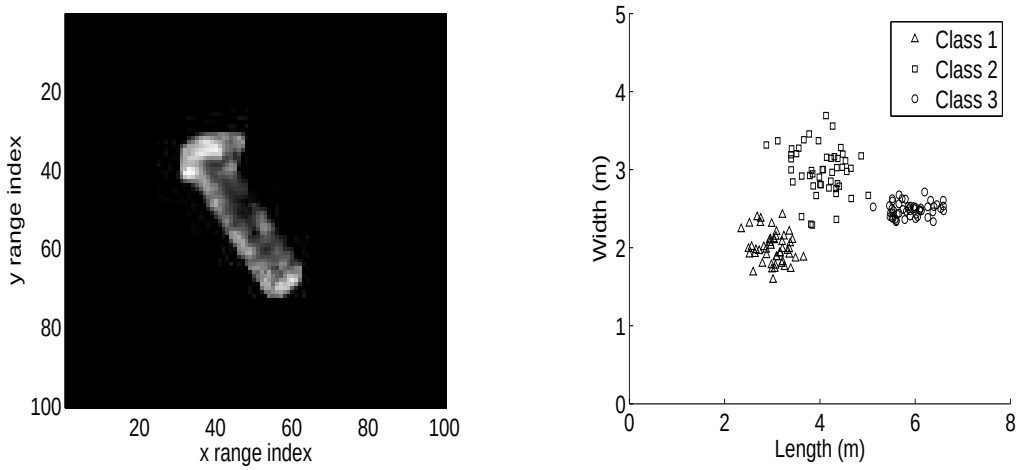


Figure 2.18: Extracting a target from the background.

High-level features provide for a further level of abstraction over templates, but potentially better represent more intrinsic target properties. High-level features measure certain properties of the target image. The image properties could relate to physical characteristics of the target such as length or width, statistical features such as mean or standard deviation of the image pixel values, or more abstract features such as FFT coefficients of the image or other complex combinations of pixel values [92]. An example of high-level features extracted for three classes of target is shown in Figure 2.19b. The two features used in this example are target length and width as measured from simulated radar images. For each class of target, 50 images were used to generate the length and width feature values. With this high-level feature representation it is easier visually to see how well the different classes of targets are separated than having to compare 50 templates per class. However, in terms of computer processing the speed of computation is related to the number of features and it does not matter whether these are derived from templates or higher-level features.

In addition to feature extraction, it may be necessary to perform feature selection. The feature selection process determines which features best allow targets to be differentiated from each other. This step is usually performed in



(a) Target template

(b) Target features

Figure 2.19: Two target representations. (a) Target template for a single target. (b) Two target features for three classes of target.

conjunction with the pattern recognition stage of the target recognition process.

2.4.4 Pattern Recognition

A large number of pattern recognition algorithms have been proposed. Popular algorithms include the Euclidean distance metric, correlation classifiers, linear discriminant analysis, artificial neural networks, support vector machines, and decision trees [155]. Here we outline only the Bayesian classifier and k-nearest-neighbour algorithm as those will be used later in this thesis when comparing super-resolution algorithms in a Bayesian framework. For a recent review of other target recognition algorithms see [27].

The Bayesian classifier is based on the statistics of target features. The probability density function of features is estimated from a database of target images. Mathematically the density function is written as $p(\mathbf{f}|T_i)$, where $\mathbf{f}$ is a vector of features, T_i $(i = 1, ..., n)$ is the target type, and n is the number of targets. There are two main methods of density estimation - parametric and non-parametric. In the parametric method a certain model of target feature variation is assumed, such as a multivariate Gaussian distribution, and density estimation proceeds by estimating parameters of the model. Different parameters are stored for each target in the database. When a new target feature set $\mathbf{f_0}$

is required to be classified, the probability density of the new feature vector is calculated for each target in the database. The probability that the new target T_0 corresponds to a given target in the database is

$$P(T_0 = T_i) = \frac{p(\mathbf{f_0}|T_i)}{\sum_{j=1}^{n} p(\mathbf{f_0}|T_j)}. \tag{2.41}$$

An example of parametric density estimation for the three target classes first introduced in Figure 2.19b is shown in Figure 2.20. Two features are used here: length and width, measured in metres, and are assumed to be drawn from a two-dimensional Gaussian distribution. For each target class the mean feature vector is plotted as a single point in feature space. The covariance for an individual class is represented in the diagram by an ellipse that includes 90% of the assumed data distribution. Figure 2.21 shows how the parametric Bayesian classifier makes a decision between classes when presented with a new feature vector, assuming equal prior probabilities for all three classes. Any target whose length and width combination falls in the black area is classified as belonging to class one, the grey area represents class two, and the white area class three. The original data are also plotted on the graph for comparison. Note that some original data points from class two are misclassified as class one or class three. This misclassification of original data is inevitable for a parametric representation of closely spaced data. However, it is hoped that the classifier maintains good generalization properties *i.e.* when presented with new data the misclassification rate does not significantly increase.

In non-parametric density estimation no particular distribution is assumed and an algorithm such as the k-nearest-neighbour (NN) method is used. In this algorithm the k feature vector samples in the database nearest to the test feature vector are examined as to which target class they belong. If the total number of nearby feature vectors corresponding to target T_i is m_i, the estimated probability that the new target corresponds to a target in the database is given by

$$P(T_0 = T_i) = \frac{m_i}{k}. \tag{2.42}$$

Figure 2.22 demonstrates operation of the k-nearest-neighbour classifier for $k = 1$ and can be compared directly with the parametric Bayesian classifier shown in Figure 2.21. The decision boundaries of the two classifiers are similar in areas near the original data points but are markedly different further away. Note

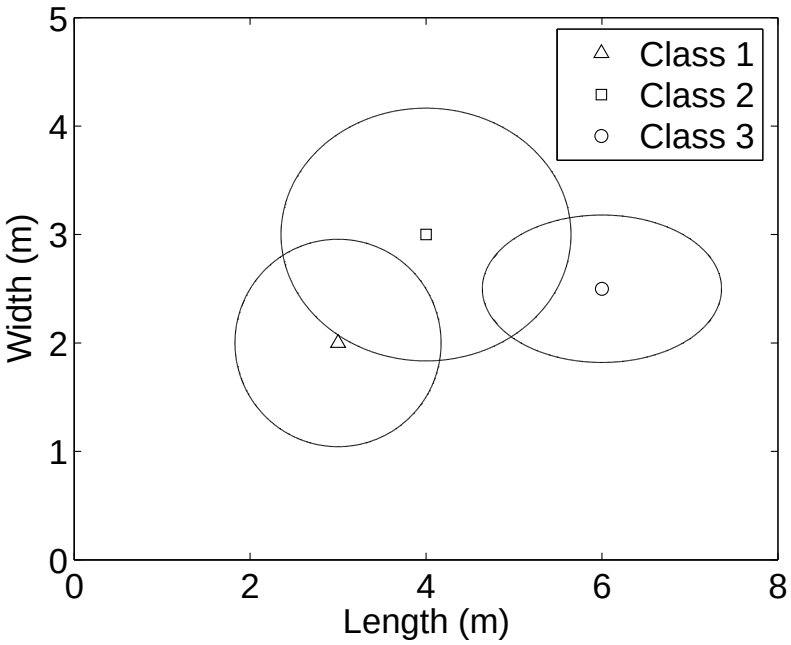


Figure 2.20: PDF of three target classes.

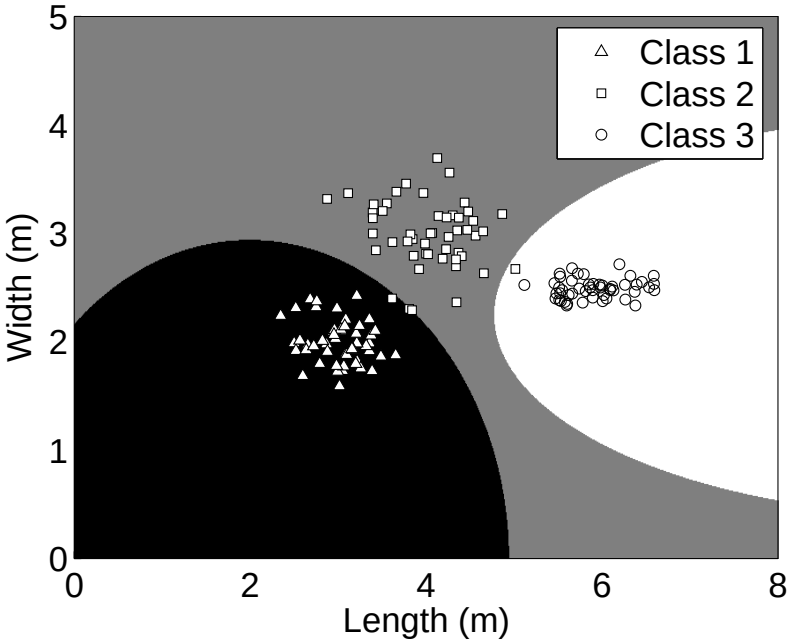


Figure 2.21: Decision areas for a Bayesian classifier based on Gaussian PDFs. Black area: class 1, grey area: class 2, white area: class 3.

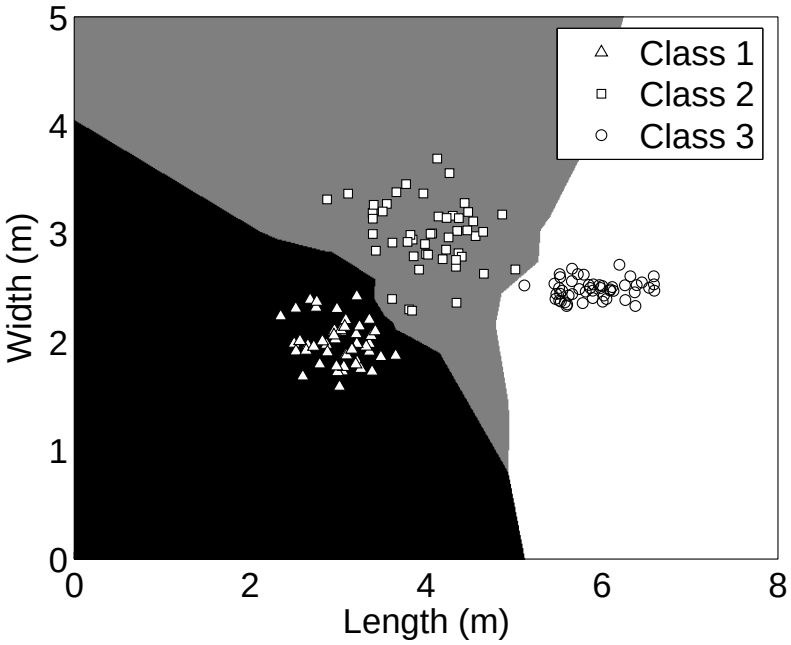


Figure 2.22: Decision areas for a nearest-neighbour classifier. Black area: class 1, grey area: class 2, white area: class 3.

that the nearest-neighbour classifier makes no misclassifications when $k = 1$. Although this may initially seem better than the parametric Bayesian classifier, it does not necessarily have good generalization properties. Generalization could potentially be improved by increasing the value of k although it is suggested in [92] that this is unnecessary. In the simple example given here the three classes are able to be separated into three contiguous areas of the parameter space. However, in the general case where there is overlap between the classes, the decision boundary for the $k = 1$ nearest-neighbour classifier is much more complicated than the Bayesian one. The issue of generalization is covered in detail in [27] and [92].

2.4.5 Results Using Measured Data

In this section the performance of three target recognition algorithms is demonstrated using measured radar data. The data used is the publicly available Moving and Stationary Target Acquisition and Recognition (MSTAR) data set [123]. Images of a variety of vehicles were recorded from angles covering a full 360^o azimuth sweep and at a variety of elevation angles. Here we use a

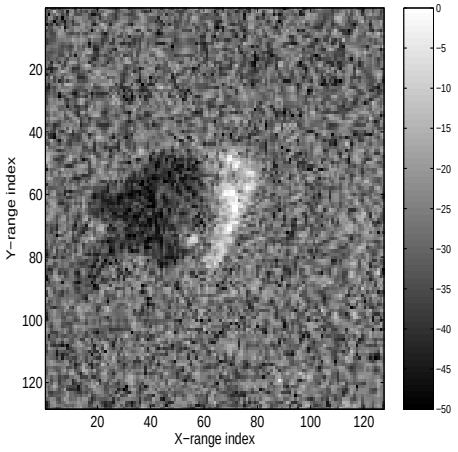
Target name	Target description
T-72	Main battle tank
2S1	Self-propelled howitzer
D7	Military bulldozer
T-62	Main battle tank
ZIL-131	General-purpose army truck

Table 2.1: Vehicles from the MSTAR data set used in these experiments.

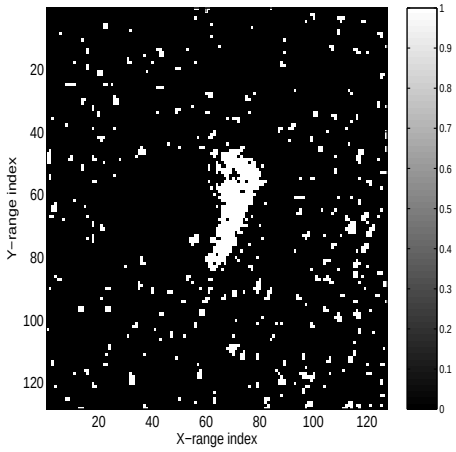
subset of the data consisting of the five vehicles shown in Table 2.1.

In each image from the data set, the target was segmented from the background using a set of morphological operations as demonstrated in Figure 2.23. Figure 2.23a shows an example image of the T-72. The statistics of this image were calculated and a detection threshold was set to detect the 5% brightest pixels. The output of the detection process is shown in Figure 2.23b. Morphological closing was then performed using a 3x3 structuring element followed by morphological opening using a 2x2 structuring element. This removes small isolated groups of pixels and fills holes in the large groups of pixels as shown in Figure 2.23c. The binary group with the brightest pixel in the original image is selected as a binary template for the target, Figure 2.23d. This binary template is then multiplied by the original image to give the segmented target, Figure 2.23e. This target segmentation process is somewhat *ad hoc* but it appears to work well on the target images tested here. It is expected that a more general algorithm would have to be used when testing other data sets.

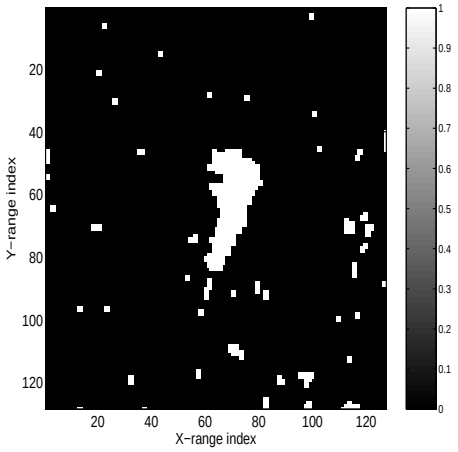
Five features were measured for each target image template. These were RCS, target length, target width, fractal dimension, and weighted rank fill ratio. The RCS is defined by the sum of the squared pixel magnitudes of the target template, which is then converted to dB. The target length and width were based on the second moment of the the binary target template about its minor and major axes, respectively. The fractal dimension and weighted rank fill ratio features were calculated as described by Novak *et al.* [104] but here used the 10% brightest pixels in the target template. The fractal dimension is a measure of the spatial distribution of the brightest pixels and the weighted rank fill ratio is a measure of the proportion of power contained in the brightest pixels.



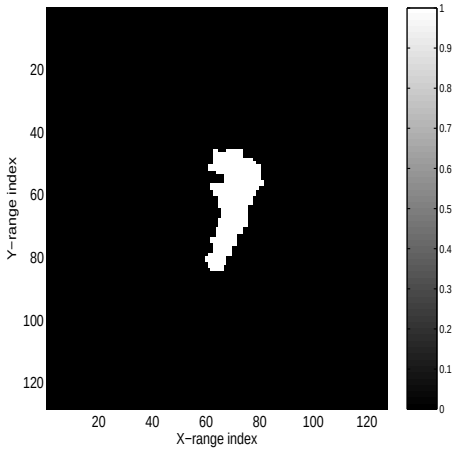
(a) Input image



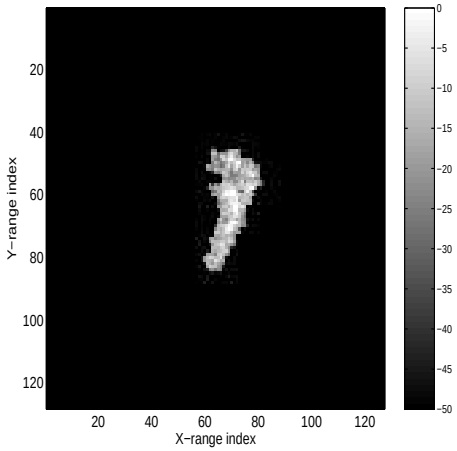
(b) Detection map



(c) Morphological processing output

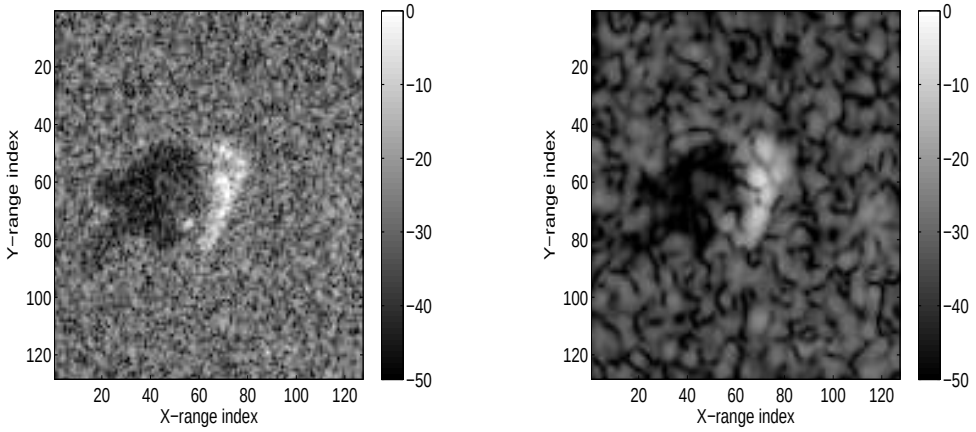


(d) Group with brightest pixel



(e) Target template

Figure 2.23: The process of segmenting the image into target and background.



(a) High-resolution image (b) Low-resolution image

Figure 2.24: High and low-resolution images of a T-72 tank.

The ability to discriminate between the targets was tested using three classifiers. The first classifier was a Bayesian classifier that assumed the features were independent and distributed normally. This was a reasonable assumption based on analysis of the data. The mean and standard deviation of target features was measured for 10^o azimuthal windows of each target and stored in a database. A new target image was classified by selecting the database target that had the highest classification probability for the correct azimuth window. The second classifier was a feature-based nearest-neighbour classifier. The test target was simply classified as the target in the correct azimuth window with the smallest Euclidean distance between test and training feature vectors. The final classifier was a template-based correlation classifier. This measured the peak value of the cross-correlation between the test target and database targets. The test target was classified as the database target in the correct azimuth window with the largest correlation value.

For all classifiers the training database was based on the set of images recorded at an elevation of 17^o. To provide some degree of data independence the test images were selected from the imagery recorded at 15^o. However, as a measure of the maximum possible performance of the classifiers, the 17^o-elevation data used for training was also used during testing.

The nominal resolution of the MSTAR data is 0.3 m and in experiments here this is referred to as high-resolution data. The effect of using low-resolution data

was also tested by creating a new data set from the original data. This was done by convolving the original data with an RCS-preserving point spread function based on the FFT of a Hamming window such that the degraded imagery had a resolution of 1.0 m. The resolution reduction process was applied only to the test imagery. Example high and low-resolution target images are shown in Figure 2.24. The resolution reduction process allows us to assess the impact of having only low-resolution imagery available to test when high-resolution data was used during training. This is the scenario considered in the introduction where different radar sensors are used for the train and test phases.

The average probability of correct classification for each of the classifiers is shown in Table 2.2. The first row corresponds to the case where high-resolution training data was used both to train and test the classifiers. It can be seen that the Bayesian classifier performs best and the template classifier performs worst. The second row corresponds to the more realistic case where the training and test data were independent sets acquired at different elevation angles. It can be seen that there has been a degradation in performance for all classifiers but their performance relative to each other has been maintained. The performance of the Bayesian and nearest-neighbour classifiers is better than that of the template classifier because, in this particular data set, high-level features are a better description of the target than templates. The distribution of target features is very close to Gaussian, which is the distribution assumed by the Bayesian classifier. This results in the Bayesian classifier having a better representation of the feature distribution than the nearest-neighbour classifier, especially in sparse regions of feature space. The third row of Table 2.2 corresponds to the case where the the training and test data resolutions are different. It can be seen that there has been a dramatic reduction in performance for the Bayesian classifier. The nearest-neighbour and template classifiers have also suffered a reduction in performance and the template classifier is now the best under these circumstances.

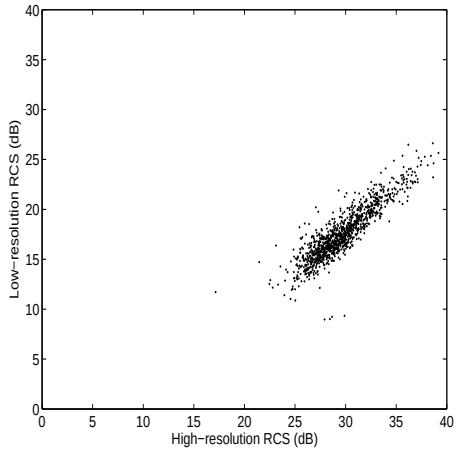
The large reduction in performance for both the Bayesian and nearest-neighbour classifiers when the resolution is changed can be explained by an analysis of the features used for those classifiers. The RCS feature value would be preserved after resolution change for a single point target because the PSF is RCS-preserving. However, the RCS can only be preserved for a single scat-

Test Data			Classifier		
Resolution	Elevation	Features	Bayesian	NN	Template
High (0.3 m)	17^o	1,2,3,4,5	80.1%	66.2%	41.5%
High (0.3 m)	15^o	1,2,3,4,5	59.3%	49.1%	38.0%
Low (1.0 m)	15^o	1,2,3,4,5	28.0%	32.7%	34.9%
High (0.3 m)	15^o	2,3	57.4%	47.0%	38.0%
Low (1.0 m)	15^o	2,3	49.0%	37.6%	34.9%

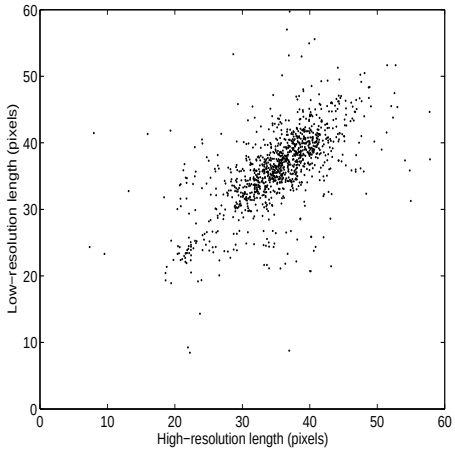
Table 2.2: Probability of correct classification using the MSTAR data set. Percentages based on 1677 test images of targets. Features 1–5 refer to RCS, target length, target width, fractal dimension, and weighted rank fill ratio, respectively. The feature database was generated using high-resolution training data gathered at an elevation of 17^o.

terer. When more than one scatterer is present, the coherent combination of the PSF and scatterers results in constructive interference in some positions and destructive interference in others. The RCS would only be preserved over the whole image if constructive interference were present throughout. However, this is unlikely to happen in practice. Thus the practical effect of resolution degradation is a reduction in the measured RCS. Notwithstanding this reduction, the values of target RCS in low- and high-resolution imagery are fairly well correlated. The degree of correlation is shown in Figure 2.25a, which is a scatter plot of RCS feature values for low- and high-resolution images of all five targets. This implies a scheme to compensate for RCS reduction could be used to recover performance by analyzing the relation between RCS and resolution and taking this into account. This has not been done here because it is a further complication that distracts from the main argument of this thesis.

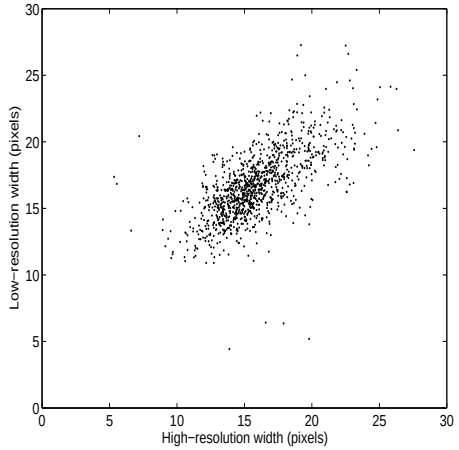
The second problem with resolution degradation is that it causes closely-spaced small bright regions to coalesce into a reduced number of larger ones. This change in distribution of bright regions has an adverse effect on the fractal dimension feature. Fractal dimension feature values in high- and low-resolution imagery are almost completely uncorrelated – see Figure 2.25d. The weighted rank fill ratio feature suffered from similar problems to the RCS feature in that interference between nearby scatterers altered the distribution of power and



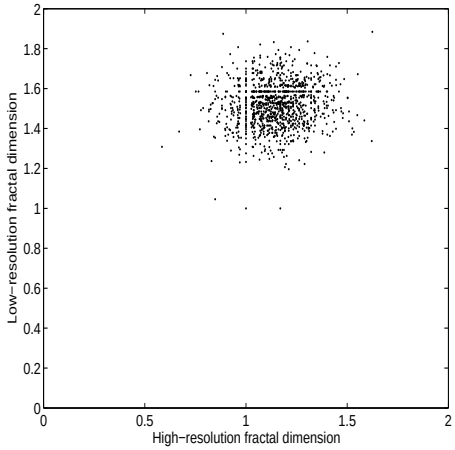
(a) RCS



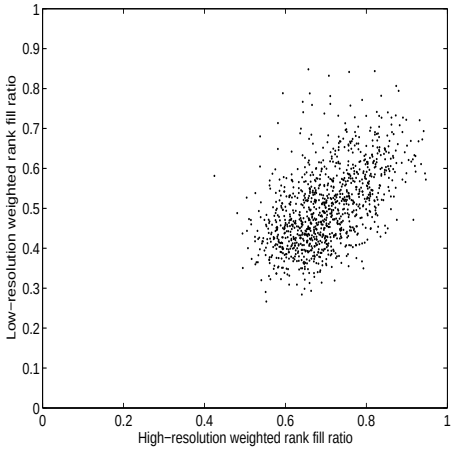
(b) Length



(c) Width



(d) Fractal dimension



(e) Weighted rank fill ratio

Figure 2.25: Comparison of feature values for high and low-resolution images of a T-72 tank.

caused a reduction in value as shown in Figure 2.25e. However, unlike the RCS, the weighted rank fill ratio feature values have a much lower correlation between low- and high-resolution imagery. The lack of strong correlation for both the fractal dimension and weighted rank fill ratio features means there is no processing available to mitigate the effects of resolution change for those features.

Given the problems associated with the RCS, fractal dimension and weighted rank fill ratio features it was decided to re-run the experiment using only the length and width features to see the effect this has on performance. These features were fairly consistent between low- and high-resolution imagery – see Figures 2.25b and 2.25c. The fourth row of Table 2.2 shows classifier performance for the case where high-resolution data was used in both testing and training, and only the length and width features were used. It can be seen that although the performance using two features is lower than that using all five features, for both the Bayesian and nearest-neighbour classifiers, the reduction is only about 2%. Indeed, this result shows that even for same-resolution data, the RCS, fractal dimension and weighted rank fill ratio features do not contribute significantly to overall classification performance. It also should be noted that the template classifier is not affected by the number of features as it does not use them during classification. The results using two features and low-resolution data during the test phase are shown in the fifth row of Table 2.2. Although there is a reduction in performance compared to using high-resolution test data, this is much less for the Bayesian and nearest-neighbour classifiers than when using all five features. This shows that the length and width features are indeed robust to changes in resolution.

When robust features are used, the relative performance of the three classifiers is restored to that in the ideal case – the Bayesian classifier is best followed by the nearest-neighbour and template classifiers. However, in all cases the change in resolution has caused a reduction in classification performance. This motivates the the use of super-resolution techniques to improve the resolution of low-resolution test imagery to match that of high-resolution imagery in the training database.

Chapter 3

Bayesian Super-resolution

3.1 Introduction

3.1.1 Chapter Outline

In this chapter a new Bayesian super-resolution algorithm is described. The algorithm is based on the Metropolis-Hastings (M-H) algorithm, which obtains samples of a distribution using a Markov chain Monte Carlo (MCMC) method. Section 3.1.2 establishes a Bayesian model of the physical processes involved when radio waves are scattered by an illuminated target, based on a radar cross-section model. In Section 3.1.3 a Bayesian model of the receiver physics and signal processing is described. Section 3.1.4 outlines the difference between the inverse cross-section and complex-field problems, which are two target parameter estimation problems. The Bayesian solution to the complex-field problem in analytic and numerical terms is given in sections 3.2.1 and 3.2.2, with simulation results in sections 3.2.3 and 3.2.4. A discussion of the two approaches is given in Section 3.2.5. The inverse cross-section problem is addressed in Section 3.3.1 with results shown in sections 3.3.2 and 3.3.3. A discussion of the algorithm and model is given in Section 3.4 and an approximate Bayesian solution based on the MMSE super-resolution algorithm is presented in Section 3.5.

3.1.2 Scattering Model

The complex scattered field $\mathbf{f}$ of a scene is generated according to a scattering model $p(\mathbf{f}|\boldsymbol{\sigma})$, where $\boldsymbol{\sigma}$ is a multi-valued variable representing the underlying

radar cross section (RCS) of the scene. It is assumed here that within each RCS element σ_k there are many scattering surfaces which each give rise to a reflected radio wave. These waves combine coherently to produce a single wave represented by an element of complex scattered field f_k. If there are a sufficient number of waves in an element then the central limit theorem applies and we may assume Gaussian statistics. This is an alternative formulation of the well-known case 1 Swerling fluctuation models [136]. Other more general models, such as the gamma distribution are possible but these are not discussed here. Following Luttrell [88], for K cross-section elements

$$\begin{aligned} p(\mathbf{f}|\boldsymbol{\sigma}) &= \prod_{k=1}^{K} \frac{\exp(-|f_k|^2/\sigma_k)}{\pi\sigma_k} \\ &= \frac{\exp(-\mathbf{f}^H\boldsymbol{\Sigma}^{-1}\mathbf{f})}{\det(\pi\boldsymbol{\Sigma})}, \end{aligned} \tag{3.1}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix with real, positive elements σ_k. The vector containing these diagonal elements is denoted by $\boldsymbol{\sigma}$. Note that $\mathbf{f}$ is a complex vector, which results in the unusual normalization in equation (3.1). This is explained in detail in [113].

The variable σ_k is a measure of target cross-section per unit area and is therefore dimensionless. In the literature this is referred to as *sigma zero*, or σ^0 and is more usually associated with radar clutter [136]. The electric field corresponding to position k in the scene is the time varying quantity $\mathrm{Re}\{f_k \exp(2\pi i f_c t)\}$, where f_c is the centre frequency of the scattered waveform, t denotes time, and $\mathrm{Re}\{z\}$ denotes taking the real part of a complex number z. Therefore the complex quantity f_k is a phasor representation of the signal. The SI unit of the complex electric field f_k is Vm^{-1}. However, in simulations the complex field has been treated as a dimensionless quantity as the specific units are irrelevant to the results. Other variables related to the complex field, such as the point spread function, complex image, and noise, are also treated as dimensionless quantities in simulations for the same reason.

The above Gaussian model for scattering allows a wide variety of complex-field values at each element position for a given value of RCS. Under this model, each time a target is imaged by a radar, a different realization of complex-field values arises. Thus this model allows for fluctuations in the target signature from image to image. This is sometimes known as multiplicative speckle noise.

In this scenario it can be seen that $\boldsymbol{\sigma}$ is a more fundamental descriptor of a target than $\mathbf{f}$.

3.1.3 Imaging Model

The imaging model used here is the same as that introduced in Section 2.2, whereby the image is considered as the convolution of a point spread function with a high-resolution target representation and the addition of Gaussian noise. However, under a Bayesian model all information must be represented by a probability density function. In this case, the PDF of the image $\mathbf{g}$ is dependent on the target representation $\mathbf{f}$ and is written as $p(\mathbf{g}|\mathbf{f})$. The imaging model is then defined as

$$p(\mathbf{g}|\mathbf{f}) = \frac{\exp\left[-(\mathbf{g}-\mathbf{T}\mathbf{f})^H\mathbf{N}^{-1}(\mathbf{g}-\mathbf{T}\mathbf{f})\right]}{\det(\pi\mathbf{N})}, \tag{3.2}$$

where the terms are defined in Section 2.2. It is important to note here that the numerical Bayesian technique described in the following sections is able to cope with non-linear and non-Gaussian systems by replacing equation (3.2) with the appropriate form.

3.1.4 Inverse Scattered Field Problems and Inverse Cross-Section Problems

In the preceding sections we saw there are two variables that describe the target – the complex scattered field $\mathbf{f}$ and the underlying cross section $\boldsymbol{\sigma}$. The existence of the two variables suggests there are two problems we could attempt to solve, known as the inverse scattered field and inverse cross-section problems [89]. The inverse scattered field problem consists of attempting to recover $\mathbf{f}$ from $\mathbf{g}$. In the Bayesian context this is done by obtaining an estimate of $p(\mathbf{f}|\mathbf{g})$. In super-resolution theory the inverse scattered field problem is almost universally favoured over the inverse cross-section problem, which instead consists of attempting to recover $\boldsymbol{\sigma}$ from $\mathbf{g}$ [89]. In the Bayesian context this done by obtaining an estimate of $p(\boldsymbol{\sigma}|\mathbf{g})$. In this case $\boldsymbol{\sigma}$ could be considered to be a hyper-parameter of a hierarchical model [50]. It has been argued in [89] that since $\boldsymbol{\sigma}$ is a more fundamental descriptor of a target than $\mathbf{f}$ it should be the

inverse cross-section problem that is studied. In this thesis we examine both problems from a Bayesian viewpoint.

3.2 Complex-Field Recovery

3.2.1 Analytic Bayesian Solution

With respect to the inverse scattered field problem, the Bayesian approach to super-resolution is a probabilistic way of modeling uncertainty in the high-resolution target representation $\mathbf{f}$. It is possible that different combinations of $\mathbf{f}$ and noise could give rise to the same image due to the interaction of several elements of $\mathbf{f}$ in a resolution width. This uncertainty is described by the probability density $p(\mathbf{f}|\mathbf{g})$ of the target representation, conditional on the image under consideration. Bayes' theorem gives the density as

$$p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f})p(\mathbf{f})}{p(\mathbf{g})}. \tag{3.3}$$

If $p(\mathbf{f})$ is a zero-mean multivariate Gaussian defined by

$$p(\mathbf{f}) = \frac{\exp(-\mathbf{f}^H\mathbf{W}^{-1}\mathbf{f})}{\det(\pi\mathbf{W})}, \tag{3.4}$$

then under this model it is possible to calculate the density $p(\mathbf{f}|\mathbf{g})$ analytically. The solution is also a multivariate Gaussian distribution and is given by [88] as

$$p(\mathbf{f}|\mathbf{g}) = \frac{\exp\left[-(\mathbf{f}-\bar{\mathbf{f}})^H\mathbf{C}^{-1}(\mathbf{f}-\bar{\mathbf{f}})\right]}{\det(\pi\mathbf{C})}, \tag{3.5}$$

where

$$\bar{\mathbf{f}} = \mathbf{C}\mathbf{T}^H\mathbf{N}^{-1}\mathbf{g} \tag{3.6}$$

is the mean of the posterior distribution and

$$\mathbf{C}^{-1} = \mathbf{W}^{-1} + \mathbf{T}^H\mathbf{N}^{-1}\mathbf{T} \tag{3.7}$$

is the inverse covariance matrix. With algebraic manipulation the mean can be re-written as $\bar{\mathbf{f}} = \mathbf{W}\mathbf{T}^H(\mathbf{T}\mathbf{W}\mathbf{T}^H + \mathbf{N})^{-1}\mathbf{g}$, which is in fact the same as the MMSE solution given in Section 2.2. However, the Bayesian solution provides more information than MMSE in the form of the covariance matrix, which determines uncertainty in the recovered scene.

3.2.2 Monte Carlo Algorithm

In general, for a non-Gaussian prior distribution or more complicated models, it will not be possible to derive a simple analytic solution similar to that of the previous section. Calculation of the normalization constant $p(\mathbf{g})$ for the posterior distribution in equation (3.3) is usually not tractable. For most physical and processing models, statistics of interest such as the mean and covariance will not be available analytically either. In such cases, rather than making simplifications to allow analytic inference on the posterior distribution, a full Bayesian approach to the problem is maintained by drawing samples from the posterior distribution $p(\mathbf{f}|\mathbf{g})$. All inferences can then be made through consideration of these samples. In most circumstances it is not possible to sample directly from the posterior distribution, therefore a Markov chain Monte Carlo (MCMC) algorithm is used. The particular algorithm used here is the Metropolis-Hastings (M-H) algorithm [24]. Note that for the Gaussian imaging model it is unnecessary to use this sampling approach as the analytic solution has already been found – see equation (3.5). However, the analytic solution allows a comparison with the MCMC solution and aids understanding of the M-H algorithm output. In subsequent sections a more advanced scattering and imaging model will be used where an analytic solution is not available.

The Metropolis-Hastings algorithm is an iterative method for generating samples of a probability distribution. For the case considered here the samples represent the probability density $p(\mathbf{f}|\mathbf{g})$. One advantage of the M-H algorithm is that it is necessary only to know the shape of the distribution $p(\mathbf{f}|\mathbf{g})$ – there is no need to calculate the normalizing factor $p(\mathbf{g})$. The likelihood of the image $p(\mathbf{g}|\mathbf{f})$ was given by equation (3.2). With choice of a suitable priors for $p(\mathbf{f})$ the quantity of interest is then

$$\pi(\mathbf{f}|\mathbf{g}) = p(\mathbf{g}|\mathbf{f})p(\mathbf{f}). \tag{3.8}$$

Each element of the vector $\mathbf{f}$ is considered to be a separate variable, although element values may be correlated. At each iteration of the algorithm it is possible to update one variable at a time or all variables in one go. Here we update each element of $\mathbf{f}$ in turn. During the update at the ith iteration, a proposed new sample for a single element of $\mathbf{f}$ is generated from a proposal distribution $q(\mathbf{f^{i+1}}|\mathbf{f^{i}})$. The proposal distribution may take a wide variety of

forms, each having their advantages and disadvantages as discussed in [24]. The proposed sample is accepted with a probability $\alpha(\mathbf{f^i}, \mathbf{f^{i+1}})$, where

$$\alpha(\mathbf{f^i}, \mathbf{f^{i+1}}) = \min\left[\frac{\pi(\mathbf{f^{i+1}}|\mathbf{g})q(\mathbf{f^i}|\mathbf{f^{i+1}})}{\pi(\mathbf{f^i}|\mathbf{g})q(\mathbf{f^{i+1}}|\mathbf{f^i})}, 1\right]. \tag{3.9}$$

For ease of notation in this equation we have omitted the dependence on field element number so that while updating the jth element of $\mathbf{f}$ we have in fact

$$\mathbf{f^i} = \left[f_1^{i+1}, f_2^{i+1}, ..., f_{j-1}^{i+1}, f_j^i, f_{j+1}^i, ..., f_m^i\right]^T \tag{3.10}$$

and

$$\mathbf{f^{i+1}} = \left[f_1^{i+1}, f_2^{i+1}, ..., f_{j-1}^{i+1}, f_j^{i+1}, f_{j+1}^i, ..., f_m^i\right]^T. \tag{3.11}$$

In other words, at each step a new sample is generated; if it is more likely (including the effect of the proposal distribution) than the current sample it is always accepted but less likely samples are also accepted with a certain probability. This avoids the problem of getting trapped in local maxima – the purpose of the algorithm is to explore the entire distribution rather than find a single optimal value. If the proposed sample is rejected then the current sample is used in the next iteration step.

Initial samples generated depend on the starting position and must be discarded if they are not reasonable values of the distribution. These samples form what is known as the burn-in period. The remaining samples are distributed from $p(\mathbf{f}|\mathbf{g})$ as required.

The M-H algorithm tends to produce time-correlated samples. This happens because of two reasons. The first is that only a certain percentage of new samples are accepted at each iteration stage. If a new sample is not accepted then the old one is retained, meaning that it is possible to have several samples in a row of exactly the same value. The second reason is that often a random walk proposal distribution is used – new samples are proposed as a step from the current sample according to a distribution such as the uniform or Gaussian distributions. If these steps are small then successive samples will have similar values. Note that it is not possible to increase the step size indefinitely to decorrelate the samples because proposed samples far from the current one often have a low probability of acceptance. Thus the random walk step size must be chosen as a compromise between a low acceptance rate and a high correlation between successive incrementally different sample values.

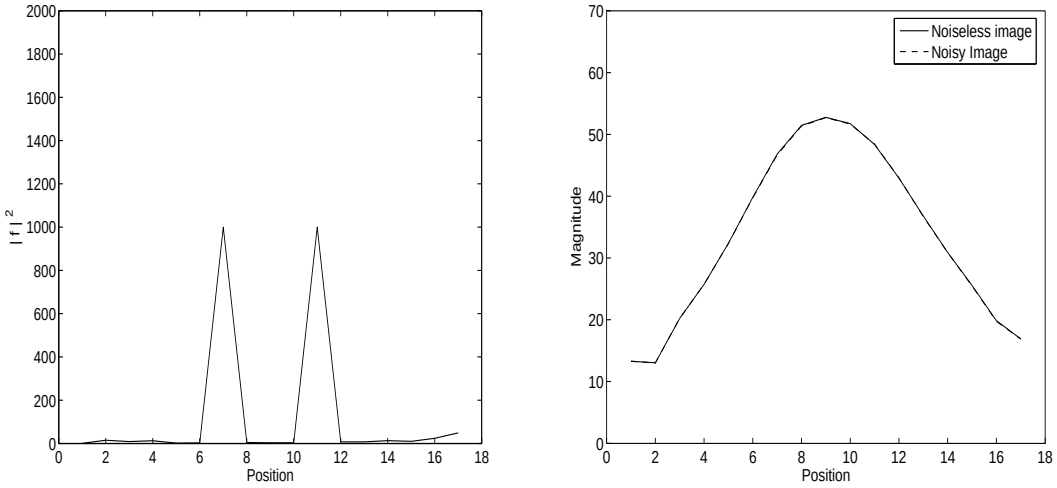
The existence of correlated samples in the M-H algorithm output need not necessarily be a problem. Indeed, correlation is irrelevant when calculating statistics such as the sample mean and variance. However, if samples are highly correlated then a large number of samples are required to be collected to ensure the samples accurately represent the whole distribution space. A large number of samples increases storage requirements and places an extra burden on post-processing of the M-H algorithm output. Therefore a commonly used technique is to save only a subset of the samples. With this technique the first sample is saved and a certain number of subsequent samples are calculated but not saved before saving another sample. The number of samples between those saved is known as the "de-correlation gap". The process is repeated throughout the algorithm run-time, alternately saving one sample and calculating several others without being saved. The result is a set of samples that have a lower correlation than those produced by the basic algorithm. The de-correlation gap is set as a compromise between longer run-times required to generate a given number of samples and the need to adequately explore the distribution space with a given number of samples.

Note that it has been assumed here the variance of the thermal noise is known. If desired, it is possible to include the noise variance as an additional unknown parameter to be estimated [27].

3.2.3 Results for a High Signal-to-Noise Ratio

A demonstration of the Metropolis-Hastings algorithm for field recovery is now given. A simulated scene of two point targets embedded in weak uncorrelated clutter at a signal-to-clutter ratio of 20dB is shown in Figure 3.1a. The two targets are in phase with a phase of 45^o with respect to an arbitrary reference. The clutter has random phase. A point spread function was applied to the scene and thermal noise added at an SNR of 50 dB. The resulting image is shown in Figure 3.1b. The point spread function used was the modified sinc function, previously used in [88], with a θ value of 0.1. This function will be discussed in more detail in Chapter 4 and is also analyzed in Appendix C.

The noisy image, point spread function, and SNR were used as inputs to the M-H algorithm. The detailed processing parameters used in the algorithm are shown in Table 3.1. The burn-in period has been set to zero for this example



(a) True Scene

(b) Low-resolution image

Figure 3.1: Field recovery set-up with arbitrary position units. (a) Power of the true-scene complex field. (b) Low-resolution noisy image with an SNR of 50 dB. Note that at this high SNR the noiseless and noisy images are visually almost indistinguishable.

to show the effect of starting position on the output samples. The random walk step size was chosen to given a sample acceptance rate of approximately 25%, which is recommended in [24] for high-dimensional problems. The prior standard deviation was set at a large value to simulate a lack of detailed prior knowledge about the scene.

The output of the M-H algorithm is a series of samples that represent the probability distribution of the complex recovered field. There are 17 range index positions in this example and each position holds a complex number. Therefore there are a total of 34 variables that describe the distribution. Since it is impossible to display all information about all 34 variables simultaneously we examine either one variable at a time or small groups of variables.

Figure 3.2a shows the sample time series of the real part of range index 7 as generated by the M-H algorithm. This position relates to the left hand scatterer in Figure 3.1a. Under ideal conditions the value we would expect for this variable is $\sqrt{1000/2} \approx 22.4$ field units. We see from the figure that the samples rapidly increase from zero to the expected value and thereafter settle into a random time-correlated pattern with the mean approximately equal to

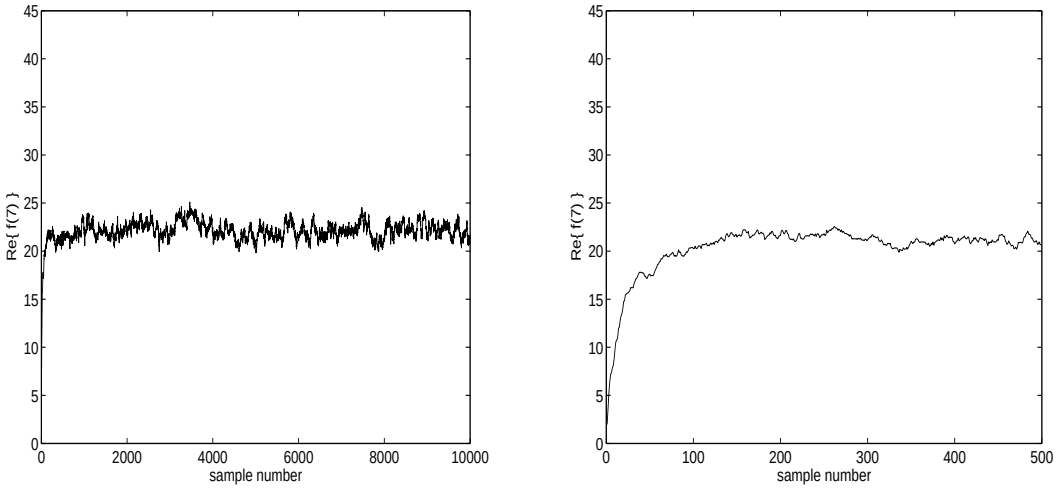
Number of saved samples	10000
Burn-in period	0
De-correlation gap	10
Proposal distribution	Uniform random walk
Random walk step size	0.5
Prior distribution p(f)	Independent Gaussian
Prior mean	0
Prior standard deviation	$\sqrt{1000}$
Sample start value	0

Table 3.1: Table of parameters used by the M-H algorithm for field recovery.

22.4. A zoomed-in graph of the first 500 samples is shown in Figure 3.2b. It can be seen from this graph that the required burn-in time is approximately 100 de-correlated samples. In those first 100 samples the values are all below 20 field units and are due to the algorithm having been initialized with values of zero. The remaining samples after the first 100 could reasonably be considered to be part of the final distribution.

The times series graph for the imaginary part of range index 7 is similar to that of the real part. A 2D plot displaying the sample-to-sample trajectory of the real and imaginary parts of the samples for range index 7 is shown in Figure 3.3a, an Argand diagram. In this plot the burn-in samples are apparent in the random walk trajectory from zero to the area of high probability density around $22.4 + 22.4i$. The majority of samples are in the area of high density as expected by design of the algorithm. The area of high density is circular in shape due to the assumption of second-order circularity as explained in Section 2.3.

An Argand diagram for range indices 10 and 11 is shown in Figure 3.3b. In this diagram the sample values are plotted instead of the sample-to-sample trajectory. Range index 11 corresponds to the right hand scatterer in Figure 3.1a. The sample values of this scatterer roughly follow the same distribution as those of the scatterer at position 7. This is because in the true scene the two scatterers have exactly the same magnitude and phase. Any difference in the sample distribution is due to noise added during the generation of the



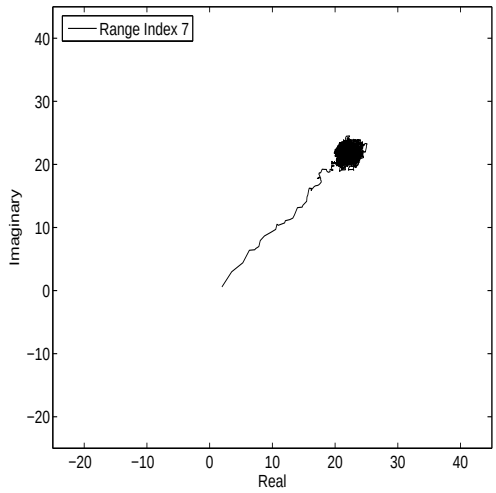
(a) Whole time series (b) Start of time series

Figure 3.2: Time series of samples generated by the M-H algorithm using a decorrelation gap of 10 samples and an SNR of 50 dB. (a) Real part time series for range index 7 of the true scene. (b) Real part time series of range index 7, zoomed-in to start of series.

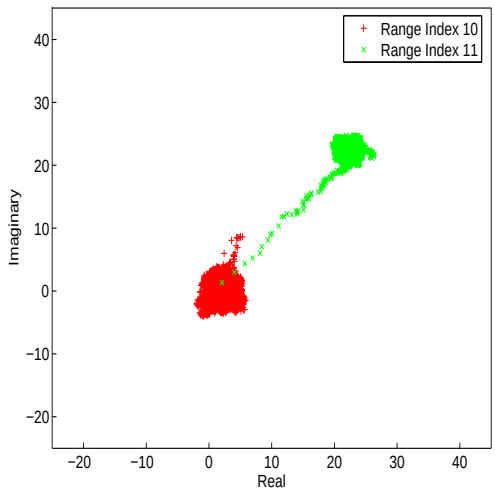
image. Range index 10 corresponds to a low-magnitude clutter sample. This is apparent in Figure 3.3b where the samples are clustered near to zero. Note that no burn-in period was required for range index 10 as all samples could be considered to be part of the final distribution. However, when discarding burn-in samples it is usual for all variables to have the same number of samples discarded. As a result, the burn-in period is determined by the variables that take longest to reach high density areas – range indices 7 and 11 in this case.

The final Argand diagrams for this example are shown in Figures 3.3c and 3.3d and relate to range indices 8 and 9 respectively. The sample distribution for position 8 is similar to that of position 10 in terms of both the mean and variance. The distribution for position 9 has a larger variance than all other positions. – the algorithm has determined that large changes to the complex-field value at this position have a small effect on the final image.

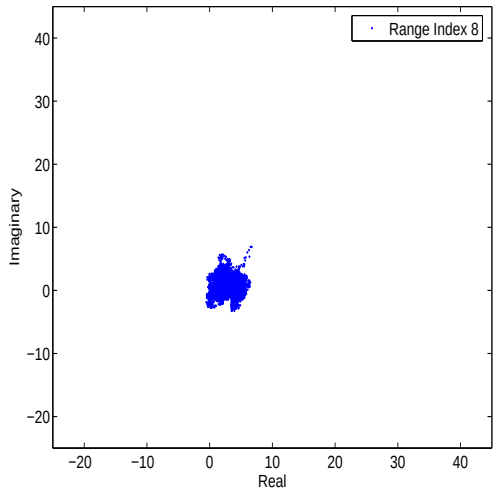
It would be possible to generate either time series graphs or Argand diagrams relating to the other positions in the scene. However, since those all relate to clutter, and their distributions are similar to those for positions 8 and 10, no further insight is gained by examining those graphs. However, it is of interest to



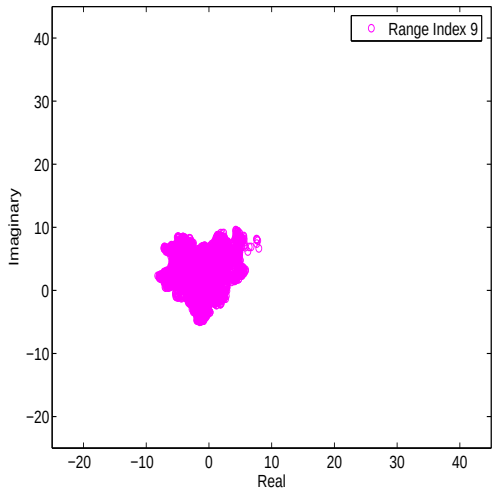
(a) Range index 7



(b) Range indices 10 and 11



(c) Range index 8



(d) Range index 9

Figure 3.3: Argand diagram of sample values for various range indices as output by the M-H algorithm with an input SNR of 50 dB.

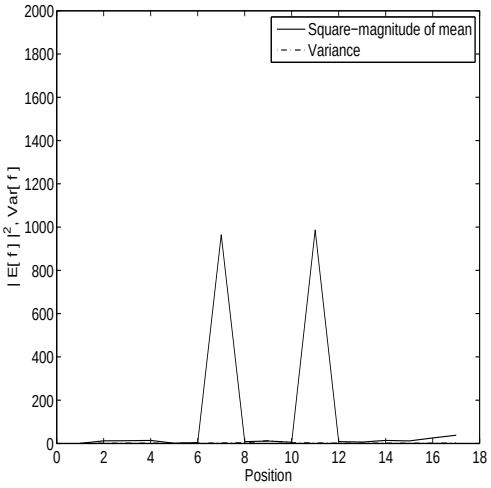


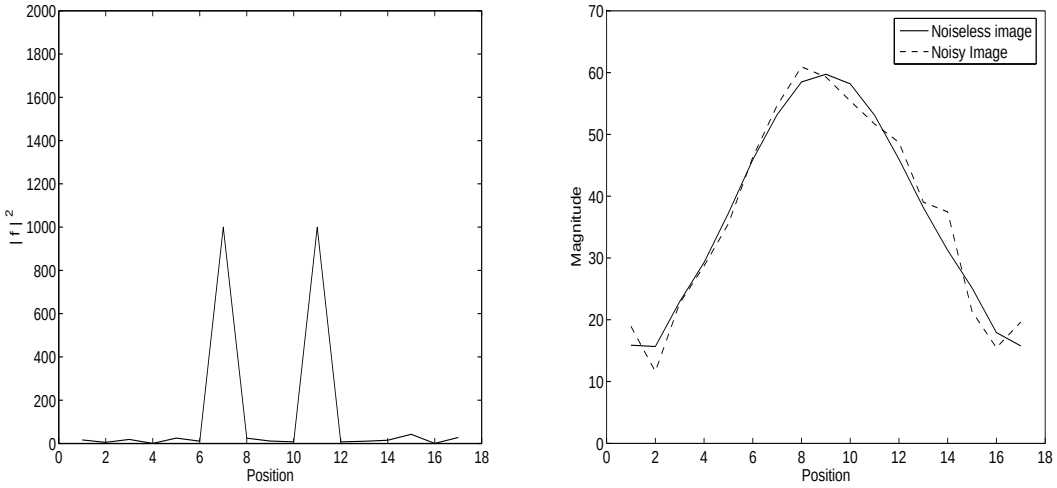
Figure 3.4: Recovered field power range profile, input SNR 50 dB. The variance at each position is low enough that it is indistinguishable from zero in the graph.

look at a range profile representative of the samples. Figure 3.4 shows the mean power range profile and the variance profile. The mean power profile is formed by calculating the mean of the complex sample values for each range index and using the squared magnitude of this value. The variance is the average squared distance of the samples from the mean. Ideally the mean power range profile should be similar to the true scene as shown in Figure 3.1a. For the high SNR of 50 dB used in this example the profile is indeed almost identical to the true scene. The variance has a maximum of about 10 units at position 9 and is visually hard to distinguish from zero in Figure 3.4.

3.2.4 Results for a Low Signal-to-Noise Ratio

We now analyze performance of the M-H algorithm with the same scene but with a lower SNR of 20 dB to see the effect of additional noise. All the parameters used in this example are the same as those used in the 50 dB example except for the noise level and the proposal distribution step length, which was set to 3 field units in order to achieve a 25% acceptance rate at this SNR. The simulated scene is shown in Figure 3.5a and the low-resolution noisy image in Figure 3.5b.

Figure 3.6a shows the sample time series of the real part of range index 7 as generated by the M-H algorithm. This position relates to the left hand scatterer in Figure 3.5a. As before, the value we would expect for this variable



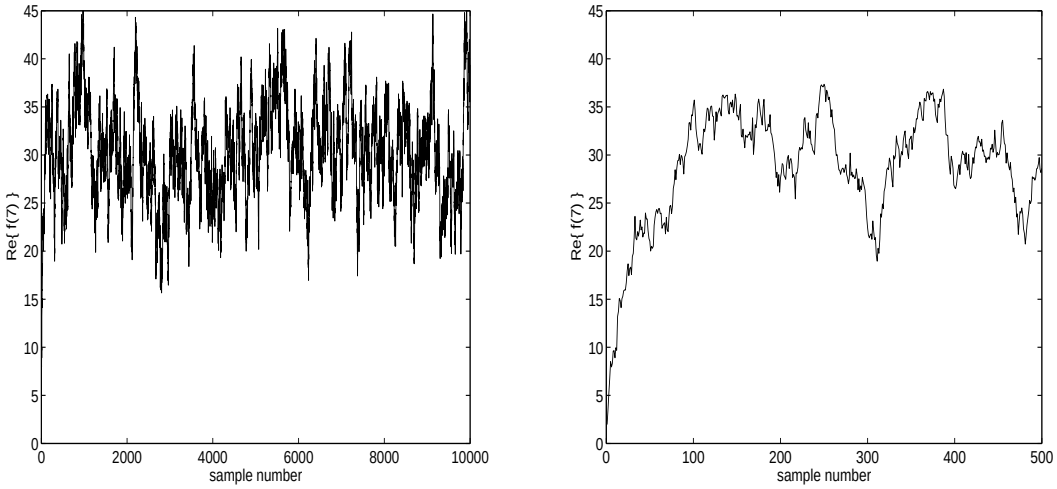
(a) True Scene

(b) Low-resolution image

Figure 3.5: Field recovery set-up with arbitrary position units. (a) Power of the true-scene complex field. (b) Low-resolution noisy image with an SNR of 20 dB.

is $\sqrt{1000/2} \approx 22.4$ field units. Again, we see from the figure that the samples rapidly move away from zero and thereafter settle into a random time-correlated pattern. A zoomed-in graph of the first 500 samples is shown in Figure 3.6b. It is apparent from both these figures that although the mean is similar to that in the high SNR example, the variance is much larger. This should be expected because a larger amount of noise results in a greater uncertainty in parameter values. The Argand diagram of the sample trajectories for range index 7 is shown in Figure 3.7a, where the increased variance is apparent for both the real and imaginary parts.

An Argand diagram for range indices 10 and 11 is shown in Figure 3.7b. Range index 11 corresponds to the right hand scatterer in Figure 3.5a. The imaginary sample values of this scatterer roughly follow the same distribution as those of the scatterer at position 7. However, the real parts of the samples at position 11 are concentrated around a lower value of approximately 5 field units instead of 22.4. The fact that the mean of the samples is not very close to the true value is due to the high levels of noise in the image. The true value is at the edge of the region of high density represented by the samples. A closer examination of the range index 7 distribution in Figure 3.7a also shows that



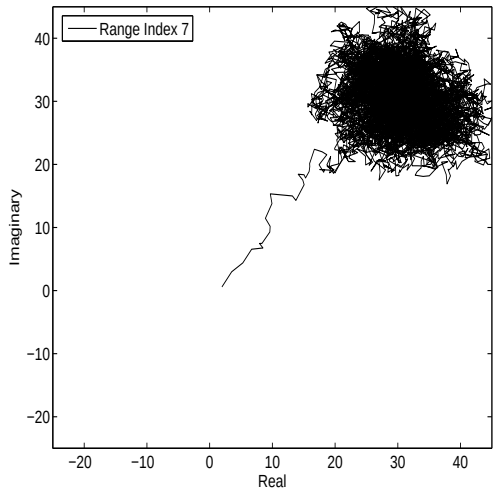
(a) Whole time series (b) Start of time series

Figure 3.6: Time series of samples generated by the M-H algorithm using a decorrelation gap of 10 samples and an SNR of 20 dB. (a) Real part time series for range index 7 of the true scene. (b) Real part time series of range index 7, zoomed-in to start of series.

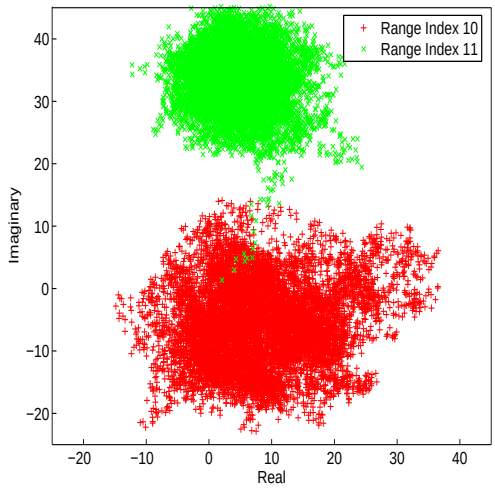
both the real and imaginary parts are slightly over-estimated. Range index 10 corresponds to a low-magnitude clutter sample. This is apparent in Figure 3.7b where the samples are clustered near to zero. However, as with indices 7 and 11 there is a certain amount of bias in the location of the sample mean.

The final Argand diagrams for this example are shown in Figures 3.7c and 3.7d and relate to range indices 8 and 9 respectively. The sample distribution for position 8 is similar to that of position 10 but the bias is in a different direction. As in the high SNR case, the distribution for position 9 has a large variance, which is even larger here due to the additional noise. Also note that coverage of the high-density complex area for position 9 is somewhat patchy. This indicates an insufficient number of samples have been collected to represent the probability distribution. In practice, this could be detected and corrected for by collecting samples until a more even coverage is available.

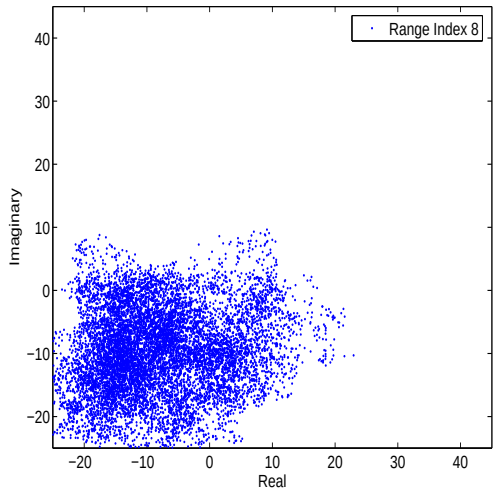
Figure 3.8 shows the mean power profile and variance. It can be seen that, as opposed to the high SNR case, the profile is somewhat different from the true scene shown in Figure 3.5a. The existence of the two targets at positions 7 and 11 have been correctly identified and their power is in the correct order of



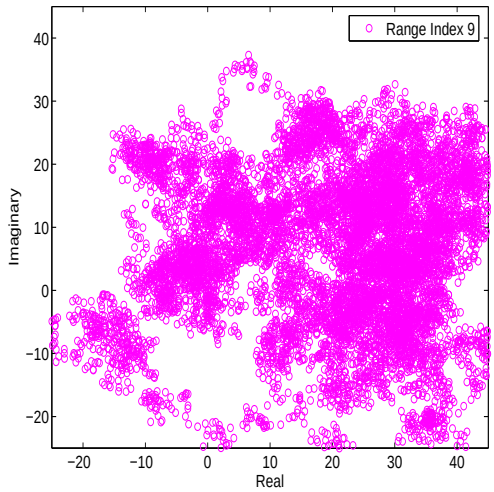
(a) Range index 7



(b) Range indices 10 and 11



(c) Range index 8



(d) Range index 9

Figure 3.7: Argand diagram for various range indices of sample values as output by the M-H algorithm with an input SNR of 20 dB.

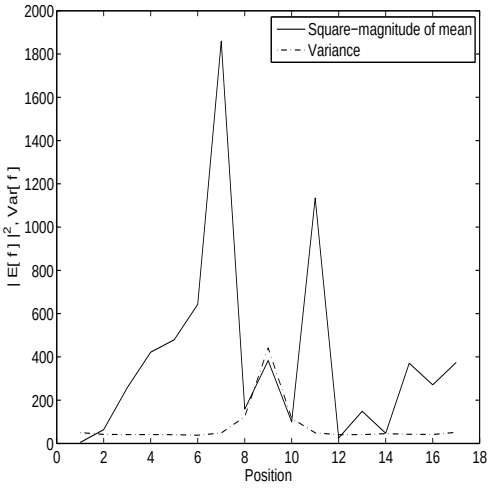


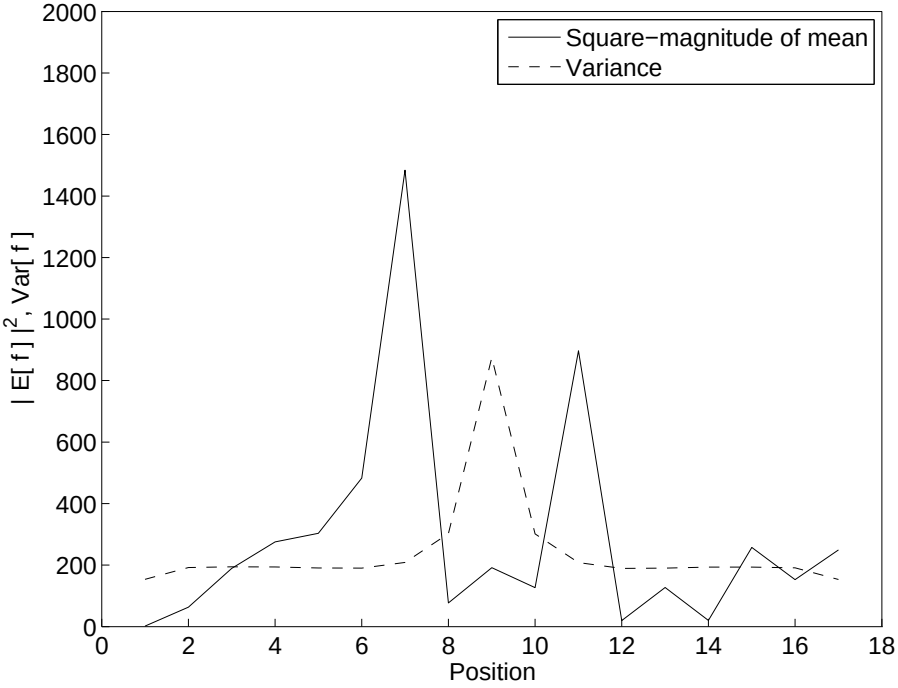
Figure 3.8: Recovered field power range profile, input SNR 20 dB.

magnitude. However, the mean profile demonstrates the bias in complex-field estimates introduced by the addition of noise in the image. The sample values at positions 4, 5 and 6 all have high mean powers, when ideally they should be close to zero. Not much can be done to alleviate this problem – the extra noise places a limit on super-resolution performance. The variance profile shows the additional uncertainty associated with the central positions when compared to the edge positions.

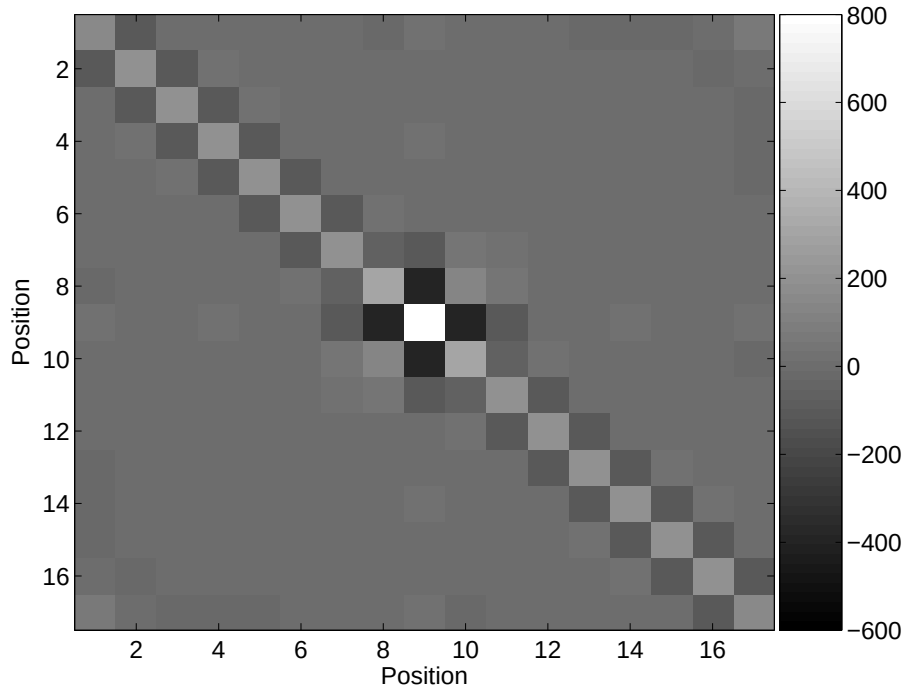
3.2.5 Comparison of Analytic and Monte Carlo Results

In the previous two sections we saw results of the Monte Carlo-based M-H algorithm for high and low SNRs. Here we present the analytic solution of the low-SNR scenario and compare it to the Monte Carlo result. The noisy image shown in Figure 3.5b was used as the input to the analytic solution given in equations (3.5)-(3.7). The prior covariance matrix used was the same as that used for the M-H algorithm – a diagonal matrix with all elements set to 1000.

The power range profile calculated from the mean of the analytic solution is shown in Figure 3.9a. This can be compared directly with the mean power range profile determined from the Monte Carlo samples, shown in Figure 3.8. It can be seen that the analytic solution closely matches the Monte Carlo solution. Also shown in Figure 3.9a is the analytically-derived variance as a function of

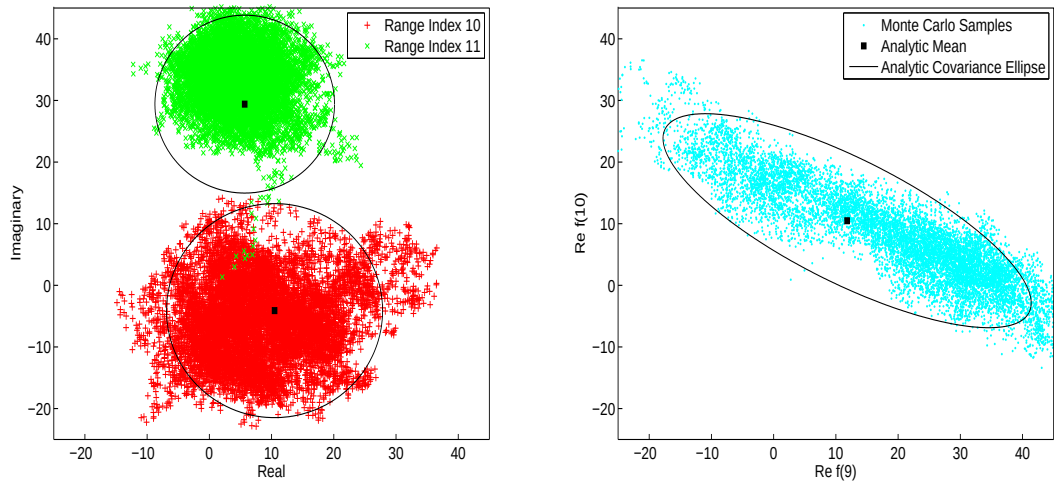


(a) Mean Range Profile



(b) Covariance Matrix

Figure 3.9: Analytic Bayesian solution. (a) Mean range profile. (b) Real part of the covariance matrix.



(a) Range index 10 and 11: complex values

(b) Range index 9 and 10: real values

Figure 3.10: Comparison of sample values output by the M-H algorithm, and covariance ellipses from the analytic Bayesian solution.

position. Most positions have a low-to-medium variance but the variance of the centre position is very high. This was seen earlier in the spread of samples for position 9 as output by the Monte Carlo algorithm.

To analyze all information about uncertainty in the recovered scene the full covariance must be examined. A graphical representation of the real part of this matrix is shown in Figure 3.9b. The variance graph in Figure 3.9a is a plot of the diagonal values of this matrix. Most values of the matrix are near to zero, which implies a low correlation between the scene values at widely separated parts of the image. However, there is a strong negative correlation between adjacent positions, evident from the dark squares on the diagonal situated one position away from the main diagonal. This is to be expected because the effect of two adjacent scene values on the image tends to get blurred out by the point spread function. Thus when either the first scene value is high and the second is low or vice versa a similar image will result. However, if both scene values are high or low the image would look different. The negative correlation is especially strong between the central position and its two adjacent positions. This provides another explanation of why the variance of the central position is so high: whether the central value is high or low, its effect is minimized by changes in the value of adjacent positions that result in a modelled image

consistent with the measured input image, within an error defined by the SNR.

The above analysis is highly detailed given that the scene considered simply contains two ideal scatterers embedded in uncorrelated clutter. This has only been possible through the use of Bayesian statistics. Other super-resolution algorithms give a single solution to the problem and do not allow uncertainty in the recovered scene to be taken into account. The discovery of correlation between adjacent positions is a particularly useful result because this explains why there are many possible solutions to the super-resolution problem consistent with noisy data. While the existence of correlations could have been conjectured from a qualitative analysis, the Bayesian approach has given precise information as to which values are correlated and by how much. When super-resolution is used without a further specific application, this information is useful in determining why the results are good or bad for a certain scene and SNR. However, the full value of the Bayesian approach is achieved when a further stage of processing, such as automatic target recognition, is used after super-resolution. In this situation uncertainty in the scene can be used as a numerical input to the next processing stage. This could be crucial in a decision-making process, where existence of uncertainty changes the decision boundary. The process of combining super-resolution and automatic target recognition is covered in Chapter 5.

A graphical comparison of the distribution given by the analytic and Monte Carlo Bayesian solutions for three positional values is shown in Figure 3.10. In each diagram the analytic result is represented by a square symbol plotted at the mean of the distribution and an ellipse representing the covariance between the two variables displayed. The size of the ellipse is selected such that its semi-major and semi-minor axes are equal to one standard deviation in a direction along each axis. The major axis of the ellipse is oriented along the line of highest correlation. The Monte Carlo result is represented as before by the sample values output by the M-H algorithm. In Figure 3.10a the diagram shows the real and imaginary parts of the recovered scene at positions 10 and 11. For both positions there is a good agreement between the location and spread as determined by the two methods. Figure 3.10b shows the correlation between the real part of the recovered scene at positions 9 and 10. Agreement between the Monte Carlo and analytic results in both the direction and degree

of correlation is also very good. The main difference between the two methods is that the samples do not fully cover the extreme tails of the distribution so the variance calculated from samples is slightly lower than the analytically derived result.

The analytic solution is relatively easy to implement because it consists simply of calculating a vector and a matrix from other quantities using addition, multiplication and the matrix inverse. The Metropolis-Hastings algorithm is somewhat more difficult to implement due to its additional complexity. The fact that the super-resolution results generated by the two methods match so well is an indication that the M-H algorithm has been implemented correctly.

The two scenarios discussed thus far have demonstrated basic operation of the Monte Carlo-based M-H algorithm where an analytic solution was available for comparison. In fact for those situations it is unnecessary to use the more computationally intensive M-H algorithm. The next section describes a scene model where an analytic solution is not available and there is no choice other than to use a Monte Carlo algorithm.

3.3 Cross-Section Recovery

3.3.1 Monte Carlo Algorithm

In Section 3.2.1 it was possible to derive an analytic expression for $p(\mathbf{f}|\mathbf{g})$. Given that $\boldsymbol{\sigma}$ is a more fundamental descriptor of a target than $\mathbf{f}$, it would be desirable to calculate $p(\boldsymbol{\sigma}|\mathbf{g})$ analytically. However, the additional complication of the scattering model makes this problem intractable even with basic forms for the prior $p(\boldsymbol{\sigma})$, such as the multivariate uniform or Gaussian distributions. Indeed, a Gaussian prior distribution for $p(\boldsymbol{\sigma})$ would be inappropriate because the elements of $\boldsymbol{\sigma}$ can take only positive values. Since an analytic solution for $p(\boldsymbol{\sigma}|\mathbf{g})$ is unavailable an algorithm such as the Metropolis-Hastings algorithm introduced in Section 3.2.2 must be used to obtain a numerical result. The M-H algorithm for cross-section recovery is now presented. This is a modified version of the algorithm for field recovery that takes into account the scattering model.

The Bayesian model describes uncertainty in the cross section $\boldsymbol{\sigma}$ by the

conditional probability density $p(\boldsymbol{\sigma}|\mathbf{g})$. Bayes' theorem gives the density as

$$p(\sigma|\mathbf{g}) = \frac{p(\mathbf{g}|\boldsymbol{\sigma})p(\boldsymbol{\sigma})}{p(\mathbf{g})}. \tag{3.12}$$

The likelihood of the image is given by

$$\begin{aligned} p(\mathbf{g}|\boldsymbol{\sigma}) &= \int p(\mathbf{g}|\mathbf{f},\boldsymbol{\sigma})p(\mathbf{f}|\boldsymbol{\sigma})\ d\mathbf{f} \\ &= \int p(\mathbf{g}|\mathbf{f})p(\mathbf{f}|\boldsymbol{\sigma})\ d\mathbf{f}, \end{aligned} \tag{3.13}$$

where we have noted that $\mathbf{g}$ does not depend directly on $\boldsymbol{\sigma}$. Using the scattering model $p(\mathbf{f}|\boldsymbol{\sigma})$ from equation (3.1) and the imaging model $p(\mathbf{g}|\mathbf{f})$ from equation (3.2) this can be shown to be [88]

$$p(\mathbf{g}|\boldsymbol{\sigma}) = \frac{\exp(-\mathbf{g}^H\mathbf{M}^{-1}\mathbf{g})}{\det(\pi\mathbf{M})}, \tag{3.14}$$

where

$$\mathbf{M} \equiv \mathbf{T}\boldsymbol{\Sigma}\mathbf{T}^H + \mathbf{N}, \tag{3.15}$$

and $\boldsymbol{\Sigma} = diag(\boldsymbol{\sigma})$. With choice of a suitable prior $p(\boldsymbol{\sigma})$ for the cross section the quantity of interest is then

$$\pi(\boldsymbol{\sigma}|\mathbf{g}) = p(\mathbf{g}|\boldsymbol{\sigma})p(\boldsymbol{\sigma}). \tag{3.16}$$

As with the field recovery algorithm, at each iteration we update each element of $\boldsymbol{\sigma}$ in turn. During the update at the ith iteration, a proposed new sample for a single cross-section element is generated from a proposal distribution $q(\boldsymbol{\sigma}^{\mathbf{i+1}}|\boldsymbol{\sigma}^{\mathbf{i}})$. The proposed sample is accepted with a probability $\alpha(\boldsymbol{\sigma}^{\mathbf{i}}, \boldsymbol{\sigma}^{\mathbf{i+1}})$, where

$$\alpha(\boldsymbol{\sigma}^{\mathbf{i}}, \boldsymbol{\sigma}^{\mathbf{i+1}}) = \min\left[\frac{\pi(\boldsymbol{\sigma}^{\mathbf{i+1}}|\mathbf{g})q(\boldsymbol{\sigma}^{\mathbf{i}}|\boldsymbol{\sigma}^{\mathbf{i+1}})}{\pi(\boldsymbol{\sigma}^{\mathbf{i}}|\mathbf{g})q(\boldsymbol{\sigma}^{\mathbf{i+1}}|\boldsymbol{\sigma}^{\mathbf{i}})}, 1\right]. \tag{3.17}$$

For ease of notation in this equation we have again omitted the dependence on cross-section element number so that while updating the jth element of $\boldsymbol{\sigma}$ we have in fact

$$\boldsymbol{\sigma}^{\mathbf{i}} = \left[\sigma_1^{i+1}, \sigma_2^{i+1}, ..., \sigma_{j-1}^{i+1}, \sigma_j^i, \sigma_{j+1}^i, ..., \sigma_m^i\right]^T \tag{3.18}$$

and

$$\boldsymbol{\sigma}^{\mathbf{i+1}} = \left[\sigma_1^{i+1}, \sigma_2^{i+1}, ..., \sigma_{j-1}^{i+1}, \sigma_j^{i+1}, \sigma_{j+1}^i, ..., \sigma_m^i\right]^T. \tag{3.19}$$

Once the burn-in samples have been removed, remaining samples are distributed from $p(\boldsymbol{\sigma}|\mathbf{g})$ as required.

3.3.2 Results for a High Signal-to-Noise Ratio

A demonstration of the Metropolis-Hastings algorithm for RCS recovery is now given. The simulated scene and image used in this demonstration are identical to those introduced in Section 3.2.3. Figure 3.1a shows the scene and Figure 3.1b the noisy low-resolution image, which has an SNR of 50 dB. The detailed processing parameters used in the M-H algorithm are shown in Table 3.2. The choice of prior in this example was difficult. If a uniform prior $p(\sigma_k) \propto 1$ is used, it is not possible to normalize the posterior distribution. This is also the case for the scale-parameter-type prior $p(\sigma_k) \propto 1/\sigma_k$ because $\boldsymbol{\sigma}$ is part of a hierarchical model [50]. A compromise between these two was to use an exponential prior.

Figure 3.11 shows the sample time series for range index 7, which corresponds to the left hand scatterer of Figure 3.1a. The immediately noticeable difference between this graph and the equivalent one for field recovery (Figure 3.2a) is that there is a lot more variation in the RCS graph than the field graph. Although many samples are near the true RCS value of 1000, there are a significant number of major variations from this value.

It is more instructive to examine the histogram of the sample series, which is shown in Figure 3.12a. Here we more clearly see the distribution of samples is indeed concentrated near a value of 1000 and also that the distribution is skewed to the right. According to the distribution, there is a very low probability the RCS for this range index takes a value near zero. This is to be expected since a scatterer exists at this position. Ideally, the peak of the distribution would be very near to the true value of 1000, especially for this high signal-to-noise ratio. However, the position of the peak has been biased to the left due to use of the exponential function as the prior probability density function. Nevertheless, the bias is not huge and a large proportion of the distribution is near the true value.

The histogram also reveals a significant tail is present in the RCS distribution. This exists because of the nature of the Gaussian scattering model introduced in Section 3.1.2. According to the model, the scattered complex field of a target at a particular position on the target could have a low value when imaged even if the underlying RCS is high, because of multiplicative Gaussian speckle. Conversely, if it is determined by the M-H super-resolution algorithm that the target has a low-magnitude scattered field at a certain position, it is not known whether this is because the underlying RCS is low or the underlying

Number of saved samples	10000
Burn-in period	0
De-correlation gap	10
Proposal distribution	Uniform random walk
Random walk step size	0.5
Prior distribution $p(\boldsymbol{\sigma})$	Independent negative exponential
Prior mean	1000
Sample start value	0

Table 3.2: Table of parameters used by the M-H algorithm for RCS recovery.

RCS is high and speckle caused destructive interference in this particular instance. Thus high values of RCS have a significant probability even if the field magnitude is relatively low.

The histogram of the RCS for range index 8, which corresponds to a low-magnitude clutter sample, is shown in Figure 3.12b. The shape of this histogram is markedly different to that of range index 7. The peak is at an RCS value of zero and the tail rapidly drops off to give a low probability for high RCS values. This is a good result – the algorithm has determined it is unlikely a high-magnitude scatterer exists at this position. The histogram of the RCS for range index 9, which corresponds to another low-magnitude clutter sample, is shown in Figure 3.12c. This is similar in shape to that of range index 8 as are the histograms of all the other clutter samples, not shown here to save space.

The final histogram for this example is shown in Figure 3.12d and corresponds to the second scatterer situated at range index 11. This histogram is similar to that of the first scatterer at range index 7 since they both have a true RCS value of 1000 units. Thus all comments about the histogram for the first scatterer apply to the second.

The mean and MAP estimated range profiles calculated from the M-H samples are shown in Figure 3.13a. The mean profile is greater in magnitude than the true profile at every position of the profile. This is because the distribution of the RCS is skewed to the right. In fact, for a skewed distribution such as this it could be considered inappropriate to use the mean profile as an estimate of the scene. The MAP profile more closely matches the shape of the true

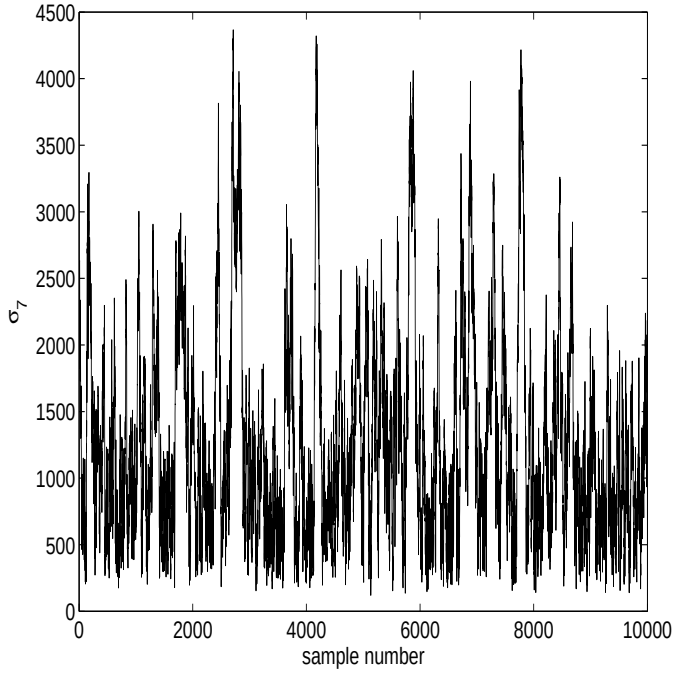
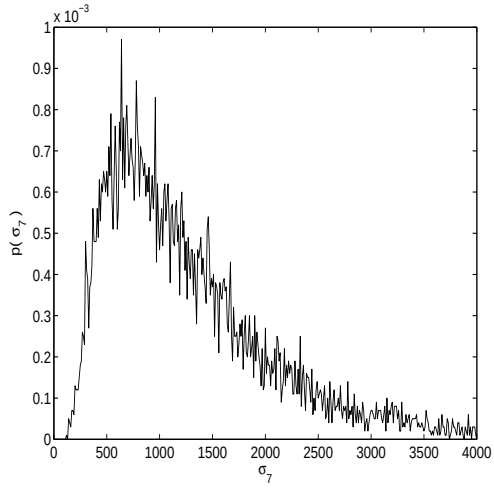


Figure 3.11: Time series of samples for range index 7 as generated by the M-H algorithm using a decorrelation gap of 10 and an input SNR of 50 dB.

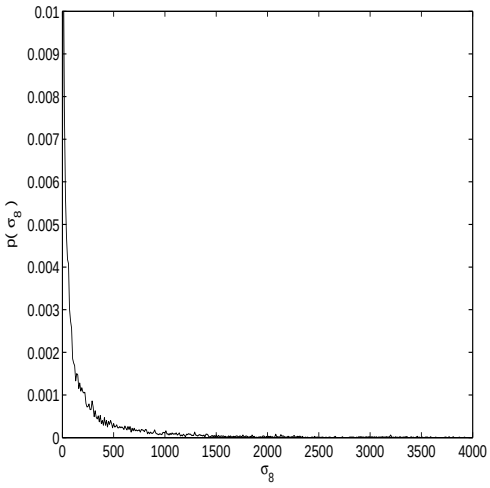
scene but the bias in the estimate of the RCS of the two scatterers introduced by the exponential prior distribution results in a slightly low estimate for the magnitude for those scatterers. However, overall the MAP estimate is a good representation of the scene.

For comparison, the standard deviation of the range profile is shown with the mean profile in Figure 3.13b. In this context the standard deviation of RCS is equivalent to the variance of the recovered field, because units of RCS are proportional to the square of the field magnitude. However, the standard deviation of RCS follows a different pattern to the equivalent field variance. We recall that the field variance did not depend on the true scene and the variance of the middle position was higher than all other positions. In contrast, due to the skewed distribution of RCS, the standard deviation of RCS does depend on the true scene, being higher at positions where there are high-magnitude scatterers[1].

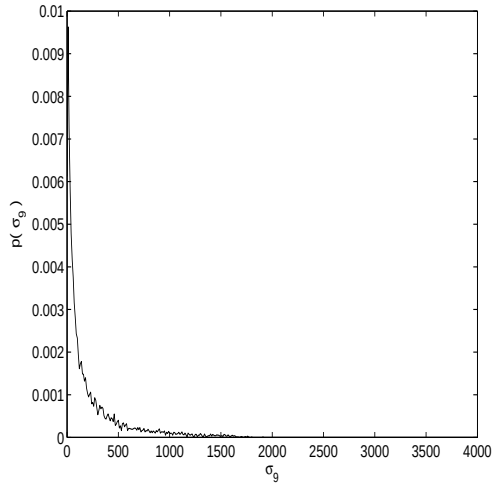
[1]One exception is at range index 5, where there is a large standard deviation even though the RCS at this position should be small. An explanation is that in skewed distributions the standard deviation estimate can be greatly affected by a few high-magnitude samples whose large distance from the mean is squared in the calculation, giving them an undue high weight.



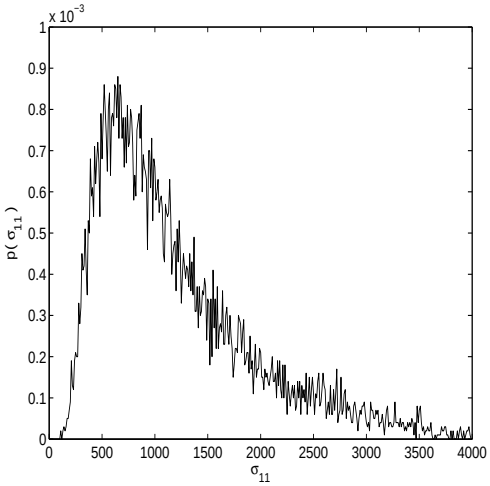
(a) Range Index 7



(b) Range Index 8

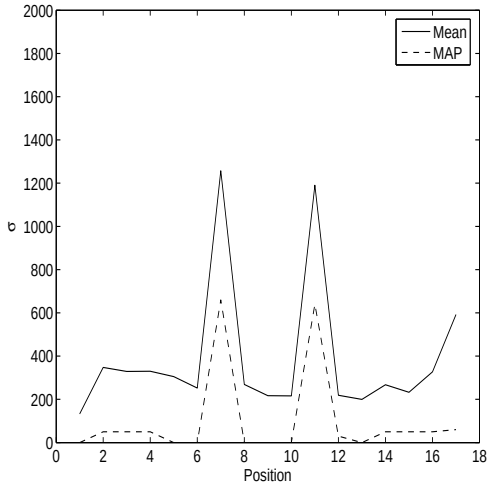


(c) Range Index 9

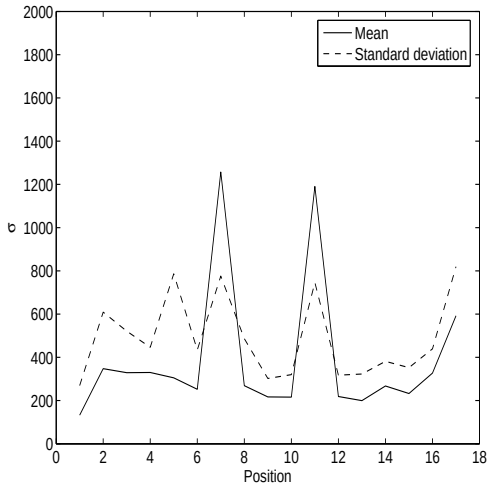


(d) Range Index 11

Figure 3.12: Histogram of samples for various range indices, generated by the M-H algorithm with an input SNR of 50 dB.



(a) Mean and MAP



(b) Mean and standard deviation

Figure 3.13: Range profile representations calculated from samples, with an input SNR of 50 dB.

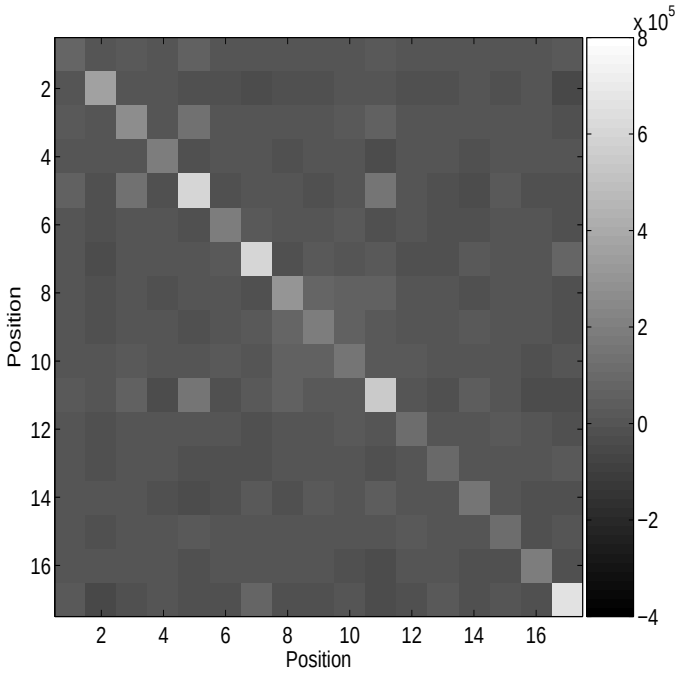


Figure 3.14: Sample covariance matrix.

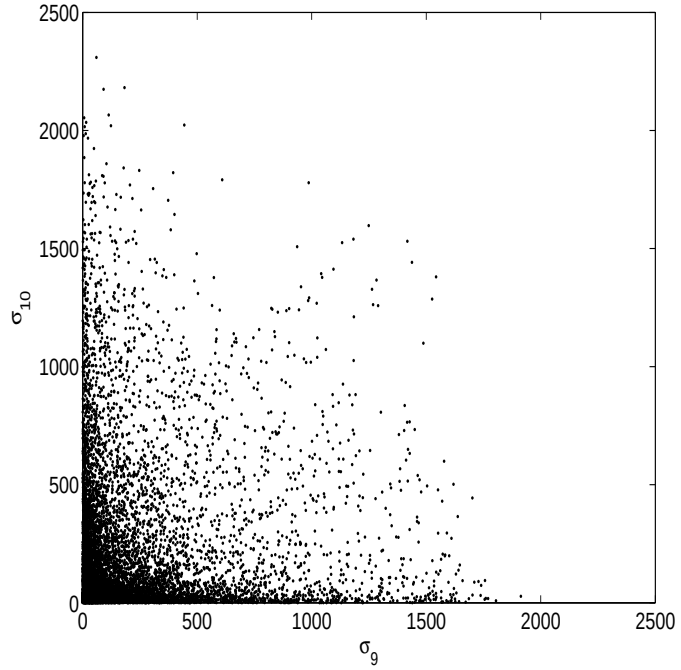


Figure 3.15: Distribution of samples for range indices 9 and 10.

The fact that the shape of the standard deviation of RCS is different to that of the field variance shows that the uncertainty in RCS dominates uncertainty in the field distribution.

The sample covariance matrix is shown in Figure 3.14. The standard deviation of RCS discussed in the previous paragraph is the square root of the main-diagonal values of this matrix. In contrast to the field covariance matrix, the RCS covariance matrix has little correlation between range indices, even those at adjacent positions. Again, this is because the large uncertainty in the RCS dominates uncertainty in the field distribution and has removed correlations. A graphical demonstration of this is shown in Figure 3.15, where the sample values of range indices 9 and 10 are plotted against each other. We clearly see the distribution is concentrated at a value of zero for both indices and there is no correlation between the indices.

3.3.3 Results for a Low Signal-to-Noise Ratio

A demonstration of the Metropolis-Hastings algorithm for RCS recovery at a lower SNR of 20 dB is now given. The simulated scene and image used in this demonstration are identical to those introduced in Section 3.2.4. Figure 3.5a shows the scene and Figure 3.5b the noisy low-resolution image. The detailed processing parameters used in the M-H algorithm are the same as those used in the high SNR case and are shown in Table 3.2.

Figure 3.16 shows the sample time series for range index 7, the position of the first scatterer. The corresponding histogram is shown in Figure 3.17a. Compared to the high SNR example, this distribution is slightly more skewed to the right, with a larger variation in RCS values. This is a result of the extra uncertainty introduced by the additional noise.

Figure 3.17b shows the histogram for range index 8, a position corresponding to low-magnitude clutter. The difference between this histogram and the equivalent high-SNR histogram in Figure 3.12b is more marked than the difference between the scatterer histograms of range index 7. In the low-SNR case the extra noise has caused significant additional uncertainty in the RCS. This is because the SNR of 20 dB is the same as the signal-to-clutter ratio and the algorithm has difficulty in determining whether high magnitude values are due to noise or clutter. Indeed at very low SNRs, the image gives very little new

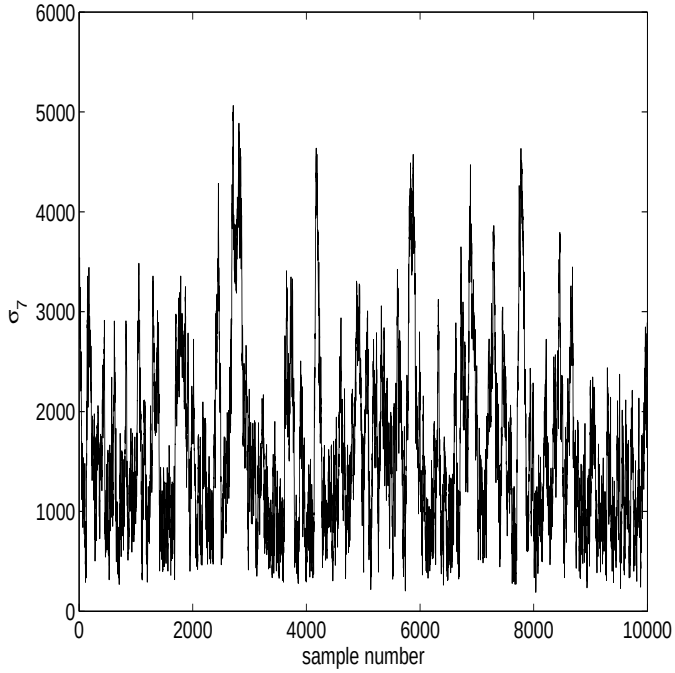
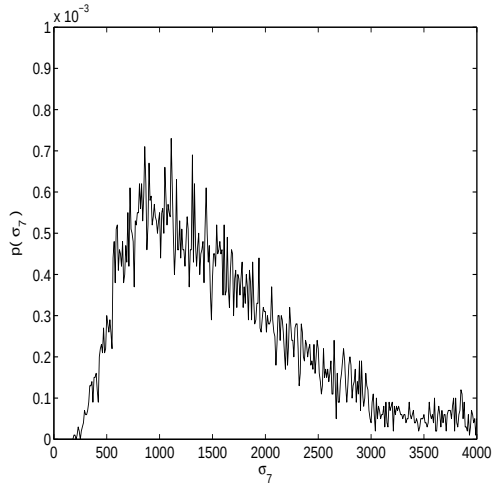


Figure 3.16: Time series of samples for range index 7 as generated by the M-H algorithm using a decorrelation gap of 10 and an input SNR of 20 dB.

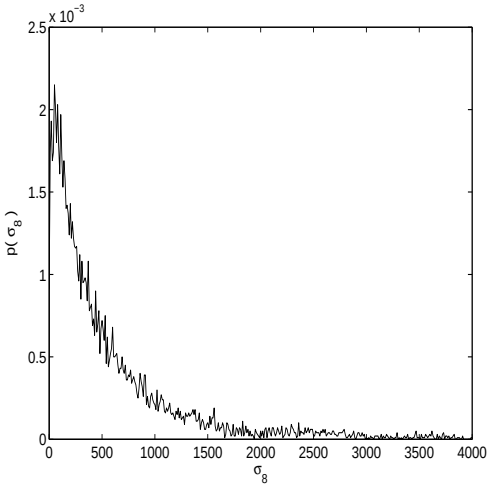
information and the posterior distribution tends towards the prior distribution, which in this case has been set to the negative exponential function. Figure 3.17c shows the histogram for range index 9, another clutter sample. Here the distribution is wider than that for range index 8 and is closer to the prior distribution.

The final histogram for this example, corresponding to the second scatterer at range index 11, is shown in Figure 3.17d. This histogram is shifted to the left a little compared to that of the first scatterer at range index 7. This result is in line with that of the 20 dB SNR field-recovery example where the second scatterer was determined to have a lower magnitude than the first for this particular noise realization. Thus bias in the scene estimate introduced by noise is present when recovering the RCS as well as the complex field.

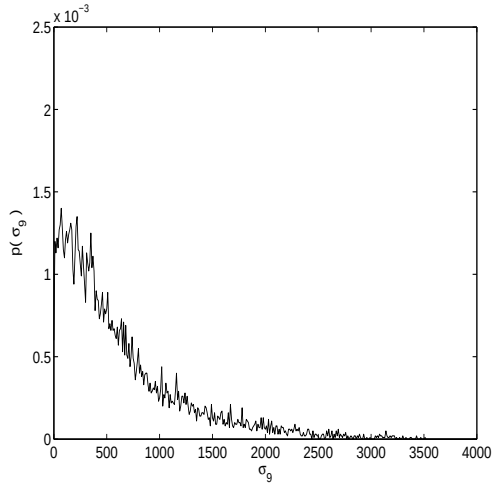
The mean and MAP estimated range profiles are shown in Figure 3.18a. The effect of a skewed RCS distribution combined with higher noise levels causing bias and an increase in uncertainty, are apparent in the distorted mean profile. Although the two scatterers at positions 7 and 11 have the highest magnitude,



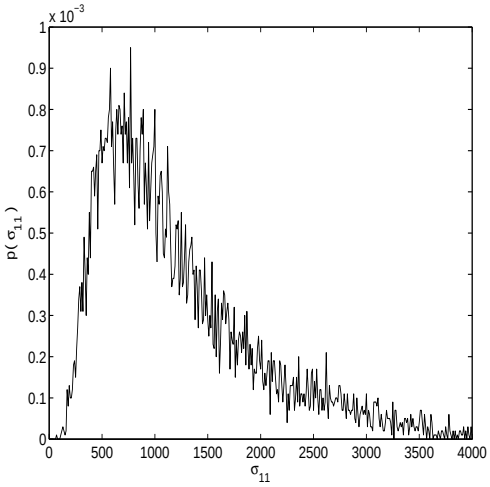
(a) Range Index 7



(b) Range Index 8



(c) Range Index 9



(d) Range Index 11

Figure 3.17: Histogram of samples for various range indices, generated by the M-H algorithm with an input SNR of 20 dB.

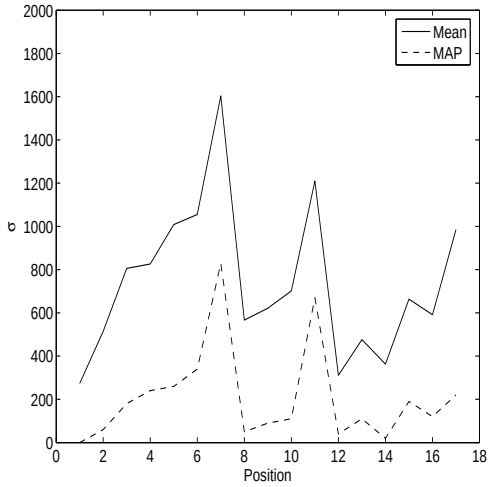
the mean estimated RCS at the clutter positions is relatively high. Looking at the more appropriate MAP profile the clutter levels are not as high as in the mean case both in absolute terms and relative to the magnitude of the scatterers. A detection algorithm using the MAP profile should be able to detect the scatterers without too high a false alarm rate.

The standard deviation of the range profile is shown in Figure 3.18b along with the mean. The skewed distribution generally results in high standard deviations where the mean is high. The covariance matrix is shown in Figure 3.19. As with the high-SNR case there is little correlation between any of the range indices – even adjacent ones. This is graphically demonstrated in Figure 3.20, where the sample RCS values at range indices 9 and 10 are plotted against each other. There is no discernable correlation.

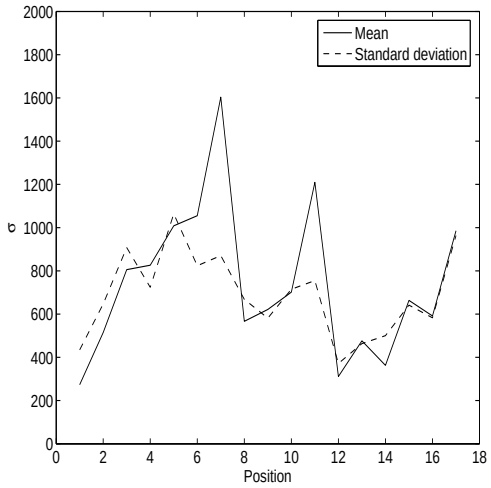
3.4 Discussion

3.4.1 Algorithm Complexity

The MCMC super-resolution algorithm provides a useful representation of the probability distribution of the RCS $\boldsymbol{\sigma}$. However, this comes at a computational cost. At the heart of the algorithm is the calculation of the matrix inverse $\mathbf{M}^{-1}$ in equation 3.14. This matrix must be recalculated at every iteration of the algorithm. It is well known that calculation of the inverse of a matrix with $m \times m$ elements is an operation of order m^3 [116]. Furthermore, with the version of the algorithm that updates a single element of RCS per iteration this operation is carried out for each of the m elements of RCS resulting in a total computation time of order m^4. The computation time is also proportional to the number of samples generated, which has to be a large number when dealing with high-dimensional data such as radar data. While it would be possible to reduce the order of the algorithm to m^3 by updating all RCS elements in one go, a smaller update step in the proposal distribution would be required to avoid rejecting too many samples. This results in highly correlated RCS values from sample to sample and a greater number of samples would have to be collected to completely explore the distribution space. The increase in computation time collecting extra samples more than offsets the reduction due to the all-in-one update. Thus it appears that no fundamental speed increase is possible for



(a) Mean and MAP



(b) Mean and standard deviation

Figure 3.18: Range profile representations calculated from samples, with an input SNR of 20 dB.

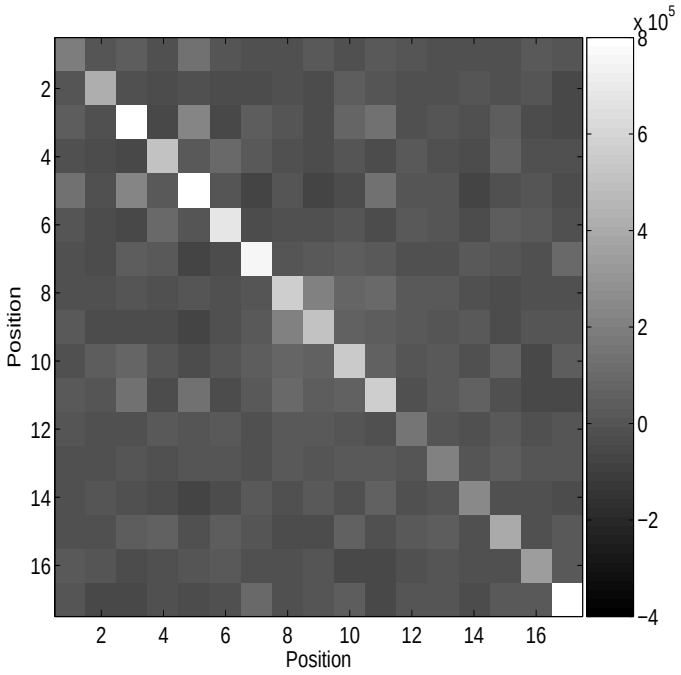


Figure 3.19: Sample covariance matrix.

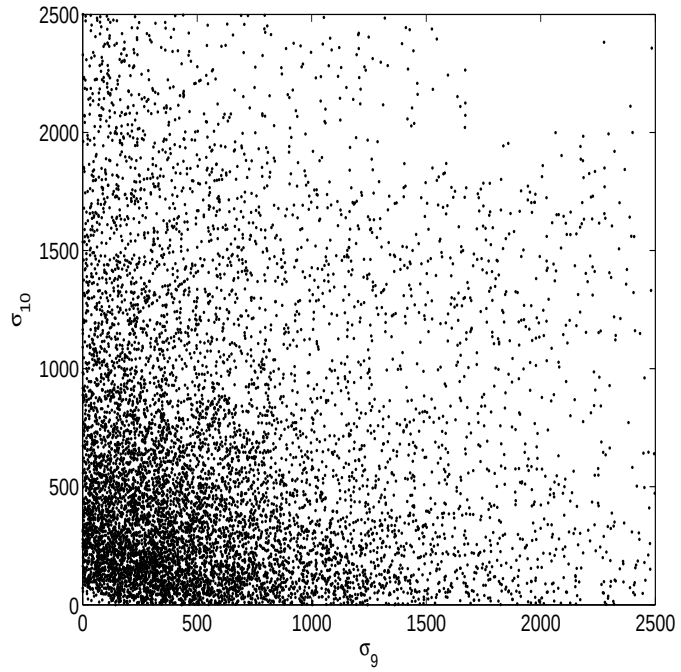


Figure 3.20: Distribution of samples for range indices 9 and 10.

the basic algorithm. The slowness of the MCMC algorithm means that at best it could be used for small range profiles. It would be impractical to use the algorithm on two-dimensional images as the larger number of pixels would result in very long computation times limiting the usefulness of the procedure.

The algorithm for field recovery also requires the calculation of two inverse matrices: $\mathbf{N}^{-1}$ and $\mathbf{W}^{-1}$ in equations 3.2 and 3.4 respectively. However, since these matrices do not depend on the current value of the complex field they only need to be calculated once at the beginning of the algorithm. Moreover, both matrices are usually diagonal in a practical implementation, hence the inverse calculation is of order m and does not make a significant contribution to the overall run-time of the algorithm. The largest contribution arises from the matrix-vector multiplications in equation 3.2, which are of order m^2. Since the algorithm updates each element of the complex field in turn, the total computation time is of order m^3. This makes operation of the algorithm feasible for realistic-sized range profiles and small 2D target image chips.

Two-dimensional super-resolution requires a high amount of processing power – if an image is NxN pixels large then the computation time is of order N^6 for complex filed recovery. If the point spread function of a system is separable into two dimensions, it may seem tempting to perform two-dimensional super-resolution in two one-dimensional steps. This would reduce the computation time to $2N^4$. However, the super-resolution process effectively adds noise in proportion to the current amount of noise in the data. If two super-resolution processes were applied to the data in turn, the constant of proportionality would be squared. In practice, the resulting super-resolution output would be too noisy to be useful.

3.4.2 Model Appropriateness

Results presented in the preceding sections demonstrated basic operation of the MCMC Bayesian super-resolution algorithm for RCS recovery. The scattering model was mentioned as the cause of the wide and skewed distribution of the recovered RCS. We now discuss further ramifications of the scattering model.

In the examples given thus far the complex scattered field consisted of two large magnitude peaks in a background of weak clutter. However, even when two scatterers are present as defined by the RCS range profile, it is perfectly possible

under the Gaussian scattering model that the scattered field could have a low magnitude at all positions. If this were to be the case, both the scattered field magnitude and RCS power profiles estimated from a low-resolution image would have low magnitudes throughout. The skewed distribution of RCS values allows consideration of the potential presence of scatterers but no indication would be given as to their position. The difficulty is that estimating the RCS from a single realization of the complex scattered field is an ill-posed inverse problem, which exists in addition to the already-present inverse problem of determining the high-resolution field from a low-resolution noisy image.

Although problems associated with estimating the underlying RCS have been considered in the general radar literature [8], they appear not to have been addressed in the super-resolution literature. For example, in a paper by Luttrell [88], which uses the same scattering model as that used here, the example given has two scatterers that happen to have a high-magnitude complex field at the appropriate points. Thus when RCS recovery is performed, reasonable estimates of the RCS are given even though in general this may not be possible.

Although it is difficult to estimate the underlying RCS from a single realization of the complex field, if there are multiple realizations available from several images or range profiles the estimation becomes easier. Every extra sample improves RCS estimation, which may alternatively be considered to be a speckle reduction problem. Two very simple speckle reduction techniques are available for SAR images – the polarimetric whitening filter (PWF) and multi-look averaging. The PWF is applicable only to polarimetric SAR and uses three polarization channels to obtain an optimal estimate of the underlying RCS [103]. Multi-look averaging uses non-overlapping sections of range-Doppler space to form several images of the same scene area. The average intensity of these independent images gives a speckle-reduced image because the probability distribution of the intensity changes from an exponential distribution for single-look images to a gamma distribution for multi-look images with the shape parameter dependent on the number of looks. This process, however, results in a loss of resolution due to the fact that smaller sections of range-Doppler space are used to form each image [9].

It may often be the case that a fully polarimetric radar is not available or the loss in resolution caused by multi-look averaging is unacceptable. In either

case more advanced speckle reduction algorithms can be used. One algorithm by Stewart *et al.* [141], which is cast as an image segmentation algorithm, attempts to estimate the average RCS of homogeneous areas of clutter. Each pixel in a region of clutter contributes independent information to the RCS estimate. While this algorithm is good at segmenting clutter regions, it is not an appropriate tool for target RCS estimation. This is because a target's RCS fluctuates much more rapidly in a spatial sense than clutter RCS. Thus there seem to be no suitable candidate techniques to obtain a significant number of independent realizations of the target's high-resolution spatial distribution at a particular aspect angle.

A second problem with target RCS estimation is the validity of the scattering model itself. The Gaussian speckle model arises from the central limit theorem – if a large number of scatterers are present in a single resolution cell then these sum coherently via a random walk process to form a zero-mean circularly complex Gaussian random variable. The limit is applicable as long as no individual scatterers dominate the sum. This model is entirely appropriate for clutter scenes such as those containing grass, trees, or sea because there are a large number of blades of grass, leaves, or variations in the sea surface to produce many scatterers of similar strength in a resolution cell. However, in man-made targets, surfaces often form dihedral or trihedral reflectors, which have a large RCS compared to other parts of the target in a resolution cell. This effect becomes more apparent as the resolving power of the radar increases and the size of a resolution cell drops. In this situation the Gaussian speckle model is no longer appropriate.

There are several candidates for a more accurate target RCS fluctuation model. One theoretically-based model is the Rice distribution. This arises when there is one dominant scatterer and many weak scatterers in a resolution cell [136]. The complex distribution is a circularly complex Gaussian random variable shifted from the origin. The magnitude of this variable follows the Rice distribution, which has two parameters - one relating to the power of the dominant scatterer and another relating to the average power of the weak scatterers. Two other models for target RCS fluctuation use the two-parameter Weibull or gamma distributions. These are empirical models that have been found to fit well to various radar data sets [3, 136, 154]. The advantage of these

models is that in general the variation in RCS is lower than that in the Gaussian speckle model. This means that, for a given number of looks at the target, the average RCS at each point on the target is known to a greater accuracy. The disadvantage of the two-parameter models is the need to estimate an additional parameter over the simpler one-parameter Gaussian speckle model. Indeed, it is not possible to obtain estimates for both parameters of the two-parameter models with a single realization of the target. Given that we would like to perform super-resolution with RCS estimation using a single image, neither of the two-parameter models seem appropriate.

3.4.3 Conclusions

We have seen that the computational cost of the super-resolution RCS recovery algorithm is much higher than that of field recovery. In addition, due to the scattering model the estimated probability distribution of the recovered high-resolution RCS is so wide and skewed it does not provide much useful information. Even if a more appropriate two-parameter model were used it would not be possible to estimate the parameters using a single image. In ideal circumstances we would like to be able to estimate the underlying RCS as it is a more fundamental target characteristic than the scattered field. However, given all these problems with RCS recovery a recommendation of this thesis is to concentrate future effort on field recovery when attempting to estimate the full probability distribution of target parameters.

3.5 MMSE Super-resolution Re-visited

3.5.1 Introduction

Since the inverse cross-section problem appears to be intractable both in terms of computation time and its double inverse nature we concentrate on recovering the scattered field. Previously it was shown that using a Gaussian distribution for the complex-field prior allowed an analytic solution to the problem. Since the MMSE algorithm generates the mean of the posterior distribution and it is a much faster algorithm than the MCMC algorithm we now re-visit the MMSE algorithm to see if it can be used as part of a Bayesian solution.

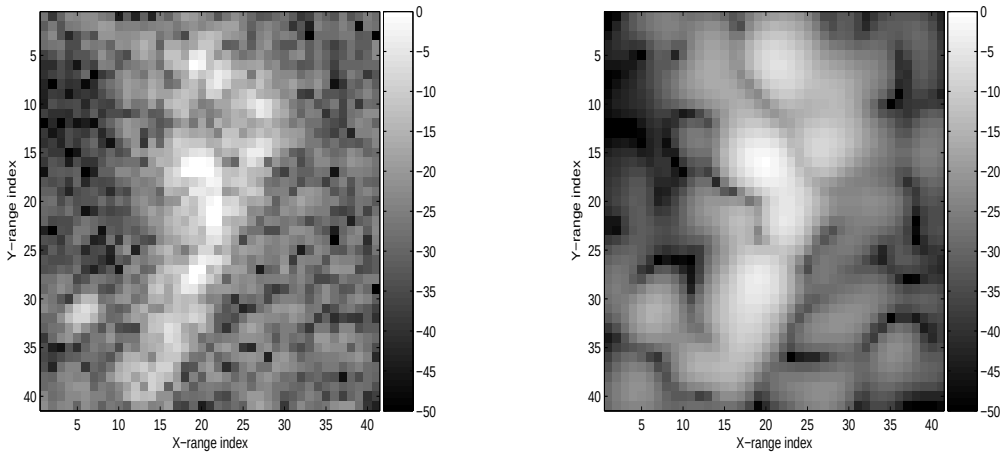
3.5.2 MMSE as an Approximate Bayesian Solution

One problem in using the true Bayesian solution to complex-field recovery is that prior knowledge of the scene is usually not well defined. This fact is the motivation behind the iterative nature of the MMSE super-resolution algorithm, which effectively estimates and refines the prior knowledge at each iteration. The MMSE algorithm on its own converges towards the mean of the posterior scene distribution. However, once the mean has been calculated via an iterative approach the covariance can be calculated in one step. Recall that the inverse covariance of the analytic solution for complex-field recovery is $\mathbf{C}^{-1} = \mathbf{W}^{-1} + \mathbf{T}^H\mathbf{N}^{-1}\mathbf{T}$. Previously, $\mathbf{W}$ was a quantity that had to be defined before the analytic solution was possible. However, since an estimate of $\mathbf{W}$ is available from the last iteration of the MMSE algorithm this can be substituted into the covariance matrix equation. The mean and covariance of the posterior scene distribution calculated in this way are all that is needed to define the complete Bayesian solution for complex-field super-resolution.

3.5.3 Two-dimensional Super-resolution with MMSE

All the super-resolution results presented thus far have been based on simulations of one-dimensional data. We now demonstrate operation of the MMSE-T super-resolution algorithm on example two-dimensional measured radar imagery from the MSTAR data set. The high-resolution scene $\mathbf{f}$ was taken to be a 41x41 pixel image of a T-72 tank. The point spread function is based on the FFT of a Hamming window as described in Section 2.4.5. The high-resolution and low-resolution images relating to this setup are shown in Figure 3.21.

The low-resolution image had noise added to it with SNRs ranging from 0 dB to 50 dB. At each SNR ten noisy images were generated, the MMSE-T algorithm was applied, and the output SNR was measured. The results of this experiment are shown in Figure 3.22, where both the peak and mean output SNRs are plotted against input SNR. As the input SNR increases, so does the output SNR for both metrics. It should be noted, however, that the peak output SNR is not directly comparable with the input SNR as they are calculated in different ways. The information-theoretic-based mean output SNR allows a better comparison because in general it is not possible for it to exceed the input



(a) High-resolution image

(b) Low-resolution image

Figure 3.21: Test imagery for the MMSE-T super-resolution algorithm.

SNR, a fact that provides an upper bound on performance.

The original and super-resolved images for the MMSE-T algorithm at an input SNR of 50 dB are shown in Figure 3.23. It can be seen that the algorithm has done a reasonable job of restoring the image from the low-resolution one shown in Figure 3.21b.

In addition to the mean field estimated by the MMSE-T algorithm, the covariance matrix was calculated as outlined in the previous section. It was found that the diagonal elements of the matrix dominated the calculation and the variance was approximately equal to the square magnitude of the mean field. The off-diagonal elements relating to adjacent pixels were slightly negative, as with the one-dimensional case. The advantages of calculating the covariance matrix are not immediately apparent here, as the matrix has no effect on the output SNR metric. Chapter 5 details how the covariance can be used to aid target recognition.

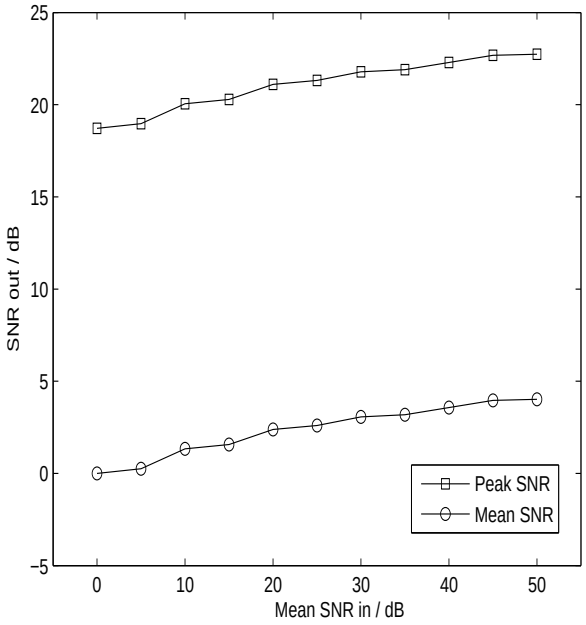
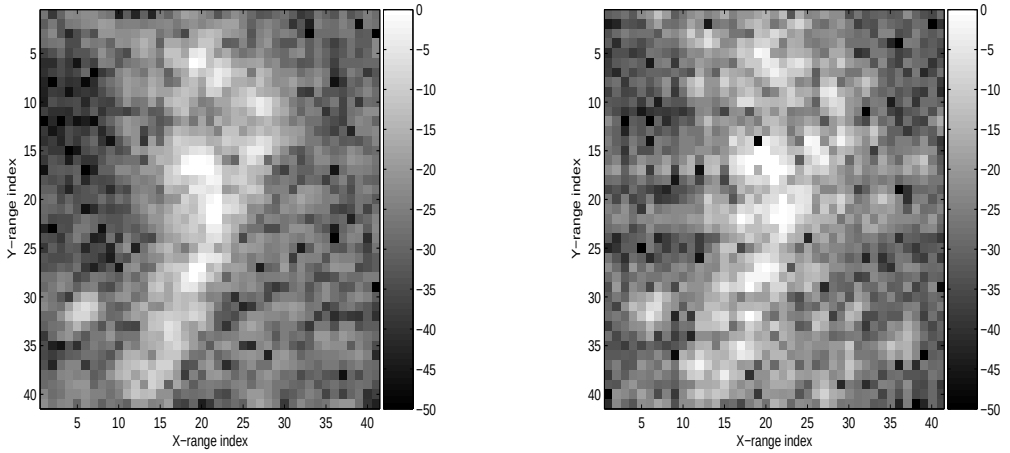


Figure 3.22: Performance of the MMSE-T super-resolution algorithm.



(a) Original high-resolution scene (b) Super-resolved high-resolution scene

Figure 3.23: Comparison of the original high-resolution scene with the MMSE super-resolved scene.

Chapter 4

Bayesian Autofocus and Super-resolution

4.1 Introduction

4.1.1 Introduction to Autofocus

The performance of super-resolution algorithms depends critically on the level of noise in the system and the accuracy with which the PSF is known. In the previous chapter all the examples given assumed a perfect knowledge of the PSF. In practical systems a variety of effects can alter the shape of the PSF. The the altered PSF usually has a reduced peak power, a broadened main lobe and increased side-lobe power.

If a target is moving then the Doppler shift induced in the signal received by the system changes the PSF - an effect characterized by the waveform ambiguity diagram as described in Chapter 2. Although Doppler processing or target tracking can provide corrections for target velocity, residual errors will remain resulting in a non-ideal PSF. If the target is vibrating due to motion of the engine this will also result in a blurred PSF even after application of autofocus techniques referred to in Chapter 2. These effects also apply if the radar imaging platform is moving instead of or in addition to the target. A detailed analysis of how the PSF is affected by platform motion is given by Blacknell [7].

Pulse eclipsing occurs due to the receiver being switched off while the system is transmitting a pulse; therefore the entire waveform will not be received for

targets that are very close to the transmitting system [164] or near the end of the range-unambiguous extent. This results in a degradation of the PSF similar to the Doppler effect [79]. Eclipsing is a particular problem in high-resolution systems with a high duty ratio, where the length of the pulse is large compared to the pulse repetition interval, because a large proportion of the range profile will be eclipsed.

Other effects that could alter the PSF are non-linearities in the radar hardware or atmospheric phase disturbances. Non-linearities are likely to be deterministic and could potentially be mitigated by measuring the form of non-linearity and compensating for this in software. However, noise or finite bandwidth would limit the degree to which this is possible. Atmospheric phase disturbance cannot be determined in advance, therefore some form of autofocus that can cope with this effect may be required.

4.1.2 Review of Autofocus Techniques

We now review a few of the more popular autofocus algorithms and those relevant to this thesis. This review is not comprehensive: it is merely intended as an introduction to the problem of autofocus. For an up-to-date review see [160].

Contrast optimization is one of the simplest autofocus algorithms. It assumes there is a residual quadratic phase error in the data and forms images with a range of values for the quadratic slope parameter. The parameter value that gives the image with the highest contrast, defined by the ratio of the standard deviation to mean, is used to phase-correct the data [106]. Contrast optimization can fail if certain contrived groups of targets are present in the scene [7]. However, it is unlikely in practice that such groups would be found in a typical scenario.

Multi-look registration, also known as map-drift autofocus, assumes that when two images of the same scene, obtained from different regions of range-doppler space, are aligned there is no phase error across the doppler bandwidth. If the images are not aligned then the alignment difference indicates what value of autofocus parameter should be used to phase correct the data. Multi-look registration and contrast optimization have similar performance and computation time [106].

One of the most popular autofocus algorithms in use today is phase gradient

autofocus (PGA) [151]. This algorithm consists of four main steps. The first step, known as centre shifting, aligns the strongest scatterers in each range bin. The second step windows the scatterer response to use only data in regions of high signal-to-noise ratio. The third step estimates the phase gradient using a minimum variance technique. The final step applies phase correction to the defocused image and repeats the first three steps until convergence. The advantage of PGA over the other techniques is that it can focus images subject to a wider variety of defocus effects such as wide-band phase error.

A recently proposed autofocus algorithm that is relevant to this thesis is a motion-compensated version of the CLEAN algorithm, known as MCCLEAN [159]. In an iterative process, this algorithm alternately uses the CLEAN algorithm to estimate the positions and amplitudes of scatterers in the scene, and determines the phase correction required to focus the image using a least squares approach. As an alternative, RELAX could be used in place of CLEAN to give the MCRELAX algorithm. This algorithm has a super-resolution capability in addition to its autofocus ability.

According to Blacknell [7], when algorithms such as contrast optimization and multi-look registration are operating at the depth-of-focus limit the residual error in the estimated PSF is sufficient to cause problems for super-resolution techniques such as MMSE. This is the case with or without the presence of noise. Performance cannot be improved even if uncertainty in the PSF is taken into account via the use of second order statistics [10]. However, Oliver [108] states that the information limit for the autofocus techniques is often considerably better than the depth-of-focus criterion suggests. As long as extended objects in the scene do not cause bias in the autofocus parameter estimate and only quadratic phase error is present, the information-limited error in the PSF does not upset super-resolution algorithms. If the radar platform does not have an inertial motion unit to compensate for high-frequency motions then the quadratic phase error assumption no longer holds and errors in the PSF again cause a degradation in super-resolution performance [107].

Luttrell [88] has proposed a joint autofocus and super-resolution algorithm, which maximizes the posterior distribution of the scene RCS and focus parameter under an information-based Bayesian framework. In simulations the algorithm is able to estimate the focus parameter to an accuracy beyond the

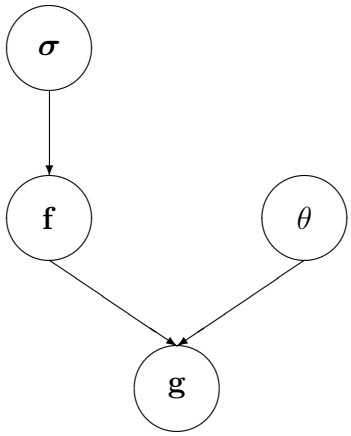


Figure 4.1: Autofocus/super-resolution model

depth-of-focus limit. However, the algorithm has not been tested on measured data and the model used may not match the true physics involved in the radar imaging process – see Appendix C and [78]. Even so, the Bayesian formulation of the algorithm has motivated the work of this thesis and the PSF is available in analytic form. Therefore Luttrell's model will be used here as a basic test of algorithms.

4.1.3 General Autofocus and Super-resolution Model

Here we present a joint autofocus and super-resolution Bayesian model based on the scattering and imaging models introduced in sections 3.1.2 and 3.1.3. This model was first used in [88] for a specific algorithm. However it is a fairly general model and could be adapted simply by using different individual scattering or imaging models.

A diagram of the overall scattering and imaging model is shown in Figure 4.1, using the notation

$$\begin{aligned} \boldsymbol{\sigma} &= \text{scattering cross section} \\ \mathbf{f} &= \text{complex scattered field} \\ \theta &= \text{focus parameters} \\ \mathbf{g} &= \text{complex image.} \end{aligned}$$

In the diagram each parameter is considered to have probability distribution

and the dependency of the parameters on each other is indicated by the arrows. The transition from $\boldsymbol{\sigma}$ to $\mathbf{f}$ is the scattering model and and the transition from $\mathbf{f}$ and θ to $\mathbf{g}$ is the imaging model. The overall model is identical to that used in the previous chapter with the exception of the focus parameters. These parameters define the PSF that generates the image $\mathbf{g}$ from the scattered field $\mathbf{f}$. Therefore the PSF matrix $\mathbf{T}$ of equation 3.2 is a function of θ. Note that in general θ could be a vector of parameters but in this thesis a dimensionless single-parameter model is used to analyze algorithms.

4.1.4 Specific Point Spread Function Model

The joint autofocus and super-resolution algorithm presented in the next section can be applied to any form of PSF model as long as it is specified probabilistically. However, for the numerical simulations in this thesis a specific single-parameter PSF model is used. Luttrell [88] states that a SAR system undergoing anomalous motion can in first order be modelled as the defocusing of a simple linear imaging system and can be considered as being "a microwave version of an optical bench experiment using coherent illumination with the lens misplaced from its correct focus". This model may not strictly be accurate for SAR systems as the physics involved is slightly different – see Appendix C. Detailed models of the PSF for defocused SAR systems are given by Blacknell [7]. However, for the purposes of demonstration of the Bayesian technique, Luttrell's simple model is used here.

The PSF as a continuous function of cross-range position x and dimensionless focus parameter θ is given in [88] as

$$\begin{aligned} T(x,\theta) &= \frac{1}{2c}\int_{-c}^{+c} \exp(ikx + i\theta k^2 x^2)\, dk \\ &\approx \frac{1}{2c}\int_{-c}^{+c} \exp(ikx)(1 + i\theta k^2 x^2)\, dk \\ &= T_0(x) + \theta T_1(x), \end{aligned} \tag{4.1}$$

where

$$\begin{aligned} T_0(x) &\equiv \frac{\sin(cx)}{cx} \\ T_1(x) &\equiv i\left[cx\sin(cx) + 2\cos(cx) - \frac{2\sin(cx)}{cx}\right]. \end{aligned} \tag{4.2}$$

For use in the imaging model this function is sampled at regular positions along the x axis to give the PSF in vector form. This vector is then converted to an appropriately formatted Toeplitz convolution matrix that applies the effect of the PSF – see Appendix A. The total convolution matrix $\mathbf{T}(\theta)$ can be written as the linear combination of the nominal convolution matrix $\mathbf{T_0}$ and the error convolution matrix $\mathbf{T_1}$:

$$\mathbf{T}(\theta) = \mathbf{T_0} + \theta\mathbf{T_1}. \tag{4.3}$$

4.2 Joint Autofocus and Super-resolution

4.2.1 Monte Carlo Algorithm for Cross-section Recovery

The Bayesian approach to autofocus and super-resolution models uncertainty in both the cross section $\boldsymbol{\sigma}$ and focus parameters θ. This is described by the joint probability density $p(\boldsymbol{\sigma}, \theta|\mathbf{g})$ of the cross section and focus parameters, conditional on the image under consideration. Bayes' theorem gives the joint density as

$$p(\boldsymbol{\sigma}, \theta|\mathbf{g}) = \frac{p(\mathbf{g}|\boldsymbol{\sigma}, \theta)p(\boldsymbol{\sigma})p(\theta)}{p(\mathbf{g})}. \tag{4.4}$$

As before, we use the Metropolis-Hastings algorithm to generate samples of a probability distribution. For the case considered here the samples represent the joint probability density $p(\boldsymbol{\sigma}, \theta|\mathbf{g})$. The likelihood of the image is given by

$$\begin{aligned} p(\mathbf{g}|\boldsymbol{\sigma}, \theta) &= \int p(\mathbf{g}|\mathbf{f}, \boldsymbol{\sigma}, \theta)p(\mathbf{f}|\boldsymbol{\sigma}, \theta)\ d\mathbf{f} \\ &= \int p(\mathbf{g}|\mathbf{f}, \theta)p(\mathbf{f}|\boldsymbol{\sigma})\ d\mathbf{f}, \end{aligned} \tag{4.5}$$

where we have noted from Figure 4.1 that $\mathbf{g}$ does not depend directly on $\boldsymbol{\sigma}$, and $\mathbf{f}$ does not depend on θ. As with the perfectly focused case, using the scattering and imaging model in equations (3.1) and (3.2) this can be shown to be [88]

$$p(\mathbf{g}|\boldsymbol{\sigma}, \theta) = \frac{\exp(-\mathbf{g}^H\mathbf{M}^{-1}\mathbf{g})}{\det(\pi\mathbf{M})}, \tag{4.6}$$

where

$$\mathbf{M} \equiv \mathbf{T}\boldsymbol{\Sigma}\mathbf{T}^H + \mathbf{N}, \tag{4.7}$$

and $\boldsymbol{\Sigma} = diag(\boldsymbol{\sigma})$. However, in this case the matrix $\mathbf{T}$ depends on θ. As it is necessary only to know the shape of the distribution $p(\boldsymbol{\sigma}, \theta|\mathbf{g})$, there is no need

to calculate the normalizing factor $p(\mathbf{g})$. With choice of suitable priors $p(\boldsymbol{\sigma})$ and $p(\theta)$ for the cross section and focus parameters the quantity of interest is then

$$\pi(\boldsymbol{\sigma}, \theta|\mathbf{g}) = p(\mathbf{g}|\boldsymbol{\sigma}, \theta)p(\boldsymbol{\sigma})p(\theta). \tag{4.8}$$

At each iteration we update each element of $\boldsymbol{\sigma}$ in turn and then perform a θ update. During the $\boldsymbol{\sigma}$ update at the ith iteration, a proposed new sample for a single cross-section element is generated from a proposal distribution $q(\boldsymbol{\sigma}^{i+1}|\boldsymbol{\sigma}^i)$. The proposed sample is accepted with a probability $\alpha(\boldsymbol{\sigma}^i, \boldsymbol{\sigma}^{i+1})$, where

$$\alpha(\boldsymbol{\sigma}^i, \boldsymbol{\sigma}^{i+1}) = \min\left[\frac{\pi(\boldsymbol{\sigma}^{i+1}, \theta^i|\mathbf{g})q(\boldsymbol{\sigma}^i|\boldsymbol{\sigma}^{i+1})}{\pi(\boldsymbol{\sigma}^i, \theta^i|\mathbf{g})q(\boldsymbol{\sigma}^{i+1}|\boldsymbol{\sigma}^i)}, 1\right]. \tag{4.9}$$

The same process is then repeated for θ with

$$\alpha(\theta^i, \theta^{i+1}) = \min\left[\frac{\pi(\boldsymbol{\sigma}^{i+1}, \theta^{i+1}|g)q'(\theta^i|\theta^{i+1})}{\pi(\boldsymbol{\sigma}^{i+1}, \theta^i|g)q'(\theta^{i+1}|\theta^i)}, 1\right]. \tag{4.10}$$

Initial samples generated during the burn-in period are discarded and the remaining samples are distributed from $p(\boldsymbol{\sigma}, \theta|\mathbf{g})$ as required.

4.2.2 Results for a Low Signal-to-Noise Ratio

A demonstration of the Metropolis-Hastings algorithm for simultaneous RCS and focus parameter recovery is now given. The simulated scene used in this demonstration is identical to that introduced in Section 3.2.4 and shown in Figure 3.5a. The true focus parameter value was set to $\theta = 0.1$ and an SNR of 20 dB was used. The noisy, low-resolution image is shown in Figure 3.5b. The detailed processing parameters used in the M-H algorithm are shown in Table 4.1.

Figure 4.2 shows the sample time series for range index 7, which is the position of the first scatterer. The corresponding histogram is shown in Figure 4.3a. Compared to the low-SNR super-resolution-only histogram in Figure 3.17a, this distribution is very similar.

Figures 4.3b, 4.3c and 4.3d show the histograms for range indices 8, 9 and 11, respectively. These are also very similar to the equivalent low-SNR super-resolution-only histograms in Figures 3.17b, 3.17c and 3.17d. However, it is noticeable that in the histograms where autofocus is being used the distribution

Number of saved samples	10000
Burn-in period	0
De-correlation gap	10
Proposal distribution, $\boldsymbol{\sigma}$	Uniform random walk
Random walk step size, $\boldsymbol{\sigma}$	0.5
Prior distribution $p(\boldsymbol{\sigma})$	Independent negative exponential
Prior mean for $p(\boldsymbol{\sigma})$	1000
Sample start value for $\boldsymbol{\sigma}$	0
Proposal distribution, θ	Gaussian random walk
Random walk step size, θ	0.02
Prior distribution $p(\theta)$	Gaussian
Prior mean for $p(\theta)$	0
Prior s.d. for $p(\theta)$	0.2
Sample start value for θ	0

Table 4.1: Table of parameters used by the M-H algorithm for simultaneous RCS and focus parameter recovery.

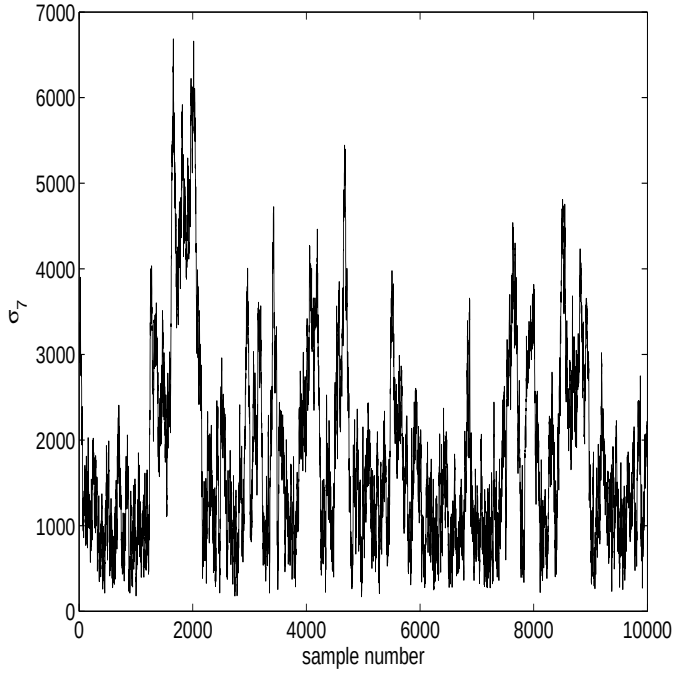
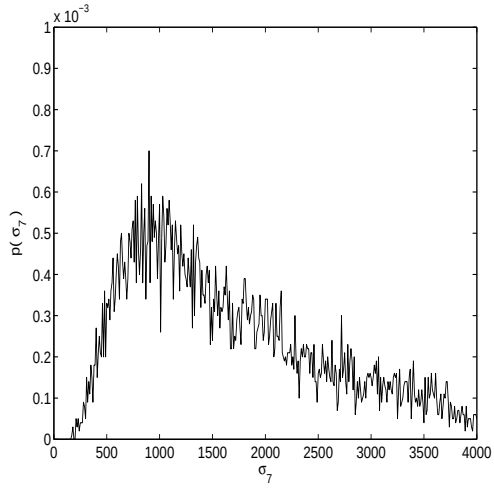


Figure 4.2: Time series of samples for range index 7 as generated by the M-H algorithm using a decorrelation gap of 10.

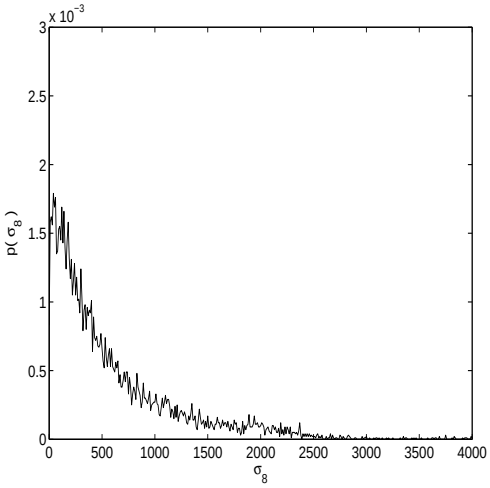
is slightly wider than those where the system is assumed to be completely focused. This is due to the additional uncertainty in the focus parameter θ.

The mean and MAP estimated range profiles are shown in Figure 4.4a. The MAP profile is very similar to the one obtained under perfect focus conditions, shown in Figure 3.18a. A detection algorithm using the MAP profile should be able to detect the scatterers without too high a false alarm rate. The mean profile is distorted by the effect of a skewed RCS distribution combined with high noise levels, as with the perfectly focused case. Although the two scatterers at positions 7 and 11 have large magnitudes, an addition phantom scatterer has appeared at position 5 in the autofocus case.

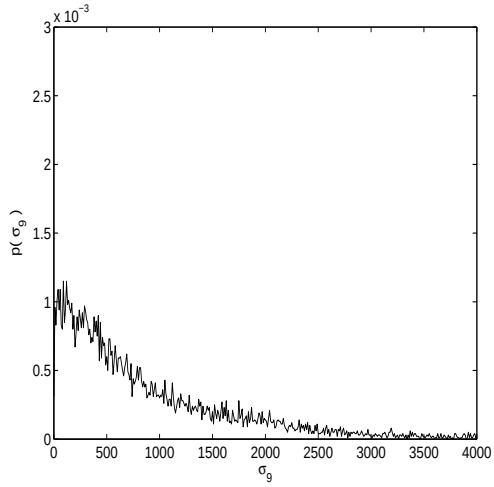
The standard deviation of the range profile is shown in Figure 4.4b along with the mean. The skewed distribution generally results in high standard deviations where the mean is high. This is particularly the case for the phantom scatterer at position 5, which has a very large uncertainty associated with it. This demonstrates one of the key advantages of analyzing the full RCS distribution under a Bayesian framework. If an algorithm estimated only the mean,



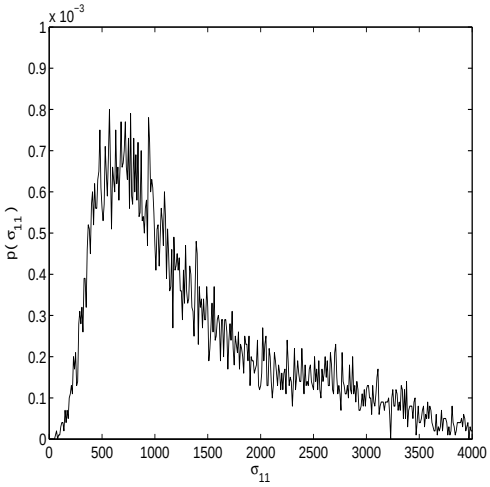
(a) Range Index 7



(b) Range Index 8

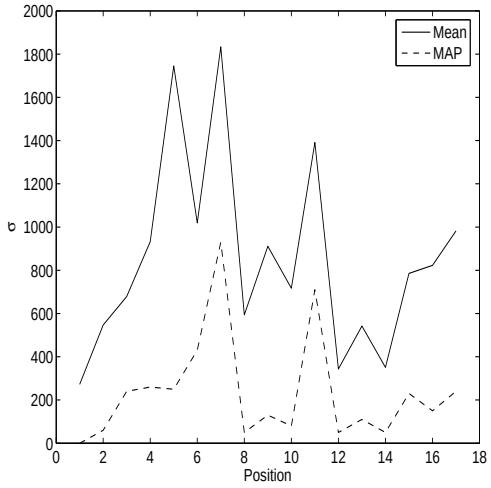


(c) Range Index 9

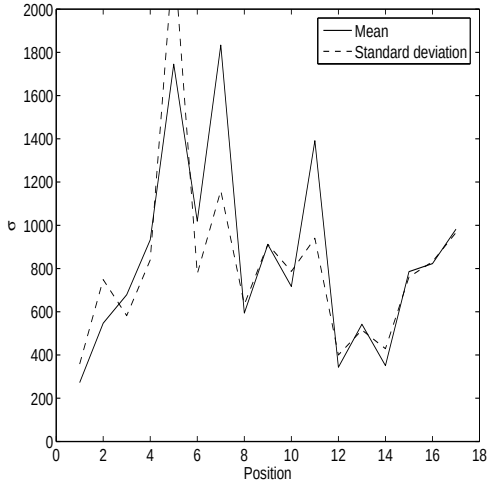


(d) Range Index 11

Figure 4.3: Histogram of samples generated by the M-H algorithm for various range indices.



(a) Mean and MAP



(b) Mean and standard deviation

Figure 4.4: Range profile representations calculated from samples.

then the phantom scatterer would have to be interpreted as a true scatterer. However, with the full distribution it can be seen that the mean of the distribution is far from its most likely value and this increases the likelihood of determining the phantom scatterer is false.

The RCS sample covariance matrix is shown in Figure 4.5. As with the perfectly focused case there is little correlation between any of the range indices. This is graphically demonstrated in Figure 4.6, where the sample RCS values at range indices 9 and 10 are plotted against each other. There is no discernable correlation.

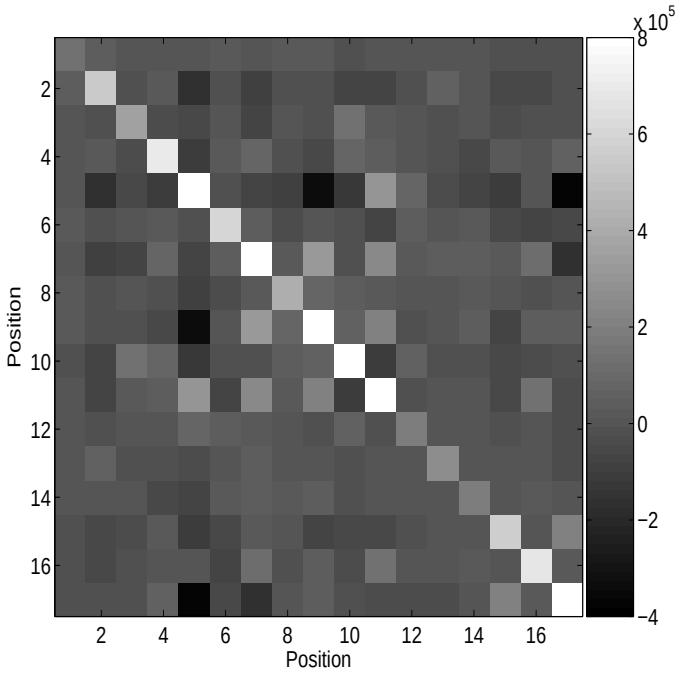


Figure 4.5: RCS sample covariance matrix.

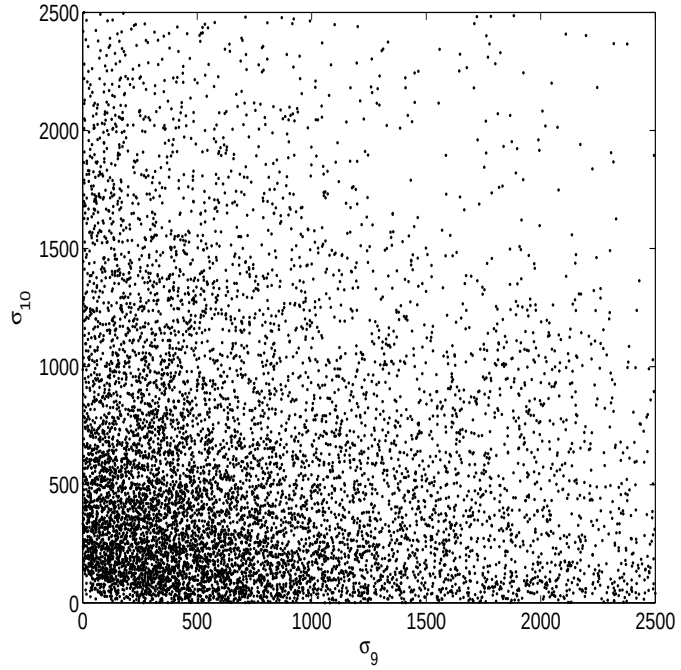


Figure 4.6: Distribution of samples for range indices 9 and 10.

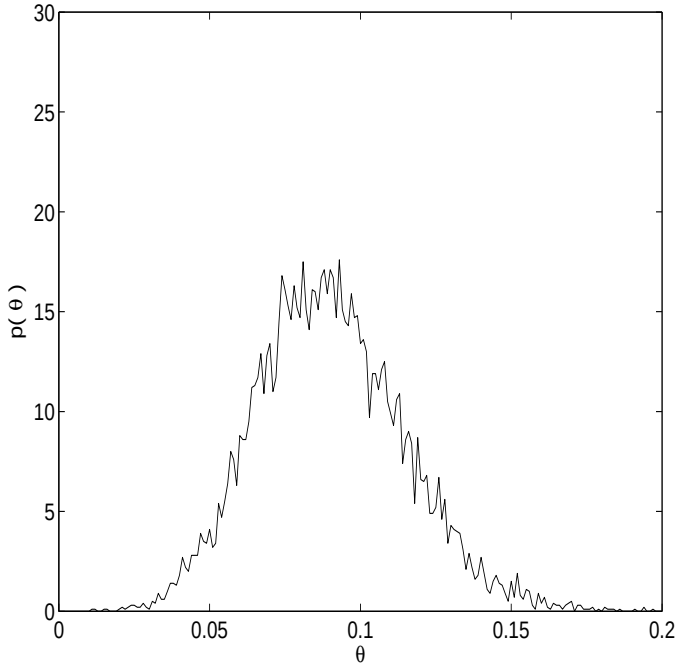


Figure 4.7: Distribution of dimensionless focus parameter θ.

So far, we have only discussed the distribution of the recovered underlying target RCS. That is what is of highest interest as the recovered RCS is a fundamental property of the target and is what is used in target recognition systems. However, it is also of interest to examine the distribution of the focus parameter θ to see how accurately it has been determined. The histogram for θ is shown in Figure 4.7. The distribution is approximately Gaussian shaped with a mean of 0.09 units and standard deviation 0.02. Recall that the true value for θ used in simulation was 0.1, which falls within the region of high probability. The small amount of bias in the distribution is due to noise. It has been stated by Luttrell [88] that a θ value of 0.1 is at the limit of the depth-of-focus criterion. Considering that the standard deviation of θ as determined by the Metropolis-Hastings algorithm is much less than this, it can be seen that it is indeed possible to determine the focus parameter to a greater accuracy than the depth-of-focus limit as suggested by Oliver [108].

4.3 Discussion

4.3.1 Complex-Field Recovery and Autofocus

In the previous chapter we saw that super-resolution RCS recovery was a more difficult problem than complex-field recovery both in terms of computational cost and the skewness of the posterior distribution. It might therefore seem desirable to examine the autofocus model using the complex-field representation of targets. However, the number of parameters in this model exceeds the number of data points available to estimate them. This is because for every complex pixel in the image there is a complex pixel in the underlying scene that needs to be estimated. The focus parameter adds another degree of freedom to the system, which potentially makes the problem intractable. It is possible that if there were a high degree of prior knowledge of either the scene or focus parameter the posterior distribution could be determined. However, a joint autofocus and field-recovery algorithm based on the Metropolis-Hastings algorithm failed to converge to a stable distribution when using the same prior knowledge as the other experiments reported in this thesis. The only way to improve the situation would be to obtain multiple images of the same scene from exactly the same aspect angle. While this could be possible from repeat passes of a SAR platform, the goal of this work is to perform super-resolution with a single image. It is therefore concluded that this type of algorithm is not useful in this context.

At first it might seem odd that complex-field recovery works better than cross-section recovery in the fully focused case, but when defocus is introduced complex-field recovery breaks down completely and cross-section recovery suffers only a small degradation in performance. This can be explained by the fact that there are fewer parameters to estimate in the cross-section recovery process because they are real valued rather than complex. Addition of the focus parameter does not therefore have a large influence on overall performance.

4.3.2 Further Work

The autofocus model used in this chapter was introduced by Luttrell [88], where it was stated that the problem of designing an algorithm that makes rapid progress towards the global maximum of $p(\boldsymbol{\sigma}, \theta|\mathbf{g})$ would be an extensive re-

search topic in it own right, requiring the topography of $p(\boldsymbol{\sigma}, \theta|\mathbf{g})$ to be investigated in detail. This investigation is essentially what has been done here and we have determined that the distribution is wide, skewed, and uncorrelated between the parameters. This analysis is a major extension of Luttrell's work and has made possible a potential link with MMSE super-resolution, via Bayesian arguments.

The model used by Luttrell may not entirely be accurate for SAR imaging. Appendix C demonstrates how a more accurate single-parameter radar-based PSF model differs from the model used by Luttrell. It would be interesting to test the Bayesian algorithm under the radar-based model to determine whether the standard deviation of the focus parameter matches that derived by Oliver [107, 108]. Even the radar-based single-parameter model may not be accurate enough to describe the radar PSF. It is likely that a realistic model would require a large number of parameters [10]. The Bayesian algorithm should be tested under such a model to determine the effect of additional parameters. However, if too many parameters are introduced then problems of under-determined systems such as those associated with joint autofocus and complex-field recovery might arise.

Overall, the problem of joint autofocus and super-resolution is difficult to solve. The problem is made less severe if the radar platform has a high-quality IMU and other electronic components such that the system's inherent point spread function is stable. Further alleviation of the problem should be possible through the use of a general purpose autofocus algorithm such as PGA, to produce the highest possible quality of images before super-resolution is attempted.

Chapter 5

A Bayesian Super-resolution Target-Recognition Framework

5.1 Introduction

When automatic target recognition systems employ multiple sensors, consideration must be given to how data from different sources is exploited. An ATR system usually requires access to a database of previously recorded or synthesized radar images for the targets of interest, or a database of features based on those images. Images used during the training phase of an ATR system might have a different resolution to those that are used during operation of the system. This could be the case, for example, where the training data consist of ISAR images formed from targets placed on a rotating turntable, and the images acquired in an operational environment are collected using an airborne platform. The need for an airborne platform to manoeuvre in operational scenarios limits the resolution that can be achieved. Higher resolutions are possible with the turntable measurements, which are recorded under more benign conditions.

As we saw in Section 2.4.5, when there is a difference in resolution between the test and training images the probability of correct classification is lower than when the resolutions are the same. It has also been shown elsewhere [105] that, in general, the higher the resolution of the data the higher the probability of correct classification. It is therefore desirable to test whether super-resolution can be used to match the resolution of the data gathered under operational conditions with that gathered during training, and improve the overall classification

performance.

A classification system that uses super-resolution will necessarily be required to pass data between various modules of the system. An amount of uncertainty is present in this data due to noise or other random effects. This uncertainty can either be ignored at each stage or information about it can be propagated between modules. Bayesian theory provides a consistent mechanism for manipulating probabilities assigned to data. This thesis therefore proposes to perform both super-resolution and classification within a Bayesian framework. The framework presented here was developed in conjunction with Copsey [27], who considers the issue of relocatable targets in addition to super-resolution and classification.

5.2 The Bayesian Framework

5.2.1 Classification Model

At the heart of the Bayesian framework is the relationship between variables for the different sensors. The autofocus and super-resolution model introduced in Section 4.1.3 gave the relationship between the radar cross section, complex scattered field, focus parameter and complex image for a single sensor. This model can be extended to include more than one sensor. The two-sensor case is shown in Figure 5.1. Extensions to a greater number of sensors are possible and repeat the same pattern as the two-sensor case. The underlying cross section of the target $\boldsymbol{\sigma}$ is constant but each time the target is imaged a different complex field $\mathbf{f}$ is scattered back towards the radar according to the scattering model in equation (3.1). The focus parameter θ for each sensor is different each time the target is imaged and depends on the relative motion between the radar and the target. The image $\mathbf{g}$ produced from a particular sensor depends on the scattered field and the point spread function for that sensor, which is determined by the focus parameter. In Figure 5.1, $\mathbf{g_1}$ represents a low-resolution test image that is required to be classified and $\mathbf{g_2}$ represents a high-resolution image of the sort used during training to form a database of target information.

A complete Bayesian solution to the classification problem would take the low-resolution test image $\mathbf{g_1}$ and perform simultaneous autofocus and super-resolution to obtain the distribution $p(\boldsymbol{\sigma}, \theta_1|\mathbf{g_1})$. The dependency of this func-

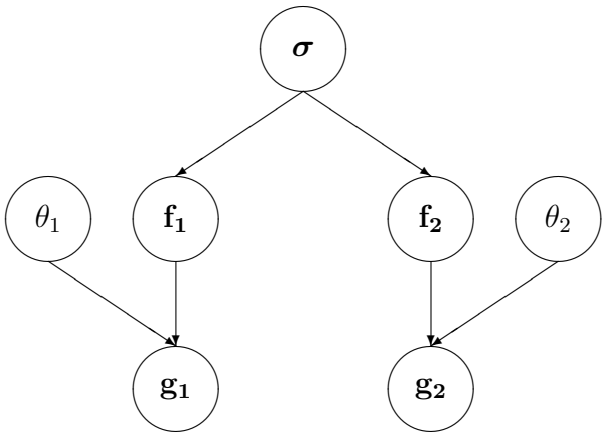


Figure 5.1: Scattering and imaging model for two sensors

tion on θ_1 can be eliminated by integrating the distribution over θ_1. The distribution $p(\boldsymbol{\sigma}|\mathbf{g_1})$ would then be passed through the scattering and imaging models using an assumed distribution for θ_2 to produce a new image distribution $p(\mathbf{g_2}|\mathbf{g_1})$, which represents images of the same resolution as those in the training database. The focus parameter θ_2 for training images should be fairly accurately known since training data is usually gathered in controlled circumstances. If θ_2 were required to be known more accurately then autofocus could be performed on the images in the training database to determine its distribution. In either case, the transformed image distribution $p(\mathbf{g_2}|\mathbf{g_1})$ would be classified against the training database. The output of a Bayesian classifier trained using the high-resolution data is a vector of class probabilities $p(\mathbf{c}|\mathbf{g_2})$. However, we require $p(\mathbf{c}|\mathbf{g_1})$. Formally, this is calculated as

$$p(\mathbf{c}|\mathbf{g_1}) = \int p(\mathbf{c}|\mathbf{g_2})p(\mathbf{g_2}|\mathbf{g_1})\ d\mathbf{g_2}. \tag{5.1}$$

The above classification procedure is theoretically ideal because $\boldsymbol{\sigma}$ is a fundamental target property under this model. If its distribution can accurately be estimated then the totality of our knowledge about the target is encapsulated in that distribution. However, as we saw in Section 3.4, the double inverse nature of determining $\boldsymbol{\sigma}$ makes this solution impractical. We therefore present a simplified model that requires no knowledge of target cross-section distributions. In this model, shown in Figure 5.2, the scattered field is assumed to be the same when all sensors image the target. This simplification ignores the scattering model and will underestimate uncertainty in target properties. This model is

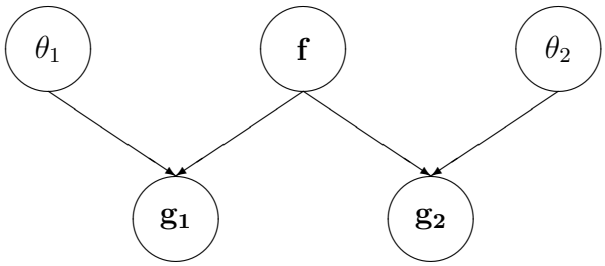


Figure 5.2: Simple imaging model for two sensors

equivalent to the Ricean model when a cross-section element has one dominant scatterer and no other weak scatterers. The classification procedure under this model would be to perform super-resolution and autofocus on the test image $\mathbf{g_1}$ to obtain the distribution $p(\mathbf{f}, \theta_1|\mathbf{g_1})$. After integrating over θ_1, the distribution $p(\mathbf{f}|\mathbf{g_1})$ would be passed through the imaging model for training sensors to construct an image distribution $p(\mathbf{g_2}|\mathbf{g_1})$. This would then be classified against the training database.

Even with the simplified model, problems arise due to defocusing of the system. As we saw in Section 4.3 the presence of unknown focus parameters in addition to the complex field makes the system under-determined. Therefore, in order for this classification procedure to be used, the issue of autofocus must be neglected for the models used here and it must be assumed that all images have had a focusing algorithm applied such that the data is already well-focused. The issue of whether it is possible to include other scattering and defocus models in a joint super-resolution and target recognition framework is addressed in Section 6.2.

5.2.2 Classification Procedure

The above classification model is theoretical in nature and a number of steps must be taken to ensure it can be implemented in a practical situation. The first step of determining $p(\mathbf{f}|\mathbf{g_1})$ is a super-resolution problem and has already been addressed in Chapters 3 and 4, where the distribution $p(\mathbf{f}|\mathbf{g_1})$ is represented by

a set of samples. In the case where an analytic solution is available, the mean and covariance matrix can be used to generate samples from the distribution. This is carried out using Cholesky decomposition of the covariance matrix [50].

To convert from $p(\mathbf{f}|\mathbf{g_1})$ to $p(\mathbf{g_2}|\mathbf{g_1})$, the imaging model for the second sensor is applied to each sample in $p(\mathbf{f}|\mathbf{g_1})$. In practice, this is the convolution of each super-resolved image $\mathbf{f}$ with an appropriate point spread function. Thus $p(\mathbf{g_2}|\mathbf{g_1})$ is represented by a set of images with a lower resolution than $\mathbf{f}$ but higher than $\mathbf{g_1}$.

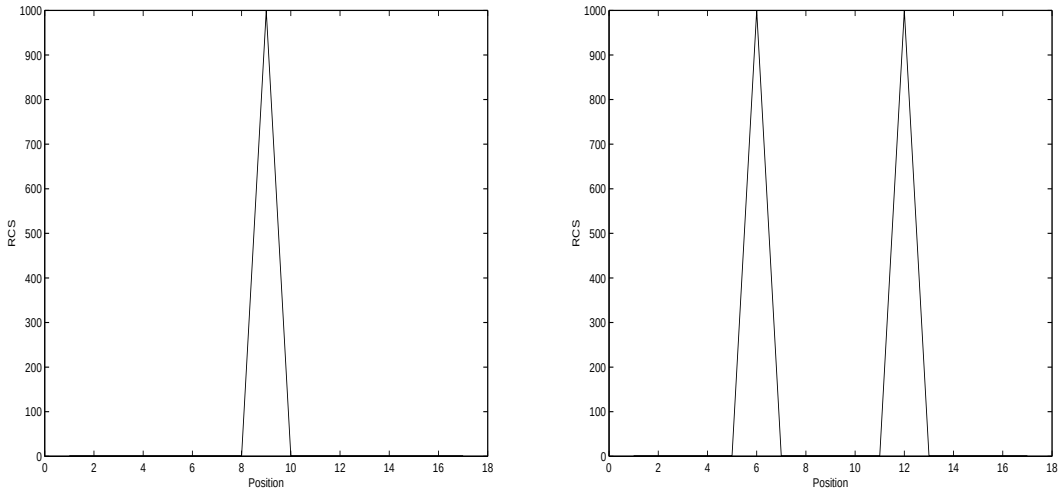
The final step in the overall classification procedure is to calculate $p(\mathbf{c}|\mathbf{g_1})$. This is done by taking each sample image in $p(\mathbf{g_2}|\mathbf{g_1})$ and determining $p(\mathbf{c}|\mathbf{g_2})$ using a standard Bayesian classifier. The mean value of $p(\mathbf{c}|\mathbf{g_2})$ calculated over all the images in $p(\mathbf{g_2}|\mathbf{g_1})$ gives the required vector of class probabilities $p(\mathbf{c}|\mathbf{g_1})$. The classifier selects the class with the highest probability. This determines which target is thought to be present in the low-resolution image $\mathbf{g_1}$.

5.3 Simulated 1D data

5.3.1 Experimental Setup

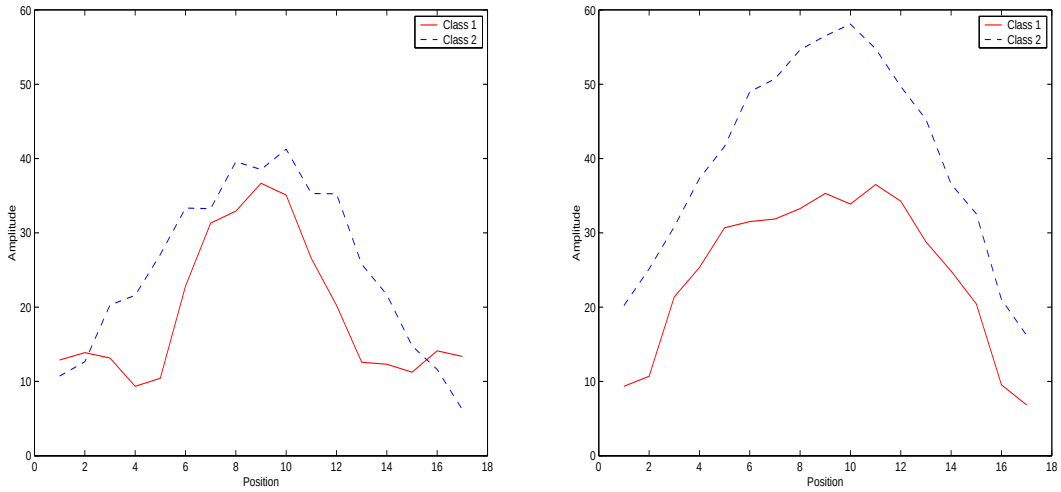
A simple two-class problem with one-dimensional data is used here to illustrate operation of the Bayesian super-resolution target-recognition framework. The two target classes are defined on a 17-element RCS range profile as displayed in Figure 5.3. The first class corresponds to a single point scatterer, while the second class corresponds to two scatterers.

The two underlying RCS profiles are used to generate complex-field profiles, in accordance with the scattering model introduced in Section 3.1.2. Simulated images are obtained by applying an appropriate sensor imaging model to each complex-field profile. The ratio of the operational sensor resolution to that of the training sensor is set to be 1.6. For both the training and operational sensors, independent Gaussian noise is added to the images at an SNR of 20 dB. Examples of the one-dimensional images for both classes and sensor types are shown in Figure 5.4. Note that for class 2 it is not immediately apparent from the images that two scatterers are present due to the wide PSF of both sensors relative to the separation between the two scatterers. This provides an opportunity to test the super-resolution algorithm within a target-recognition



(a) Class 1 (b) Class 2

Figure 5.3: Underlying radar cross sections



(a) Training sensor (b) Operational sensor

Figure 5.4: Magnitude of sensor images

framework. However, it should be noted that the scatterer separation and PSF widths in this simulation have been chosen to demonstrate the potential of the Bayesian approach rather than to accurately reflect real targets or training or operational conditions.

For each sensor the images from the two classes appear quite distinct, as can be seen in Figure 5.4. Thus a classifier trained using data from the training sensor should be able to successfully classify training sensor data from the two classes. Similarly, a classifier trained using data from the operational sensor should be able to separate operating sensor data into the two classes. However, for each class there is a considerable difference in the images between the two sensors. Visually, the operational sensor image from class 1 is more similar to the training sensor image from class 2, than it is to the training sensor image from class 1. This indicates that using the training sensor classifier directly on the operational sensor data will lead to poor classification performance for class 1. Therefore the Bayesian super-resolution approach is required.

A Bayesian classifier was trained using 100 high-resolution training sensor images from each class. The features consisted of the complex values of the image at each position. Preprocessing of the images prior to their use in the classifier was performed by taking the amplitudes of the complex images, and normalizing each complex input vector such that the average amplitude was unity. This is a commonly performed processing step used in target recognition systems. Note that for the operational sensor such preprocessing would remove some of the separability between the classes – see Figure 5.4.

The test data consisted of 100 low-resolution operational sensor images from each class. The Bayesian joint autofocus and super-resolution algorithm for complex-field recovery, based on the Metropolis-Hastings algorithm, was applied to each of these images. Parameters of the M-H algorithm are shown in Table 5.1. It is possible to perform autofocus and super-resolution for complex-field recovery in this case because the prior distribution for the focus parameter θ has been defined tightly around the true value of 0.1. Each super-resolved image had a PSF applied such that the resulting image had the same resolution as the training data. Each of those images was then classified using the Bayesian classifier.

In addition to the Bayesian super-resolution procedure, two baseline results

Number of saved samples	1000
Burn-in period	5000
De-correlation gap	10
Proposal distribution, $\mathbf{f}$	Gaussian random walk
Random walk step size, $\mathbf{f}$	0.5
Prior distribution $p(\mathbf{f})$	Independent Gaussian
Prior mean for $p(\mathbf{f})$	0
Prior s.d. for $p(\mathbf{f})$	10
Sample start value for $\mathbf{f}$	0
Proposal distribution, θ	Gaussian random walk
Random walk step size, θ	0.02
Prior distribution $p(\theta)$	Gaussian
Prior mean for $p(\theta)$	0.1
Prior s.d. for $p(\theta)$	0.03
Sample start value for θ	0.1

Table 5.1: Table of parameters used by the M-H algorithm for 1D super-resolution target-recognition experiments.

were obtained for comparison. The first baseline case used high-resolution data for both training and testing. This corresponds to a scenario where the operational sensor has the same resolution as the training sensor. The second baseline case trained the classifier using high-resolution data from the training sensor, and applied the classifier directly to the lower-resolution operational sensor data, without super-resolution.

5.3.2 Results

The set of results for baseline classification and super-resolution classification performance are shown in Table 5.2. The first baseline procedure shows that if it were possible to measure both training and test data with the same resolution, then the two classes can be classified correctly all the time. The second baseline results show that, when the resolution of the two sensors is different, ignoring the change in sensor resolution is undesirable. In particular, the observation

Method	Operational/Test Data	Class 1	Class 2	Average
Baseline 1: direct high/high comparison	High resolution	100%	100%	100%
Baseline 2: direct low/high comparison	Low resolution	3%	100%	51.5%
Bayesian super-resolution	Low-res. → super-res. → high-res.	97%	100%	98.5%

Table 5.2: Probability of correct classification for idealized targets. Percentages based on 100 test images. Training data was based on the high-resolution sensor.

that the operational sensor images from class 1 are more similar to the training sensor images from class 2 than those from class 1, is evidenced by the fact that only 3% of the class 1 images are correctly identified. In contrast, the proposed Bayesian super-resolution target-recognition procedure recovers much of the class separability that is present when using high-resolution data for both sensors. Thus, with appropriate sensor imaging models for this example, we have successfully demonstrated procedures that enable a classifier to be used in situations where the operating sensor resolution differs from the training data resolution. It should be emphasized, however, that the above classification performance is for idealized targets only, and in practical situations a much lower classification performance would be expected for both the baseline and Bayesian methods.

5.4 Measured 2D data

5.4.1 Experimental Setup

The previous section demonstrated the advantage of the Bayesian superresolution target-recognition procedure over standard target recognition under idealized conditions. However, in practice, targets are more complicated and we would like to classify two-dimensional images rather than one-dimensional range profiles. To test the Bayesian procedure under more realistic conditions the subset of the MSTAR data set analyzed in Section 2.4.5 has been used here as a basis for target images.

Four classification procedures have been used to test the various parts of the Bayesian framework. The first of these is a baseline procedure that defines an upper limit on performance. This procedure follows that described in Section 2.4.5, where targets are segmented from the background, features are extracted from the segmented target and a classifier is trained and tested using those features. In the experiments reported in this section noise was also added to each test image at a certain SNR prior to features being extracted. This is referred to as high-resolution test data and has a resolution of 0.3 m. The performance of the Bayesian and nearest-neighbour classifiers was recorded to allow a comparison between different classifier types.

The second classification procedure was the same as the first with the exception that the resolution of the test imagery was degraded from 0.3 m to 1.0 m before noise was added. This set of data is referred to as low-resolution data.

The third classification procedure tests the utility of the MMSE-T super-resolution algorithm. Ideally, the algorithm would be applied to each entire low-resolution image. The super-resolved image would then be classified as normal. This, however, is not possible due to the size of the images, which are typically 128x128 pixels. The algorithm runs too slowly on images of this size to be of practical use. Even if the algorithm were able to run more quickly, it is thought that as the image size increases, super-resolution performance decreases [43]. The solution to this was to extract a 41x41 pixel image centered on the brightest pixel in the original MSTAR image and perform super-resolution on this reduced-size image. For the vast majority of images examined, this window size was sufficient to contain all the target detail of interest. The procedure of performing super-resolution on sub-images of a larger MSTAR image has also been used in [42].

An additional problem with the MMSE-T algorithm was that it tended to over-super-resolve targets so that they were estimated to be constructed from several isolated scatterers. This effect was mitigated by applying a Hamming window to the data in the frequency domain to limit the resolution to 0.3 m, which is the original resolution of the MSTAR data. Once the super-resolved 41x41 pixel image with a 0.3 m resolution was formed, it was inserted back into the low-resolution image. This step is clearly sub-optimal since the rest of the image is not super-resolved. However, during initial testing, it was found that

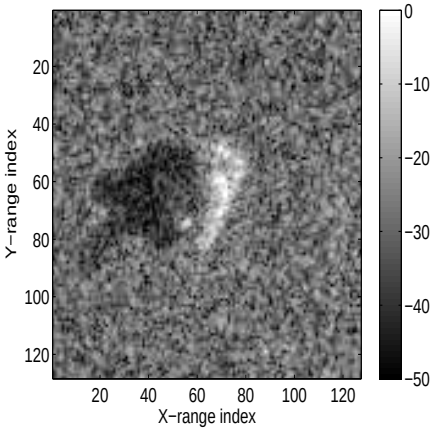
the feature extraction algorithm was able to extract features from the inserted super-resolved image to a reasonable degree of accuracy. Figure 5.5 illustrates the super-resolution target chip insertion process. Visually, the inserted super-resolved target chip in Figure 5.5d matches well with the original high-resolution image in Figure 5.5a. The feature extraction algorithm is not severely affected by the super-resolution insertion process because the target pixels are segmented from the surrounding image before the features are calculated, as was shown in Figure 2.23. Therefore it is unnecessary to perform super-resolution on other parts of the image. The features extracted from the modified super-resolution image were used as an input to the Bayesian and nearest-neighbour classifiers. This data is referred to as super-resolution data.

The final classification procedure tests whether it is better to use a Bayesian super-resolution algorithm than a non-Bayesian one. The same steps as the previous procedure were followed up to the production of a 41x41 pixel super-resolved image with its resolution adjusted to 0.3 m. At this point the covariance of the super-resolved image was calculated using the approximation in Section 3.5.2. The mean and covariance were used to generate 100 super-resolved images and were inserted back into the low-resolution image as before. Each of these images was classified as normal by the Bayesian and nearest-neighbour classifiers and the target with the highest probability of correct classification over all 100 realizations was selected as the target thought to be in the image. This data is referred to as Bayesian super-resolution data.

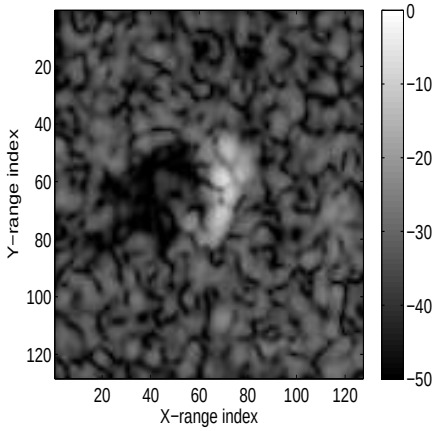
In all classification procedures the system was assumed to be perfectly focused such that the correct point spread function for both the training and operational sensors was used in all processing.

5.4.2 Results

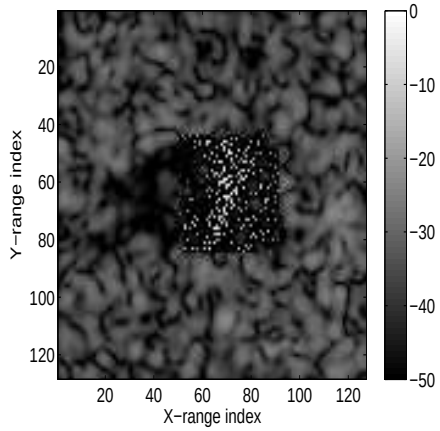
The experiments described in the previous section were run with SNRs ranging from 10 dB to 50 dB, using the length and width features only. The results are shown numerically in Table 5.3 and graphically in Figures 5.6 and 5.7. A detailed analysis of the results for a 50 dB SNR is now given, followed by a general analysis for the whole set of results. Throughout the discussion it should be noted that with five targets present in the database a classifier picking randomly between targets would achieve a 20% probability of correct classification.



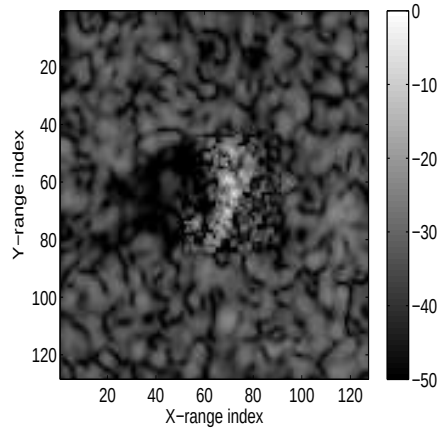
(a) High-resolution image.



(b) Low-resolution image.



(c) Super-resolution image.



(d) Super-resolution image with a Hamming window applied to target chip.

Figure 5.5: The process of inserting a super-resolved target chip into the background. (a) The original high-resolution image with a resolution of 0.3 m. (b) The low-resolution image with a resolution of 1.0 m. (c) The super-resolved target chip with a resolution better than 0.3 m has been inserted into the low-resolution image. (d) A Hamming window has been applied to the central target chip such that its resolution is 0.3 m.

The first line of Table 5.3 is directly comparable with line four of Table 2.2, which shows results for the case where no noise was added, but had all other parameters set to the same values. It can be seen that the addition of noise at 50 dB has led to a drop in classification performance of 5.1% points for the Bayesian classifier and 7.1% points for the nearest-neighbour classifier when using high-resolution imagery. The second line of Table 5.3 can be compared to the fifth of Table 2.2. Here, it is seen that the addition of noise has decreased performance by 5.6% points for the Bayesian classifier and 0.6% points for the nearest-neighbour classifier when using low-resolution imagery.

The third line of Table 5.3 shows a large reduction in performance when super-resolution is used compared to either the ideal case where high-resolution data is used or the operational scenario considered here where only low-resolution test data is available. This is seen as a disappointing result because the idea behind using super-resolution was that performance of super-resolved imagery is supposed to be better than that of low-resolution imagery.

There are a number of explanations as to why using super-resolved imagery results in a low performance. The first of these is that the process of chipping out a small target image from a larger image before performing super-resolution is sub-optimal. Although the majority of the energy in the large image is contained in the target chip there is a significant amount of energy in the rest of the image. It is possible that some of this energy could leak back into the target chip via sidelobes. This energy is not taken into account in the imaging model and could upset the super-resolution algorithm. This potential problem has previously been pointed out by Dickey *et al.* [43].

The second potential reason for low super-resolution performance is that although super-resolution improves the resolution of an image by design, a side effect is the addition of noise – the output SNR is always worse than the input SNR. It may be the case that any gain in performance due to resolution enhancement is more than offset by a reduction in performance due to the additional noise.

A third explanation is that the features used for classification are not robust with respect to super-resolution processing. In fact, during initial processing it was found that the output of the MMSE-T algorithm tended to produce images with a few isolated scatterers. Given that the true target images were

Test Data	SNR	Features	Classifier	
			Bayes	NN
High-resolution test data	50 dB	2,3	52.3%	39.9%
Low-resolution test data	50 dB	2,3	43.4%	37.0%
Super-resolution test data	50 dB	2,3	23.8%	29.7%
Bayesian super-resolution data	50 dB	2,3	30.6%	31.9%
High-resolution test data	40 dB	2,3	45.8%	41.6%
Low-resolution test data	40 dB	2,3	43.4%	35.9%
Super-resolution test data	40 dB	2,3	23.8%	25.0%
Bayesian super-resolution data	40 dB	2,3	32.3%	33.0%
High-resolution test data	30 dB	2,3	41.1%	42.8%
Low-resolution test data	30 dB	2,3	43.9%	39.1%
Super-resolution test data	30 dB	2,3	26.2%	29.8%
Bayesian super-resolution data	30 dB	2,3	32.3%	32.9%
High-resolution test data	20 dB	2,3	35.1%	35.7%
Low-resolution test data	20 dB	2,3	42.8%	35.9%
Super-resolution test data	20 dB	2,3	24.4%	27.4%
Bayesian super-resolution data	20 dB	2,3	30.7%	32.2%
High-resolution test data	10 dB	2,3	14.3%	16.7%
Low-resolution test data	10 dB	2,3	15.6%	24.1%
Super-resolution test data	10 dB	2,3	26.1%	26.8%
Bayesian super-resolution data	10 dB	2,3	30.7%	30.1%

Table 5.3: Probability of correct classification using the MSTAR data set, with extra noise added at various SNRs. Percentages based on 167 test images of targets. Features 2 and 3 refer to target length and target width, respectively. The feature database was generated using high-resolution training data gathered at an elevation of 17^o. All test features are based on data gathered at an elevation of 15^o.

generally smoother than the MMSE-T output it may seem that the algorithm was missing important detail. However, when the super-resolved images were passed through the imaging model and compared to the original "measured" low-resolution images it was found that the difference between the two was commensurate with the noise level. The isolated scatterer representation was therefore a reasonable description of the target. However, the effect of having isolated scatterers had a catastrophic effect on the feature extraction algorithm. When attempting to segment the target from the background the algorithm would select the brightest scatterer and its close neighbourhood only, missing out all the other major scatterers. This fundamentally altered all the feature values enough to make them useless. This was the motivation of applying the Hamming window to the data to restore the resolution to 0.3 m. The Hamming-processed images tended to connect the scatterers enough for the segmentation algorithm to appear to work correctly. However, it is possible that feature values were still altered enough to result in the degradation of classification performance. This effect could potentially be mitigated either by finding features robust to super-resolution processing or by providing an alternative to Hamming-window processing that does not severely alter feature values.

The fourth line of Table 5.3 shows the performance of the Bayesian super-resolution classification process. It is seen that, in comparison to standard super-resolution, Bayesian super-resolution has increased performance by 6.8% points for the Bayesian classifier and 2.2% points for the nearest neighbour classifier. This result justifies use of Bayesian super-resolution – by taking into account uncertainty in the super-resolved image classification performance is improved. This happens because whenever a large spike is present in the super-resolved image, there is a large uncertainty associated with it. This means that if the spike is related to noise, any detrimental effect on classification performance is mitigated. If the spike corresponds to a true target property then the fact that it is present helps classification performance. It should be noted, however, that even though classification performance using Bayesian super-resolution is better than standard super-resolution it is worse than simply using a low-resolution image. The comments as to why standard super-resolution performance is low also apply to Bayesian super-resolution.

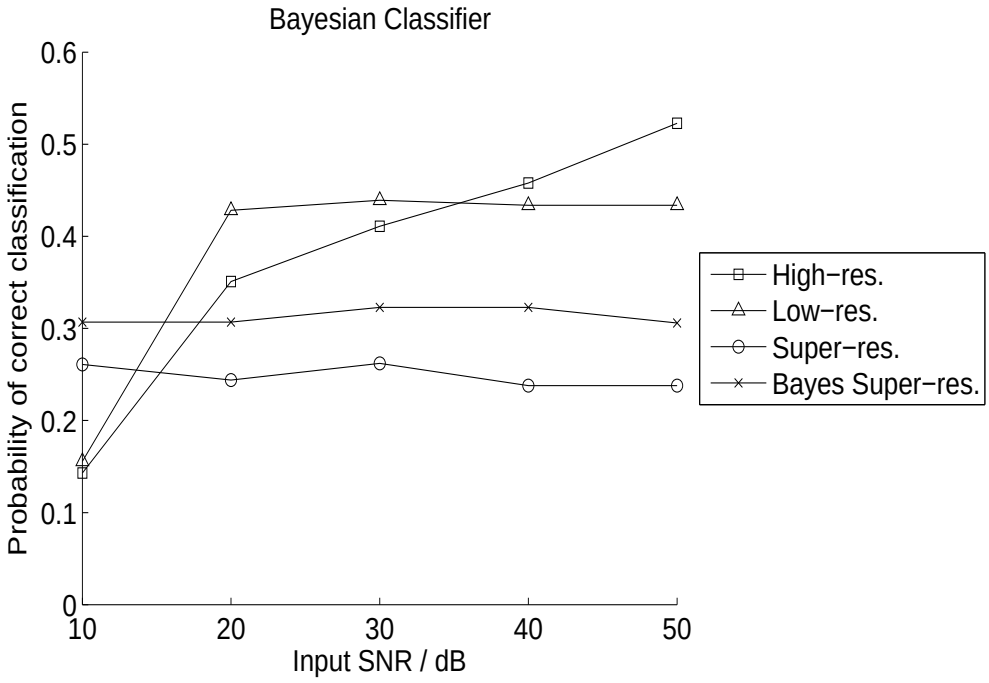


Figure 5.6: Probability of correct classification for the Bayesian classifier.

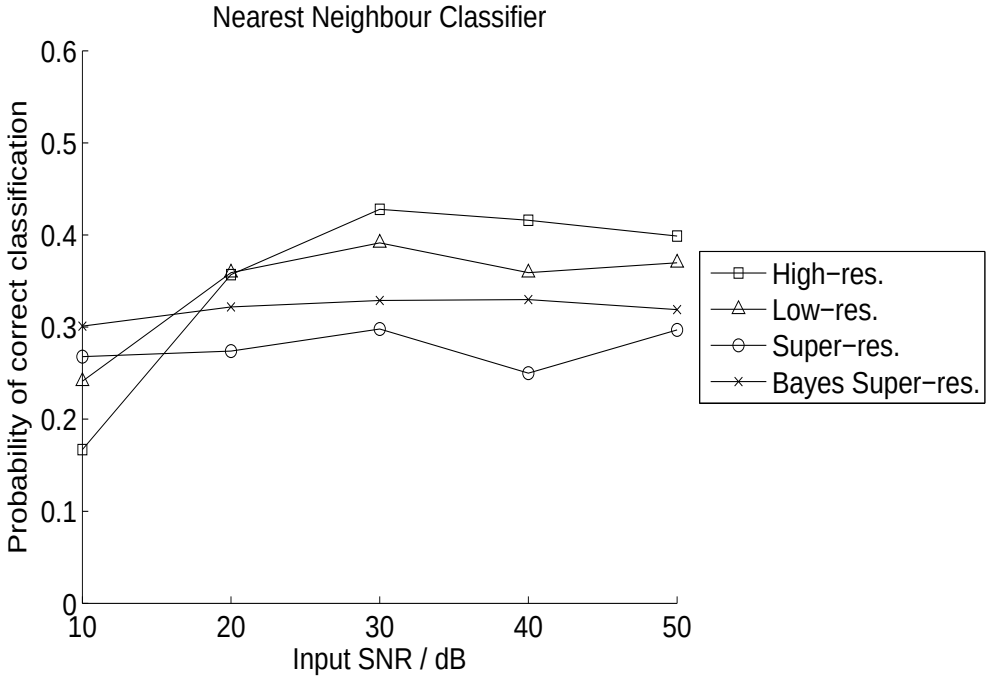


Figure 5.7: Probability of correct classification for the nearest-neighbour classifier.

The effect of SNR on the Bayesian classifier is shown graphically in Figure 5.6. It can be seen that, when using high-resolution data, performance drops smoothly as the SNR is reduced, with a significant drop at an SNR of 10 dB. The classifier has slightly different behaviour whilst using low-resolution data. The probability of correct classification is fairly constant until an SNR of 10 dB is reached, where a sharp drop in performance is also present. Classification performance using both standard and Bayesian super-resolution data is fairly constant over the range of SNRs tested, with Bayesian super-resolution always having the better performance of the two.

Performance as a function of SNR is shown for the nearest-neighbour classifier in Figure 5.7. The probability of correct classification is relatively stable over an SNR range of 20 dB to 50 dB for all resolutions but, as with the Bayesian classifier, classification performance significantly drops at an SNR of 10 dB when using high-resolution or low-resolution data.

It may initially seem surprising that classification performance does not vary more dramatically with SNR than it does. However, at high SNRs the addition of more noise does not have a large effect because the noise power is such a small fraction of the target power. At an SNR of 10 dB, the noise changes target images enough to significantly affect feature extraction and hence classification performance. The probability of correct classification for both standard and Bayesian super-resolution is fairly constant because the noise introduced by super-resolution processing is much greater than the thermal noise modelled in the system.

Chapter 6

Conclusions

6.1 Conclusions

The goal of this thesis has been to formulate methods that facilitate automatic target recognition using images generated from different radar sensors. The specific problem studied is the situation where the resolutions of the two sensors are different. It is found that a difference in resolution between the two systems causes a reduction in target recognition performance. It has been proposed, firstly, to use super-resolution to match the resolution from the two sensors in order to recover recognition performance, and, secondly, to combine information from the two sensors in a Bayesian framework.

Chapter 2 showed that standard super-resolution algorithms are able to improve the resolution of images. However, performance of the algorithms is limited by noise and the extent to which the point spread function is known. As noise levels increase, the algorithms fail in different ways. The matrix inverse technique tends to produce large spikes in the output because it over-fits the noisy data. The SVD algorithm is more stable with respect to noise and the resolution improvement gracefully degrades as fewer singular values are used in the algorithm. The MMSE-T algorithm appears to be the technique most robust with respect to noise, retaining a good performance at low SNRs by not generating a large number of spurious spikes or degrading resolution.

One problem with standard super-resolution techniques is the generation of spurious artefacts in the data, especially at low SNRs. When viewing the super-resolution output it is not possible to tell whether spikes in the data relate to

noise or are genuine scatterers from targets of interest. Chapter 3 presented a new Bayesian algorithm that calculates the full probability distribution of values each pixel can take. If a spike is due to noise then the Bayesian algorithm will determine that there is a large uncertainty associated with this spike and this additional information can be used when interpreting the image.

The Bayesian algorithm has been developed for both the inverse complex-field problem and the inverse cross-section problem. It was found that an analytic solution was possible in the inverse complex-field problem when noise follows a Gaussian model. The solution is itself Gaussian and can thus be represented by a mean vector and a covariance matrix. The Bayesian solution to the inverse cross-section problem was not able to be solved analytically so a numerical Monte Carlo algorithm based on the Metropolis-Hastings algorithm was used. It was found that the posterior distribution of target RCS had a large variance, was skewed and was uncorrelated between target pixels. This led to a re-examination of the Gaussian scattering model and it was conjectured that this model is inappropriate for high-resolution representations of targets as the central limiting theorem does not apply in the way it does for low-resolution data.

The computational cost of the Bayesian algorithm was analyzed and it was found to be proportional to the fourth power of the number of pixels being processed. Although the algorithm is able to be applied to small one-dimensional profiles it was too slow to be used with two-dimensional images of practical size. To alleviate the computational burden it was proposed to use the MMSE-T algorithm as a basis for an approximate analytic solution to complex-field recovery. In this approximation the mean of the posterior distribution was given by the MMSE-T output and the covariance matrix was a function of the mean and other image parameters.

In coherent radar processing there are a number of effects that can alter the shape of the point spread function. These are collectively known as defocus effects. The primary cause of defocus is relative motion between the radar platform and targets that has not been taken into account during processing. For an airborne radar platform this arises from deviations of the aircraft from a straight-line trajectory. The change in point spread function has a detrimental effect on standard super-resolution algorithms because they rely on knowledge

of the PSF.

In Chapter 4, a new Bayesian algorithm was developed that simultaneously estimates the high-resolution RCS representation of targets, and the focus parameter in a single-parameter focus model. This algorithm takes into account uncertainty in the focus parameter when estimating the high-resolution scene and thus offers an advantage over the situation in standard super-resolution where changes in the PSF are usually ignored. The posterior RCS distribution had a similar shape to that generated in the perfectly focused case but had a slightly larger variance and was a little more skewed.

It was found that a version of the Bayesian algorithm that attempts to estimate the complex field as well as the focus parameter failed to converge when prior knowledge about the focus parameter was diffuse. This is because the system is under-determined. The joint autofocus and complex-field super-resolution algorithm only worked when the prior distribution of the focus parameter was defined in a tight region about the true value – a situation unlikely to be possible in practice.

Although the Bayesian autofocus and super-resolution algorithm was shown to work in principle for RCS recovery, in realistic situations it is likely that the detailed nature of defocus is determined by more than one parameter, possible a large number of parameters. The Bayesian algorithm could be extended to deal with multiple parameters but the algorithm would take longer to run and care must be taken otherwise the system could become under-determined for RCS recovery as well as complex-field recovery.

Chapter 5 outlined a complete Bayesian super-resolution target-recognition framework for the case where the resolution of imagery measured in operational situations differed from that used in the training phase of the system. A number of approximations were required to be made for the system to be practically implementable. It was shown that use of a Bayesian super-resolution algorithm gave a higher probability of correct classification than using standard super-resolution. However, classification performance using super-resolution data was worse than using low-resolution data, for both types of super-resolution algorithm. Three reasons why this might be the case are that the approximations required to make operation of the Bayesian framework implementable were too severe, the advantages of resolution improvement were outweighed by the noise

introduced during super-resolution processing, or that the features used for target recognition were not robust to super-resolution processing.

There are three reasons why the Bayesian super-resolution target-recognition framework as presented here would not be worth implementing in a real radar system, without further modification. The first of these has just been mentioned – namely that super-resolution has degraded classification performance under the non-parametric model. The second reason is that the results based on measured data, presented in Chapter 5, did not analyze defocusing effects. It is likely that simple defocus would further degrade classification performance and, even if it were taken into account, the single-parameter focus model is probably insufficient to capture all significant effects. The third reason is that the computational burden of even the approximate Bayesian solution is high. The computation time is proportional to the third power of the number of pixels used – if an image measures $N \times N$ pixels then the computation time is proportional to N^6. Even with advances in processor power the ability to process large images would remain beyond reach for the foreseeable future.

Although the Bayesian super-resolution target-recognition framework as presented in this thesis appears to be unsuccessful in solving the multi-resolution target-recognition problem there are a number of modifications to it that could enable it to work. These possibilities are outlined in the next section where recommendations for further work are given.

6.2 Recommendations for Further Work

One of the reasons given for the relatively low classification performance when using super-resolution images was that feature values are not robust with respect to super-resolution processing. This is a major obstacle to successful execution of the Bayesian super-resolution target recognition framework. However, only five features were tested and is possible that other features could be more robust. The target recognition literature has many examples of different features. These should be analyzed to see how they are affected by changes in resolution – both for resolution degradation and super-resolution. One feature that may be of use would be the RCS coefficient of variation. This is the standard deviation of a target's distributed RCS, $\mathrm{Var}[\sigma^0(x,y)]^{1/2}$, divided by its mean

RCS, $\mathrm{E}[\sigma^0(x, y)]$. This feature is not affected by image amplitude scaling, and could provide a certain amount of robustness when changes in resolution cause changes overall image amplitude. A separate possibility is combining features from both low-resolution and super-resolution images. The low-resolution images may be better at determining large-scale features such as those related to target shape, and the super-resolution images better at determining fine-scale detail such as the position of closely placed scatterers.

Another problem with the super-resolution algorithms presented in this thesis is the computational burden associated with them. This arises from the need either to calculate the inverse of large matrices or to multiply large matrices together. It is shown in Appendix A that the convolution matrix is Toeplitz-block Toeplitz. The repetitive nature of this type of matrix means that processing advantages are possible if the matrix structure is taken into account [65,153]. This form of processing should be investigated to determine whether the algorithms can be made to run significantly faster than using a standard matrix multiplication implementation. Another way of potentially speeding up algorithms is through the use of FFTs. Convolution can be performed as a multiplication in the Fourier domain, which is faster than a straight-forward implementation of convolution in the spatial domain. A number of super-resolution algorithms do use Fourier processing – see [122, 139], for example. However, some work is necessary to determine how the Fourier representation of signals would fit into the Bayesian framework.

A third major problem with the super-resolution algorithms presented here is that they are based on the low-level continuum scattering model [90,140]. Under this model the amplitude of every pixel in the super-resolved image is calculated. This means a large number of parameters must be estimated – often as many parameters as there are data points in the input image. The addition of focus parameters makes the system under-determined for autofocus and complex-field recovery. A solution to this is to use the high-level scattering centre model [90, 140]. Under this model the scene can be described by a small number of point scatterers. This has two advantages. The first is that the number of parameters required to be estimated is dramatically reduced, especially for large images. This would allow more complex focusing models to be used and each parameter should be able to be estimated more accurately. The second

advantage is that the computational burden of algorithms using the scattering centre model depends on the number scatterers in the scene rather than the size of the scene image. Therefore it is quicker to process large images using the scattering centre model than the continuum model as long as the whole image is not completely filled with objects that must be modelled.

The disadvantage of the scattering centre model is that the number of scatterers present must be estimated for each image – an order selection problem. Since the number of scatterers is unknown in advance this would result in a variable number of dimensions under the Bayesian framework. This requires extra work to keep track of the dimensions and is the reason the scattering centre model has not been analyzed in this thesis. However, several papers mentioned in the literature survey use the reversible jump Markov chain Monte Carlo (RJMCMC) method to estimate model order and it seems that this approach is worth pursuing further in conjunction with the target-recognition framework.

It has been assumed throughout this thesis that a Gaussian scattering model applies to each pixel value. This model is valid when there are a large number of scatterers in a single resolution cell and the central limit theorem applies. However, at very high resolutions, there may only be one dominant scatterer in each resolution cell. In this case the central limit theorem no longer applies and some model other than Gaussian scattering ought to be used. This issue should be investigated further to determine a more accurate scattering model using measured data. This model should then be tested in the Bayesian super-resolution target-recognition framework to see whether recognition performance can be improved under this model.

Another area where the modeling could be more accurate is the focus model. The work in this thesis assumed a single-parameter was sufficient to describe focusing effects. Simulations were based on Luttrell's approximate model [88]. A more accurate model derived from first principles of radar signal processing is presented in Appendix C. In that model the focus parameter can be related directly to uncompensated cross-track acceleration. However, in practice, a number of other defocus effects are likely to be present and a multi-parameter model should be investigated.

Under the imaging model used for this thesis the thermal noise was uncorrelated and independent of the point spread function. This model is used by

many other authors in super-resolution or image restoration work. However, in a radar receiver, noise is generated in the front-end electronics before application of the matched filter. This means the noise is correlated with the filter weights. The detailed effect of correlated noise on super-resolution algorithms is unknown but Blunt has partially addressed this issue in his version of MMSE super-resolution [11]. It has also been assumed that the scattering model for speckle is independent of the PSF and can effectively be modelled as multiplicative noise. However Kuan *et al.* [76] state that speckle is spatially correlated and the correlation function depends on the coherent point spread function of the imaging system and the original image intensity. Correlated thermal noise and speckle should be analyzed under the Bayesian super-resolution target-recognition framework to determine whether they have an effect on target recognition performance and, if so, they should be included future models.

The presence of correlations in the image can also be attributed, to some extent, to properties of the target, as opposed to the point spread function. An example of using spatial correlations with the MMSE super-resolution technique is given in [22]. In a more general Bayesian analysis, the scene can be modelled by a Markov random field (MRF), in which each pixel depends on adjacent pixels in a defined neighbourhood [51]. It is required to specify the form of prior information for the MRF. One possibility is to use a Gaussian MRF that directly defines the correlation between adjacent pixels [148]. Another possibility is to use a Gibbs distribution, which utilizes a potential function that is a function of image gradients [58]. Three possibilities for the potential function are the hyper-surface convex function [66], the Huber function [114], or a log-cosh function [58]. All these functions have quadratic behaviour near zero but are linear further out. This type of prior can be preferable to Gaussian priors because image edges are not penalized so severely. Example applications of MRFs applied to radar data can be found in [36] and [131]. These techniques should be considered for further development in conjunction with Bayesian super-resolution analysis.

Finally, the performance of super-resolution algorithms can be measured in a number of different ways. Two measures used in this thesis are output SNR and the effect on classification performance. Other super-resolution performance metrics have been proposed, such as super-resolution gain [43], the

ability to segment an image into target and background regions [42], and Blacknell's heuristic performance metric based on the position and amplitude of recovered scatterers [7]. There seems to be no consensus as to which is the best performance metric. Indeed, different metrics might be useful for different applications. The MMSE super-resolution algorithm is derived by attempting to maximize the output SNR metric. Other super-resolution algorithms could be designed to maximize performance on other metrics. It would be especially interesting to see whether an algorithm specifically designed to maximize classification performance could be implemented. Such an algorithm would be a good candidate for the Bayesian super-resolution target-recognition framework presented in this thesis.

Appendices

Appendix A

The Point Spread Function Matrix is Toeplitz

A.1 One-dimensional PSF

Let the sampled one-dimensional point spread function be defined by

$$\mathbf{h} = [h_1, h_2, \ldots, h_c, \ldots, h_n]^T, \tag{A.1}$$

where c is the centre co-ordinate. The one-dimensional image $\mathbf{g}$ is the convolution of $\mathbf{h}$ and the one-dimensional scattered field vector $\mathbf{f}$. This convolution may be written in the form

$$\mathbf{g} = \mathbf{T}\mathbf{f}, \tag{A.2}$$

where

$$\mathbf{T} = \begin{pmatrix} h_c & h_{c-1} & \ldots & h_2 & h_1 & 0 & \ldots & 0 & 0 \\ h_{c+1} & h_c & \ldots & h_3 & h_2 & h_1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ h_{n-1} & h_{n-2} & \ldots & h_c & h_{c-1} & h_{c-2} & \ldots & h_1 & 0 \\ h_n & h_{n-1} & \ldots & h_{c+1} & h_c & h_{c-1} & \ldots & h_2 & h_1 \\ 0 & h_n & \ldots & h_{c+2} & h_{c+1} & h_c & \ldots & h_3 & h_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & h_n & h_{n-1} & h_{n-2} & \ldots & h_c & h_{c-1} \\ 0 & 0 & \ldots & 0 & h_n & h_{n-1} & \ldots & h_{c+1} & h_c \end{pmatrix}. \tag{A.3}$$

The structure of $\mathbf{T}$ is such that its ijth element T_{ij} is a function of only $(i-j)$ and thus has identical elements along its main diagonal and sub-diagonals. This

is the definition of a Toeplitz matrix [153]. The special structure of Toeplitz matrices allows savings in memory storage and processing requirements. The work in this thesis has used truncated convolution where the scene is assumed to be zero-valued outside the analysis region. Sometimes the PSF is assumed to "wrap around" from one end of the scene to the other. If this were the case the top right values of the $\mathbf{T}$ matrix would have values $[\ldots, h_2, h_1, h_n, h_{n-1}, \ldots]^T$ and the matrix would be a special kind of Toeplitz matrix known as a circulant matrix. In that case further processing advantages would be possible via the use of an FFT [65].

A.2 Two-dimensional PSF

Application of the two-dimensional point spread function is somewhat more complicated than that for one-dimensional data. Let the two-dimensional scattered field be defined by

$$\mathbf{F} = \begin{pmatrix} f_{11} & f_{12} & \ldots & f_{1M} \\ f_{21} & f_{22} & \ldots & f_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ f_{N1} & f_{N2} & \ldots & f_{NM} \end{pmatrix}. \tag{A.4}$$

This is converted to vector form by scanning along the columns of $\mathbf{F}$ to form

$$\mathbf{f} = [f_{11}, \ldots, f_{N1}, f_{12}, \ldots, f_{N2}, \ldots, f_{1M}, \ldots, f_{NM}]^T. \tag{A.5}$$

The two-dimensional image $\mathbf{G}$ and vector form $\mathbf{g}$ are defined in a similar manner to $\mathbf{F}$ and $\mathbf{f}$, respectively. The two-dimensional PSF with centre co-ordinate (d, c) is defined by

$$\mathbf{H} = \begin{pmatrix} h_{1,1} & h_{1,2} & \ldots & h_{1,c} & \ldots & h_{1,M-1} & h_{1,M} \\ h_{2,1} & h_{2,2} & \ldots & h_{2,c} & \ldots & h_{2,M-1} & h_{2,M} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ h_{d,1} & h_{d,2} & \ldots & h_{d,c} & \ldots & h_{d,M-1} & h_{d,M} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ h_{N-1,1} & h_{N-1,2} & \ldots & h_{N-1,c} & \ldots & h_{N-1,M-1} & h_{N-1,M} \\ h_{N,1} & h_{N,2} & \ldots & h_{N,c} & \ldots & h_{N,M-1} & h_{N,M} \end{pmatrix}. \tag{A.6}$$

If $\mathbf{g} = \mathbf{Tf}$, the $MN \times MN$ two-dimensional convolution matrix $\mathbf{T}$ is given by

$$\mathbf{T} = \begin{pmatrix} \mathbf{T_c} & \mathbf{T_{c-1}} & \dots & \mathbf{T_2} & \mathbf{T_1} & 0 & \dots & 0 & 0 \\ \mathbf{T_{c+1}} & \mathbf{T_c} & \dots & \mathbf{T_3} & \mathbf{T_2} & \mathbf{T_1} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{T_{M-1}} & \mathbf{T_{M-2}} & \dots & \mathbf{T_c} & \mathbf{T_{c-1}} & \mathbf{T_{c-2}} & \dots & \mathbf{T_1} & 0 \\ \mathbf{T_M} & \mathbf{T_{M-1}} & \dots & \mathbf{T_{c+1}} & \mathbf{T_c} & \mathbf{T_{c-1}} & \dots & \mathbf{T_2} & \mathbf{T_1} \\ 0 & \mathbf{T_M} & \dots & \mathbf{T_{c+2}} & \mathbf{T_{c+1}} & \mathbf{T_c} & \dots & \mathbf{T_3} & \mathbf{T_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{T_M} & \mathbf{T_{M-1}} & \mathbf{T_{M-2}} & \dots & \mathbf{T_c} & \mathbf{T_{c-1}} \\ 0 & 0 & \dots & 0 & \mathbf{T_M} & \mathbf{T_{M-1}} & \dots & \mathbf{T_{c+1}} & \mathbf{T_c} \end{pmatrix}, \tag{A.7}$$

where each $N \times N$ sub-matrix $\mathbf{T_i}$ is the one-dimensional convolution matrix determined from the ith column of $\mathbf{H}$. The sub-matrices $\mathbf{T_i}$ form a Toeplitz structure meaning $\mathbf{T}$ is block Toeplitz. Furthermore, since each sub-matrix is also individually a Toeplitz matrix, the overall convolution matrix $\mathbf{T}$ is Toeplitz-block Toeplitz [153].

Appendix B

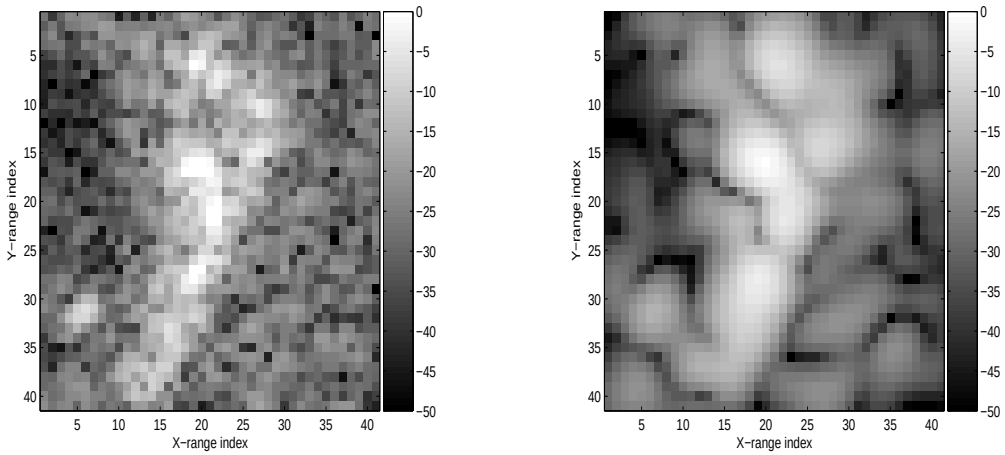
Choosing the Singular Value Threshold

B.1 Experimental Setup

Performance of the SVD super-resolution algorithm is analyzed here in terms of the input and output SNRs as defined in Section 2.2. The high-resolution scene $\mathbf{f}$ was taken to be a 41x41 pixel image of a T-72 tank taken from the MSTAR data set. The point spread function is based on the FFT of a Hamming window as described in Section 2.4.5. The high-resolution and low-resolution images relating to this setup are shown in Figure B.1.

The low-resolution image had noise added to it with SNRs ranging from 0 dB to 50 dB. At each SNR ten images were generated with different noise realizations, the SVD algorithm was applied, and the mean output SNR was measured. Five versions of the SVD algorithm were tested, each of which used a different threshold to determine the number of singular values used. These were:

- use all singular values in the calculation;
- use the top 80% singular values;
- use the top 50% singular values;
- use the top 20% singular values; and
- use singular values s such that $s^2 > \max(s^2)/SNR$.



(a) High-resolution image

(b) Low-resolution image

Figure B.1: Test imagery for the SVD super-resolution algorithm.

B.2 Results and Conclusions

The singular value spectrum of the convolution matrix is shown in Figure B.2, where it can be seen that the smallest values have a very low magnitude compared to the largest value. Results of the super-resolution experiment, except for the all-singular-value version, are shown in Figure B.3. The best performing algorithm was the one based on a variable squared-singular-value threshold. This was the only version to maintain an output SNR of greater than 0 dB for the whole range of input SNRs. At low SNRs, the majority of the singular values were excluded, which tends to produce images that are not fully super-resolved but do not contain any spurious noise spikes. At higher SNRs more singular values are used, which allows a more accurate super-resolution result. The next best performing version was the one that used the top 20% singular values. At low SNRs more singular values were used than the variable threshold version. Using too many singular values allows the creation of noise spikes, which explains the lower performance of this algorithm at low SNRs. At the high end of SNRs tested here the algorithm uses a similar number of singular values as the variable threshold version and has a similar performance. Ultimately as the SNR approaches infinity more than 20% of the singular values should be used and the performance of this version would be expected to be lower than the

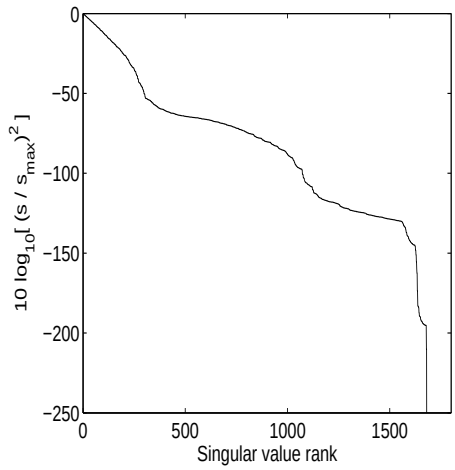


Figure B.2: Singular value spectrum of the 1681x1681 convolution matrix.

variable threshold version of the algorithm. Similar comments to those made for the 20% version apply to the other versions of the algorithm that use a fixed percentage of singular values, except that they clearly use too many values for the range of SNRs tested. Performance of the algorithm that used all singular values is not shown in Figure B.3 as its output SNR was constantly about 100 dB below that of the 80% version.

The original and super-resolved images for the variable-threshold SVD algorithm at an input SNR of 50 dB are shown in Figure B.4. Although the output SNR of the super-resolved image is about 5 dB, the algorithm has visually done a reasonable job of restoring the image from the low-resolution one shown in Figure B.1b. An SNR of 5 dB may seem low, but it should be remembered that this is an information-theoretic SNR based on the whole image rather than the peak SNR used in detection, which is always higher. The resolution of the super-resolved image is slightly lower than the original – it can be seen that a little detail has been lost. Although it may be possible to increase performance of the SVD algorithm with this particular image by changing the number of singular values used, this would not necessarily apply to other images in general. In fact it has previously been shown that resolution improvement using SVD is limited by noise [7]. Given that the variable-threshold version of the SVD algorithm takes SNR into account it seems like a good candidate for a general super-resolution algorithm.

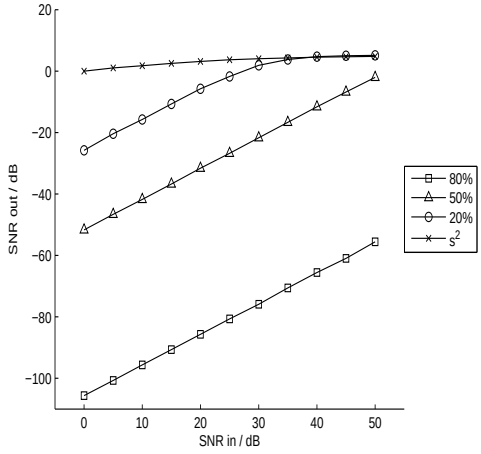
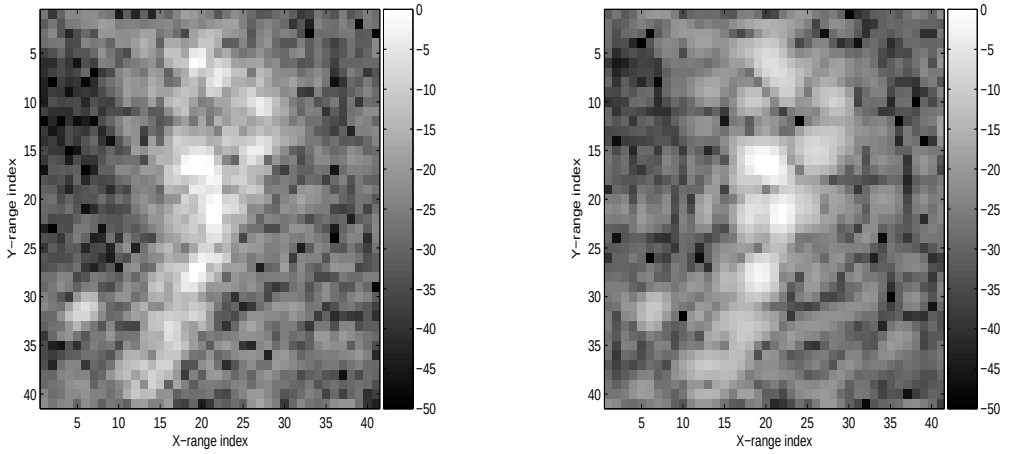


Figure B.3: Performance of four versions of the SVD super-resolution algorithm.



(a) Original high-resolution scene

(b) Super-resolved high-resolution scene

Figure B.4: Comparison of the original high-resolution scene with the SVD super-resolved scene.

Appendix C

The Radar Point Spread Function

C.1 Introduction

The point spread function model used in this thesis to test super-resolution algorithms is the single-parameter model introduced by Luttrell [88]. It is stated by Luttrell that a SAR system undergoing anomalous motion can in first order be modelled as the defocusing of a simple linear imaging system and can be considered as "a microwave version of an optical bench experiment using coherent illumination with the lens misplaced from its correct focus". The point spread function derived from Luttrell's model shall therefore be referred to here as an optics-based PSF. However, this model may not strictly be accurate for SAR systems as the physics involved is slightly different. Here we present the similarities and differences between Luttrell's model and a simple single-parameter model derived from first principles of radar signal processing. This allows us to relate Luttrell's focus parameter θ to the more conventional focus parameter $\delta\beta$ used by Oliver [106–108]. A more detailed analysis of the radar point spread function from a focusing point of view is given by Blacknell [7]. This appendix was published in a shortened form in [78].

C.2 Radar-based PSF

The precise nature of the radar PSF is dependent on several aircraft and radar parameters. Where necessary, reasonable values for the parameters have been assumed and correspond to those used in the work of Oliver [106]. The relevant parameters are: minimum range $R_0 = 40$ km, wavelength $\lambda = 0.03$ m, antenna length $d = 2$ m, aircraft velocity $v_x = 200$ ms^{-1} and synthetic-aperture time $T = 3$ s.

For a sideways looking radar traveling in a perfectly straight line the range-to-target is given by

$$R^2 = R_0^2 + x^2, \tag{C.1}$$

where x is the distance traveled by the radar platform from the position of closest approach. Since $x \ll R_0$ this may be approximated by

$$R \approx R_0 \left(1 + \frac{x^2}{2R_0^2}\right). \tag{C.2}$$

The amount the range-to-target differs from the minimum range is then

$$\Delta R = \frac{x^2}{2R_0}. \tag{C.3}$$

If the radar platform is undergoing constant cross-track acceleration a_y, there is an additional displacement of $a_y t^2/2$, where t is the time elapsed since closest approach. Thus overall the variation in phase with time is

$$\phi(t) = \frac{4\pi\Delta R}{\lambda} = (\beta_0 + \delta\beta)t^2, \tag{C.4}$$

where

$$\beta_0 = \frac{2\pi v_x^2}{R_0\lambda} \tag{C.5}$$

and

$$\delta\beta = \frac{2\pi a_y}{\lambda}. \tag{C.6}$$

The signal received by the radar for a point target is $s(t) = \exp(-i\phi(t))$. The matched filter output $g(t)$ is given by the cross-correlation of $s(t)$ with a reference signal $h^*(t) = \exp(i\beta_0 t^2)$:

$$g(t) = \frac{1}{T}\int_{-T/2}^{T/2} s(t+\tau)h^*(\tau)d\tau \tag{C.7}$$

$$= \frac{1}{T}\int_{-T/2}^{T/2} \exp\left[-i(\beta_0 + \delta\beta)(t+\tau)^2 + i\beta_0\tau^2\right] d\tau \tag{C.8}$$

$$= \frac{1}{T}\exp\left[-i(\beta_0 + \delta\beta)t^2\right]\int_{-T/2}^{T/2} \exp\left[-i(\beta_0 + \delta\beta)2t\tau - i\delta\beta\tau^2\right] d\tau. \tag{C.9}$$

For zero cross-track acceleration $\delta\beta$=0 and this simplifies to the familiar modulated sinc function

$$g(t) = \mathrm{sinc}(\beta_0 T t)\exp(-i\beta_0 t^2). \tag{C.10}$$

When the cross-track acceleration is non-zero the integral in (C.9) can be Fourier transformed to give

$$\begin{aligned} I(\omega) &= \frac{1}{T}\int_{-T/2}^{T/2}\exp\left(-i\delta\beta\tau^2\right)\int_{-T/2}^{T/2}\exp\left[-i(\beta_0+\delta\beta)2t\tau\right]\exp(i\omega t)dtd\tau \\ &= \int_{-T/2}^{T/2}\exp\left(-i\delta\beta\tau^2\right)\mathrm{sinc}\left(\left[(\beta_0+\delta\beta)2\tau-\omega\right]\frac{T}{2}\right)d\tau. \end{aligned} \tag{C.11}$$

If $\frac{\pi}{(\beta_0+\delta\beta)T} \ll 1$, which holds well for the parameters used here, the width of the sinc function is much less than its height and it may be approximated by the Dirac delta function

$$\mathrm{sinc}\left(\left[(\beta_0+\delta\beta)2\tau-\omega\right]\frac{T}{2}\right) \approx \frac{\pi}{(\beta_0+\delta\beta)T}\delta\left(\tau-\frac{\omega}{2(\beta_0+\delta\beta)}\right). \tag{C.12}$$

Using this approximation gives the Fourier domain representation of the integral as

$$I(\omega) = \frac{\pi}{(\beta_0+\delta\beta)T}\exp\left[\frac{-i\delta\beta\omega^2}{4(\beta_0+\delta\beta)^2}\right]. \tag{C.13}$$

The inverse Fourier transform of this function cannot be evaluated in terms of simple functions. However, the analytical form when using integration limits $\omega = \pm W/2$ is given by Mathematica 4.0 [157] as

$$\begin{aligned} I(t) &= \left(\frac{\pi}{\beta T}\right)\left(\frac{\sqrt{\pi}(-1)^{1/4}\beta}{W\sqrt{\delta\beta}}\right)\exp\left[\frac{i\beta^2t^2}{\delta\beta}\right]\times ... \\ &\quad \left\{\mathrm{erfi}\left[\frac{(-1)^{\frac{3}{4}}\left[4\beta^2t-\delta\beta W\right]}{4\sqrt{\delta\beta}\beta}\right]-\mathrm{erfi}\left[\frac{(-1)^{\frac{3}{4}}\left[4\beta^2t+\delta\beta W\right]}{4\sqrt{\delta\beta}\beta}\right]\right\}, \end{aligned} \tag{C.14}$$

where $\beta = \beta_0 + \delta\beta$, $\mathrm{erfi}(z) = \mathrm{erf}(iz)/i$ and $\mathrm{erf}(z) = (2/\sqrt{\pi})\int_0^z \exp(-t^2)dt$. This can be substituted for the integral in (C.9) to give the desired PSF as a function of time. To obtain the PSF in spatial co-ordinates t is replaced with x/v_x. In practice, for the simulations to be presented, the inverse Fourier transform has been performed numerically. The numerical result has been verified by comparing it to the analytic form of the PSF.

C.3 Optics-based PSF

Luttrell's model of the point spread function is given in [88] as

$$g(x) = \frac{1}{2c}\int_{-c}^{+c} \exp(ikx + i\theta k^2 x^2)dk. \tag{C.15}$$

The paper goes on to make a linear approximation valid when $|\theta c^2 x^2| < 1$ to solve this analytically. It is stated that for super-resolution within the main lobe $|x| < \pi/c$ and it is required that $|\theta| < 0.1$. However, it should be noted that in practice the image obtained from a radar will also contain energy in the side-lobes and either a much smaller value of θ should be used, which would limit the range of motion that can be accommodated, or the approximation would have to be be expanded to quadratic or higher orders of θ. If the approximation were expanded then Luttrell's autofocus/super-resolution algorithm would no longer apply because it depends on a PSF linear in θ. It is in fact possible to evaluate the integral (C.15) in terms of special functions and is given by Mathematica as:

$$\begin{aligned} g(x) &= \frac{(-1)^{3/4}e^{-i/(4\theta)}\sqrt{\pi}}{4c\sqrt{\theta}x} \times ... \\ &\left[\text{erfi}\left(\frac{(-1)^{1/4}(1-2c\theta x)}{2\sqrt{\theta}}\right) - \text{erfi}\left(\frac{(-1)^{1/4}(1+2c\theta x)}{2\sqrt{\theta}}\right)\right]. \end{aligned} \tag{C.16}$$

Again, the integral has been implemented numerically and verified for the following simulations.

C.4 PSF Comparison

A comparison between the optics model and the radar cross-track acceleration model was made by setting integration limits such that the -3dB resolution with no distortion was 1m for both models, setting θ to various values and then adjusting $\delta\beta$ until the two PSFs had the same first side-lobe level. The corresponding values of θ and $\delta\beta$ are displayed in Figure C.1 showing a mildly non-linear relationship between the two focus parameters. From the graph and using the assumed radar parameters, the residual cross-track acceleration corresponding to θ=0.1 (the value used in this thesis and [31, 80, 81, 88]) is 4.0×10^{-3} ms^{-2}.

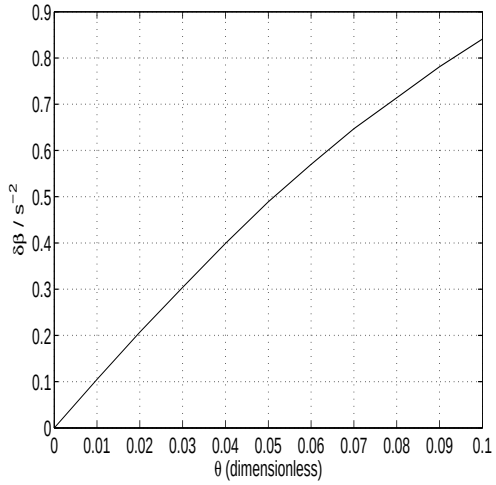
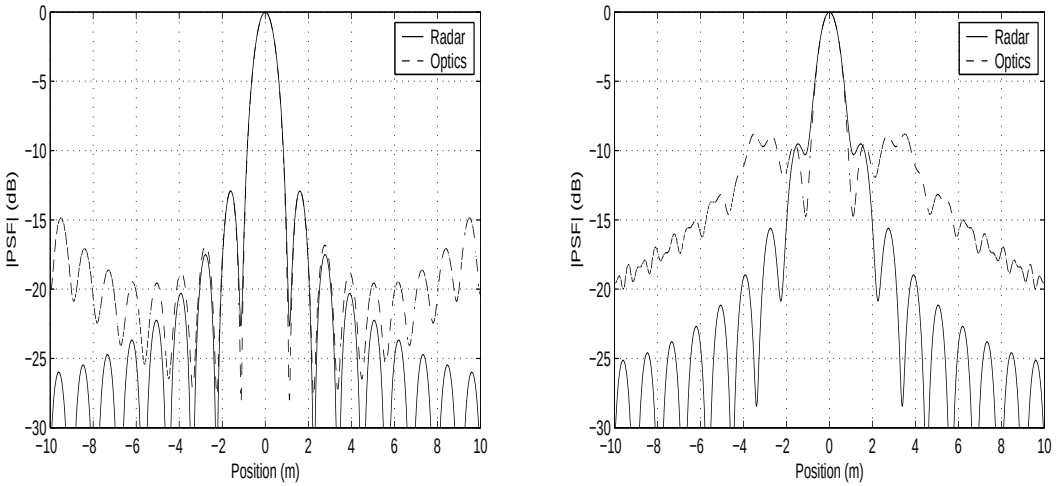


Figure C.1: Relationship between focus parameters when matching first sidelobe levels



(a) $\theta = 0.02$, $\delta\beta = 0.207s^{-2}$ (b) $\theta = 0.10$, $\delta\beta = 0.831s^{-2}$

Figure C.2: Comparison of radar and optics PSFs for two levels of defocus

Examples of the PSFs are shown in Figures C.2a and C.2b where θ=0.02 and 0.10 respectively. For θ=0.02 the level of defocus is low and the two PSFs are similar both to each other and to a completely focused sinc function (not shown) for the first few side-lobes. However, at larger distances from the main lobe the PSFs start to diverge with the optics PSF showing more distortion. For θ=0.10 the PSFs are substantially different. The optics PSF has a much higher level of distortion and the first side-lobe is no longer the strongest. In comparison, the radar cross-track acceleration PSF has deteriorated only slightly and is still similar to a sinc function. The resolution of the two PSFs degraded by about 3.6% for $\theta = 0.10$ and 5.0% for $\delta\beta = 0.831s^{-2}$.

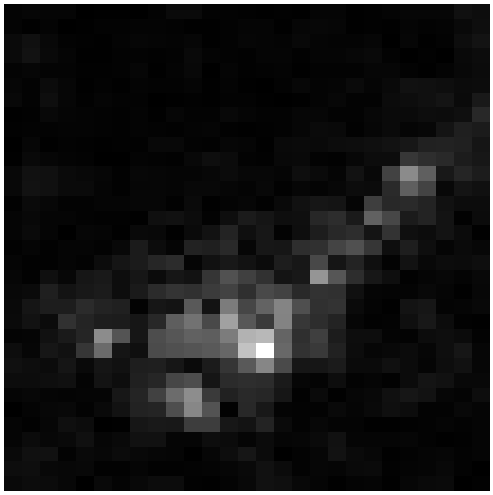
C.5 Super-resolution Implications

The effect of the differing PSFs is analyzed here using three standard superresolution algorithms: matrix inverse (INV), singular value decomposition (SVD) and thresholded minimum mean square error (MMSE-T). The super-resolution performance metric used to assess performance was the output signal-to-noise ratio defined as $SNR_{out} = ||\mathbf{f}||^2/||\hat{\mathbf{f}} - \mathbf{f}||^2$, where $\mathbf{f}$ is the true high-resolution scene and $\hat{\mathbf{f}}$ is the estimated scene using any particular algorithm. A Monte Carlo assessment has been carried out using measured SAR data. A high-resolution target image of 27x33 pixels had its resolution degraded using the cross-track acceleration PSF with $\delta\beta = 0.831s^{-2}$ and noise was added at an SNR of 30dB. Each super-resolution algorithm was then executed using the correct cross-track acceleration PSF, the optical PSF with the correct first side-lobe level ($\theta = 0.10$), and a sinc PSF. This was repeated 100 times with different noise realizations and the mean and standard deviations of the output SNR were measured.

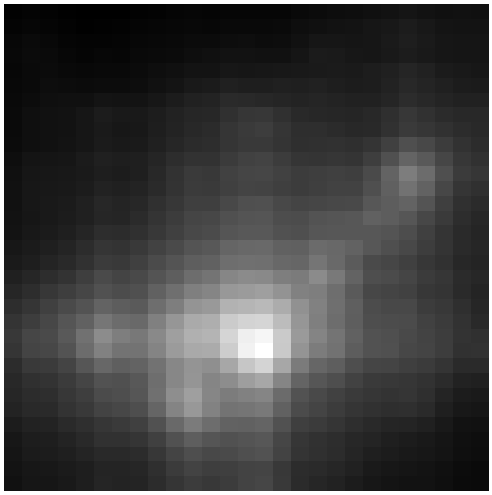
Results of the assessment are shown in Table C.1. The radar PSF gives the best performance for SVD and MMSE-T, which are the most reliable algorithms. This is to be expected as cross-track acceleration is the correct model used in the simulation. Using the optics PSF gives a worse performance than using a sinc function. This is because the optics PSF is less similar to the radar PSF than is the sinc function. Example images before and after super-resolution using MMSE-T and the radar PSF are shown in Figure C.3.

	Radar PSF		Optics PSF		Sinc PSF	
Algorithm	Mean	S.D.	Mean	S.D.	Mean	S.D.
INV	-11.83	1.16	-1.46	0.31	-4.16	1.10
SVD	8.85	0.08	2.14	0.03	8.10	0.11
MMSE-T	10.52	0.44	2.22	0.10	8.19	1.40

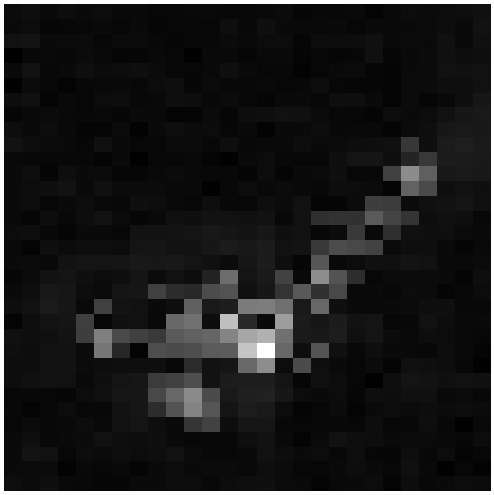
Table C.1: Output SNR in dB



(a) Original high-resolution image



(b) Low-resolution image



(c) Super-resolved image

Figure C.3: Stages of super-resolution. (a) Original image. (b) blurred low-resolution image. (c) Super-resolved image.

C.6 Conclusions

A radar and optics PSF have been compared using analytic expressions and numerical evaluations for the functions. The effect of using the optics PSF during super-resolution when the actual PSF is due to cross-track acceleration was found to be worse than using an idealized sinc function when no cross-track acceleration is present. Out of the algorithms tested MMSE-T was the best using the output SNR metric, which confirms previous results [81].

It should be noted that the single-parameter model based on cross-track acceleration may be insufficient to describe realistic motion of an aircraft. If this were the case, a multi-parameter model would have to be developed. Effects other than anomalous motion between the radar and target may also alter the PSF. Phase noise or non-linearities due to imperfections of the radar receiver, quantization noise and atmospheric phase disturbances all increase side-lobe levels. Also, scattering centres whose properties vary with frequency and imaging geometry result in non-ideal PSFs. These effects would further reduce the performance of super-resolution algorithms.

Bibliography

[1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19(6):716–723, December 1974.

[2] C. Andrieu and A. Doucet. Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC. *IEEE Transactions on Signal Processing*, 47(10):2667–2676, October 1999.

[3] J. P. Ballard. Optimum solution methods in range profile classification and finding the distribution of range cell fluctuations. *International Radar Conference (RADAR 97), Edinburgh, UK*, pages 639–642, October 1997.

[4] P. Barone and R. Ragona. Bayesian estimation of parameters of a damped sinusoidal model by a Markov chain Monte Carlo method. *IEEE Transactions on Signal Processing*, 45(7):1806–1814, July 1997.

[5] G. R. Benitz. High-definition vector imaging. *The Lincoln Laboratory Journal*, 10(2):147–169, 1997.

[6] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[7] D. Blacknell. *Synthetic aperture radar motion compensation using autofocus with implications for super-resolution*. PhD thesis, University of Sheffield, UK, March 1990.

[8] D. Blacknell and C. J. Oliver. Information content of coherent images. *J. Phys. D: Appl. Phys.*, 26:1364–1370, 1993.

[9] D. Blacknell and C. J. Oliver. Recognizing chaos in radar images. *J. Phys. D: Appl. Phys.*, 27:1608–1618, 1994.

[10] D. Blacknell and S.Quegan. SAR super-resolution with a perturbed point spread function. *IEE Colloquium on the Role of Image Processing in Defence and Military Electronics*, pages 3/1–3/4, April 1990.

[11] S. D. Blunt and K. Gerlach. Adaptive pulse compression via MMSE estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 42(2):572–584, April 2006.

[12] B. Borden. Maximum entropy regularization in inverse synthetic aperture radar imagery. *IEEE Transactions on Signal Processing*, 40(4):969–973, April 1992.

[13] E. C. Botha. Classification of aerospace targets using superresolution ISAR images. *Proceedings of the IEEE South African Symposium on Communications and Signal Processing (COMSIG'94)*, pages 138–145, October 1994.

[14] R. K. Bryan and J. Skilling. Deconvolution by maximum entropy, as illustrated by application to the jet of M87. *Mon. Not. R. Astr. Soc.*, 191:69–79, April 1980.

[15] J. P. Burg. *Maximum entropy spectral analysis.* PhD thesis, Department of Geophysics, Stanford University, USA, May 1975.

[16] F. M. Candocia. *A unified superresolution approach for optical and synthetic aperture radar images.* PhD thesis, University of Florida, USA, May 1998.

[17] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE*, 57(8):1408–1418, August 1969.

[18] W. G. Carrara, R. S. Goodman, and R. M. Majewski. *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms.* Artech House, London, October 1995.

[19] G. Casella and E. George. Explaining the Gibbs sampler. *The American Statistician*, 46(3):167–174, August 1992.

[20] M. Çetin. *Feature-enhanced synthetic aperture radar imaging.* PhD thesis, College of Engineering, Boston University, USA, 2001.

[21] M. Çetin and W. C. Karl. Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization. *IEEE Transactions on Image Processing*, 10(4):623–631, April 2001.

[22] R. Chellapa and R. L. Kashyap. Digital image restoration using spatial interaction models. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30(3):461–472, June 1982.

[23] C. Chesnaud, P. Réfrégier, and V. Boulet. Statistical region snake-based segmentation adapted to different physical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11), November 1999.

[24] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335, November 1995.

[25] T. J. Connolly and R. G. Lane. Gradient methods for superresolution. *IEEE International Conference on Image Processing*, 1:917–920, October 1997.

[26] T. J. Connolly and R. G. Lane. Constrained regularization methods for superresolution. *IEEE International Conference on Image Processing*, 3:727–731, October 1998.

[27] K. D. Copsey. *Bayesian Approaches for Robust Automatic Target Recognition*. PhD thesis, Imperial College London, UK, August 2004.

[28] K. D. Copsey, N. J. Gordon, and A. D. Marrs. Bayesian analysis of generalized frequency-modulated signals. *IEEE Transactions on Signal Processing*, 50(3):725–735, March 2002.

[29] K. D. Copsey, R. O. Lane, S. Manchanda, and A. R. Webb. Bayesian approach to exploiting prior targeting information within a weapon seeker. *NATO RTO SET Symposium SET-080, Target Identification and Recognition using RF Systems, Oslo, Norway*, October 2004.

[30] K. D. Copsey, R. O. Lane, S. Manchanda, and A. R. Webb. Bayesian approach to recognising relocatable targets. *NATO RTO SET Symposium SET-080, Target Identification and Recognition using RF Systems, Oslo, Norway*, October 2004.

[31] K. D. Copsey, R. O. Lane, and A. R. Webb. Designing NCTR algorithms when operating sensor conditions differ from training conditions. *International conference on radar systems (Radar2004), Toulouse, France*, October 2004.

[32] T. Cornwell and A. Bridle. Deconvolution tutorial. *Technical Report, National Radio Astronomy Observatory*, November 1996.

[33] N. Cotuk, S. Ture, and M. Çetin. Application of point enhancement technique for ship target recognition by HRR. *Proceedings of SPIE*, 5095:185–193, September 2003.

[34] J. Cui, J. Gudnason, and M. Brookes. Radar shadow and superresolution features for automatic recognition of MSTAR targets. *IEEE International Conference on Radar, Arlington, Virginia*, pages 534–539, May 2005.

[35] A. Curtis and A. Lomax. Prior information, sampling distributions, and the curse of dimensionality. *Geophysics*, 66(2):372–378, March–April 2001.

[36] M. Datcu, K. Seidel, and M. Walessa. Spatial information retrieval from remote-sensing images - part I: information theoretical perspectives. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1431–1445, September 1998.

[37] M. Datcu and M. Walessa. Maximum entropy methods for despeckling and resampling synthetic aperture radar images of rough terrain. *Proceedings of SPIE*, 3217:76–83, September 1997.

[38] F. Daum and J. Huang. Curse of dimensionality and particle filters. *Proceedings of the IEEE Aerospace Conference*, 4:1979–1993, March 2003.

[39] J. F. G. de Freitas. *Bayesian Methods for Neural Networks*. PhD thesis, Engineering Department, University of Cambridge, UK, 1999.

[40] S. R. DeGraaf. SAR imaging via modern 2-D spectral estimation methods. *IEEE Transactions on Image Processing*, 7(5):729–761, May 1998.

[41] L. M. Delves, G. C. Pryde, and S. P. Luttrell. A super-resolution algorithm for SAR images. *Inverse Problems*, 4(3):681–703, 1988.

[42] S. D'Ercole. Super-resolution techniques applied to MSTAR data. *NATO RTO SET Symposium SET-080, Target Identification and Recognition using RF Systems, Oslo, Norway*, October 2004.

[43] F. M. Dickey, L. A. Romero, J. M. DeLaurentis, and A. W. Doerry. Super-resolution, degrees of freedom and synthetic aperture radar. *IEE Proceedings – Radar, Sonar and Navigation*, 150(6):419–429, December 2003.

[44] A. W. Doerry and F. M. Dickey. Synthetic aperture radar. *Optics and Photonics News*, 15(11):28–33, November 2004.

[45] D. L. Donoho. High-dimensional data analysis: the curses and blessings of dimensionality. *American Mathematical Society conference on Mathematical Challenges of the 21st Century*, August 2000.

[46] L. Dou and R. J. W. Hodgson. Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation: I. *Inverse Problems*, 11(5):1069–1085, 1995.

[47] L. Dou and R. J. W. Hodgson. Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation: II. *Inverse Problems*, 12(2):121–137, 1996.

[48] J. R. Fienup. Detecting moving targets in SAR imagery by focusing. *IEEE Transactions on Aerospace and Electronic Systems*, 37(3):794–809, July 2001.

[49] B. R. Frieden. Restoring with maximum likelihood and maximum entropy. *Journal of the Optical Society of America*, 62(4):511–518, April 1972.

[50] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall, 2nd edition, 2004.

[51] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.

[52] P. J. Green. Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[53] V. Guglielmi, F. Castani, and P. Piau. Application of super-resolution methods to synthetic aperture radar data. *International Geoscience and Remote Sensing Symposium, IGARRS95*, 3:2289–2291, July 1995.

[54] S. Gulam-Razul, W. J. Fitzgerald, and C. Andrieu. Bayesian model selection and parameter estimation of nuclear emission spectra using RJMCMC. *Nuclear Instruments and Methods in Physics Research A*, 437:492–510, February 2003.

[55] S. F. Gull and G. J. Daniel. Image reconstruction from incomplete and noisy data. *Nature*, 272:686–690, April 1978.

[56] S. F. Gull and J. Skilling. Maximum entropy method in image processing. *IEE Proceedings Part F*, 131(6):646–659, October 1984.

[57] J. L. Harris. Diffraction and resolving power. *Journal of the Optical Society of America*, 54(7):931–936, July 1964.

[58] D. M. Higdon, J. E. Bowsher, V. E. Johnson, T. G. Turkington, D. R. Gilland, and R. J. Jaszczak. Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data. *IEEE Transactions on Medical Imaging*, 16(5):516–526, October 1997.

[59] J. A. Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement*, 15:417–426, June 1974.

[60] A. M. Howatson, P. G. Lund, and J. D. Todd. *Engineering Tables and Data*. Chapman and Hall, London, 2nd edition, 1991.

[61] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(5):814–824, May 1990.

[62] B. Huether, T. Kempf, and M. Çetin. A fast hybrid approach for SAR classification. *Europen Conference on Synthetic Aperture Radar, Ulm, Germany*, 1:443–446, May 2004.

[63] C. Hülsmeyer. Hertzian-wave projecting and receiving apparatus adapted to indicate or give warning of the presence of a metallic body, such as a ship or a train, in the line of projection of such waves. *British Patent 13170*, 22nd September 1904.

[64] S. W. Isakson. *Bayesian superresolution.* PhD thesis, University of California, Santa Barbara, USA, June 2001.

[65] A. K. Jain. Fast inversion of banded Toeplitz matrices by circular decompositions. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-26(2):121–126, April 1978.

[66] A. Jalobeanu, L. Blanc-Féraud, and J. Zerubia. Hyperparameter estimation for satellite image restoration using a MCMC maximum-likelihood method. *Pattern Recognition*, 35(2):341–352, February 2002.

[67] E. T. Jaynes. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, 4(3):227–241, September 1968.

[68] E. T. Jaynes. Confidence intervals vs. Bayesian intervals. In W. L. Harper and C. A. Hooker, editors, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, pages 175–257. D. Reidel Publishing, Dordrecht, Holland, 1976.

[69] E. T. Jaynes. On the rationale of maximum entropy methods. *Proceedings of the IEEE*, 70(9):939–952, September 1982.

[70] E. T. Jaynes. Bayesian methods: general background. In J. H. Justice, editor, *Maximum Entropy and Bayesian Methods in Applied Statistics*, pages 1–25. Cambridge University Press, 1985.

[71] S. M. Kay and S. L. Marple. Spectral analysis – a modern perspective. *Proceedings of the IEEE*, 69(11):1380–1419, November 1981.

[72] K. T. Kim, J. H. Bae, and H. T. Kim. Effect of AR model-based data extrapolation on target recognition performance. *IEEE Transactions on Antennas and Propagation*, 51(4):912–914, April 2003.

[73] K. T. Kim, D. K. Seo, and H. T. Kim. Efficient radar target recognition using the MUSIC algorithm and invariant features. *IEEE Transactions on Antennas and Propagation*, 50(3):325–337, March 2002.

[74] M. Kirscht. Detection and imaging of arbitrarily moving targets with single-channel SAR. *IEE Proceedings – Radar, Sonar and Navigation*, 150(1):7–11, February 2003.

[75] D. E. Kreithen, S. D. Halverson, and G. J. Owirka. Discriminating targets from clutter. *The Lincoln Laboratory Journal*, 6(1), 1993.

[76] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel. Adaptive restoration of images with speckle. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-35(3):373–383, March 1987.

[77] R. O. Lane. Estimating radar cross section using Bayesian image restoration. *Proceedings of the London Communications Symposium*, pages 1–4, September 2003.

[78] R. O. Lane. Super-resolution and the radar point spread function. *Proceedings of the London Communications Symposium*, pages 5–8, September 2005.

[79] R. O. Lane. The effects of doppler and pulse eclipsing on sidelobe reduction techniques. *IEEE National Radar Conference, Verona, NY*, April 2006.

[80] R. O. Lane, K. D. Copsey, and A. R. Webb. A Bayesian approach to simultaneous autofocus and super-resolution. *Proceedings of SPIE*, 5427:133–142, April 2004.

[81] R. O. Lane, K. D. Copsey, and A. R. Webb. Assessment of a Bayesian approach to recognising relocatable targets. *NATO RTO SET-096 specialists' meeting on the millimeterwave advanced target recognition and identification experiment (MATRIX 2005), Oberammergau, Germany*, May 2005.

[82] P. M. Lee. *Bayesian Statistics, An Introduction.* Arnold, London, 2nd edition, 1997.

[83] J. Li and P. Stoica. Efficient mixed-spectrum estimation with applications to target feature extraction. *29th Asilomar conference on Signals, Systems and Computers*, 1:428–432, November 1995.

[84] J. Li and P. Stoica. An adaptive filtering approach to spectral estimation and SAR imaging. *IEEE Transactions on Signal Processing*, 44(6):1469–1484, June 1996.

[85] X. Liao and Z. Bao. Radar target recognition using superresolution range profiles as features. *Proceedings of SPIE*, 3545:397–400, September 1998.

[86] Z.-S. Liu and J. Li. Implementation of the RELAX algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 34(2):657–664, April 1998.

[87] S. P. Luttrell. Prior knowledge and object reconstruction using the best linear estimate technique. *Optica Acta*, 32(6):703–716, June 1985.

[88] S. P. Luttrell. A Bayesian derivation of an iterative autofocus / superresolution algorithm. *Inverse Problems*, 6(6):975–996, 1990.

[89] S. P. Luttrell. The theory of Bayesian super-resolution of coherent images: a review. *International Journal of Remote Sensing*, 12(2):303–314, February 1991.

[90] S. P. Luttrell and C. J. Oliver. Prior knowledge in synthetic-aperture radar processing. *J. Phys. D: Appl. Phys.*, 19:333–356, 1986.

[91] P. A. C. Marques and J. M. B. Dias. Moving targets in synthetic aperture images: a Bayesian approach. *IEEE International Conference on Image Processing*, 1:685–688, September 2000.

[92] G. T. Maskall. *Feature extraction for robust automatic target recognition.* PhD thesis, School of Electronic, Electrical and Computer Engineering, University of Birmingham, UK, November 2006.

[93] S. Maskell. *Sequentially structured Bayesian solutions.* PhD thesis, Engineering Department, University of Cambridge, UK, February 2004.

[94] J. Mather. The incremental multi-parameter algorithm. *24th Asilomar conference on signals, systems and computers*, 1:368–372, November 1990.

[95] A. L. McLean. *Applications of Maximum Entropy Data Analysis*. PhD thesis, Department of Physics, University of Southampton, UK, September 1995.

[96] R. K. Mehra, A. Gandhe, M. Huff, and B. Ravichandran. A comparison of superresolution algorithms. *Proceedings of SPIE*, 3462:252–260, July 1998.

[97] R. K. Mehra, B. Ravichandran, and M. Huff. Survey of radar superresolution methods with applications to automatic target recognition. *Proceedings of SPIE*, 3374:186–193, April 1998.

[98] A. Mohammad-Djafari, J.-F. Giovanelli, G. Demoment, and J. Idier. Regularization, maximum entropy and probabilistic methods in mass spectrometry data processing problems. *International Journal of Mass Spectrometry*, 215(1):175–193, April 2002.

[99] T. G. Moore, B. W. Zuerndorfer, and E. C. Burt. Enhanced imagery using spectral-estimation based techniques. *The Lincoln Laboratory Journal*, 10(2):171–186, 1997.

[100] F. D. Neeser and J. L. Massey. Proper complex random processes with applications to information theory. *IEEE Transactions on Information Theory*, 39(4):1293–1302, July 1993.

[101] D. H. Nguyen, G. R. Benitz, J. H. Kay, B. J. Orchard, and R. Whiting. Superresolution HRR ATR with high definition vector imaging. *IEEE Transactions on Aerospace and Electronic Systems*, 37(4):1267–1286, October 2001.

[102] M. K. Nguyen and A. Mohammad-Djafari. Bayesian approach with the maximum entropy principle in image reconstruction from microwave scattered field data. *IEEE Transactions on Medical Imaging*, 13(12):254–262, June 1994.

[103] L. M. Novak and M. C. Burl. Optimal speckle reduction in polarimetric SAR imagery. *IEEE Transactions on Aerospace and Electronic Systems*, 26(2):293–305, March 1990.

[104] L. M. Novak, G. J. Owirka, and C. M. Netishen. Performance of a high-resolution polarimetric SAR automatic target recognition system. *The Lincoln Laboratory Journal*, 6(1):11–24, 1993.

[105] L. M. Novak, G. J. Owirka, and A. L. Weaver. Automatic target recognition using enhanced resolution SAR data. *IEEE Transactions on Aerospace and Electronic Systems*, 35(1):157–175, January 1999.

[106] C. J. Oliver. Synthetic-aperture radar imaging. *J. Phys. D: Appl. Phys.*, 22:871–890, 1989.

[107] C. J. Oliver. High-frequency limits on SAR autofocus and phase correction. *Int. J. Remote Sensing*, 14(3):495–519, February 1993.

[108] C. J. Oliver. The limits on SAR resolution imposed by autofocus uncertainty. *Int. J. Remote Sensing*, 14(3):485–494, February 1993.

[109] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36, May 2003.

[110] D. Pastina, A. Farina, J. Gunning, and P. Lombardo. Two-dimensional spectral analysis applied to SAR images. *IEE Proceedings – Radar, Sonar and Navigation*, 145(5):281–290, October 1998.

[111] D. N. Pedlar and D. Blacknell. Target delineation and classification using a region-based active-contour and a support vector machine classifier on SAR imagery. *5th European Conference on Synthetic Aperture Radar (EUSAR 2004), Ulm, Germany*, pages 141–144, May 2004.

[112] R. P. Perry, R. C. Dipietro, and R. L. Fante. SAR imaging of moving targets. *IEEE Transactions on Aerospace and Electronic Systems*, 35(1):188–200, January 1999.

[113] B. Picinbono. Second-order complex random vectors and normal distributions. *IEEE Transactions on Signal Processing*, 44(10):2637–2640, October 1996.

[114] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman. Bayesian image super-resolution, continued. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 1089–1096. MIT Press, Cambridge, MA, 2007.

[115] L. C. Potter and R. L. Moses. Attributed scattering centers for SAR ATR. *IEEE Transactions on Image Processing*, 6(1):79–91, January 1997.

[116] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.

[117] E. Radoi and A. Quinquis. Superresolution imagery based classification of some types of scale reduced radar targets. *The 31st internationally attended scientific conference: Modern Technologies in the XXI century, Bucharest, Romania*, November 2005.

[118] E. Radoi, A. Quinquis, F. Totir, and F. Pellen. Automatic radar target recognition using superresolution MUSIC 2D images and self-organising neural network. *Proceedings of the European signal processing conference*, pages 2139–2142, September 2004.

[119] B. D. Rao and K. V. S. Hari. Performance analysis of ESPRIT and TAM in determining the direction of arrival of plane waves in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12):1990–1995, December 1989.

[120] B. Ravichandran and R. K. Mehra. FOPEN radar ATR using superresolution and Fishertemplates. *Proceedings of SPIE*, 3721:253–260, April 1999.

[121] Q. S. Ren and A. J. Willis. High resolution array processing algorithm for characterisation of extended sources. *Electronics Letters*, 34(21):2006–2007, October 1998.

[122] M. A. Richards. Iterative noncoherent angular superresolution. *Proceedings of the IEEE National Radar Conference*, pages 100–105, April 1988.

[123] T. Ross, S. Worrell, V. Velten, J. Mossing, and M. Bryant. Standard SAR ATR evaluation experiments using the MSTAR public release data set. *Proceedings of SPIE*, 3370:566–573, April 1998.

[124] R. Roy and T. Kailath. ESPRIT–estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7), July 1989.

[125] J. J. Sacchini, A. Romano, and W. M. Steedly. Evaluation of single and full-polarization two-dimensional prony techniques applied to radar data. *Proceedings of SPIE*, 2234:91–105, April 1994.

[126] T. K. Sarkar. An ultra-low sidelobe pulse compression technique for high performance radar systems. *IEEE National Radar Conference*, pages 111–114, May 1997.

[127] T. K. Sarkar and O. Pereira. Using the matrix pencil method to estimate the parameters of a sum of complex exponentials. *IEEE Antennas and Propagation Magazine*, 37(1):48–55, February 1995.

[128] T. K. Sarkar, D. D. Weiner, and V. K. Jain. Some mathematical considerations in dealing with the inverse problem. *IEEE Transactions on Antennas and Propagation*, AP-29(2):373–379, March 1981.

[129] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, March 1986.

[130] P. J. Schreier and L. L. Scharf. Second-order analysis of improper complex random vectors and processes. *IEEE Transactions on Signal Processing*, 51(3):714–725, March 2003.

[131] M. Schröder, H. Rehrauer, K. Seidel, and M. Datcu. Spatial information retrieval from remote-sensing images - part II: Gibbs-Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1446–1455, September 1998.

[132] U. J. Schwarz. Mathematical-statistical description of the iterative beam removing technique (method CLEAN). *Astronomy and Astrophysics*, 65:345–356, April 1978.

[133] Y. Selén and P. Stoica. Estimation of semi-sparse radar range profiles. *Digital Signal Processing*, 2008.

[134] S. H. W. Simpson, T. D. Welsh, and P. E. R. Galloway. Coherent superresolution and its role in target recognition for PCR. *Multinational conference on passive and covert radar, Roke Manor Research, UK*, June 2002.

[135] J. Skilling and R. K. Bryan. Maximum entropy image reconstruction: general algorithm. *Mon. Not. R. Astr. Soc.*, 211:111–124, November 1984.

[136] M. I. Skolnik. *Introduction to Radar Systems*. McGraw-Hill Book Company, 2nd edition, 1980.

[137] P. F. Smith, M. A. Player, and D. A. L. Collie. The performance of the maximum entropy method: deconvolution and the frequency content of data. *J. Phys. D: Appl. Phys.*, 22:906–914, 1989.

[138] H. C. Stankwitz, R. J. Dallaire, and J. R. Fienup. Nonlinear apodization for sidelobe control in SAR imagery. *IEEE Transactions on Aerospace and Electronic Systems*, 31(1):267–279, January 1995.

[139] H. C. Stankwitz and M. R. Kosek. Sparse aperture fill for SAR using super-SVA. *IEEE National Radar Conference*, pages 70–75, May 1996.

[140] G. Stawinski, A. Doucet, and P. Duvaut. Reversible jump Markov chain Monte Carlo for Bayesian deconvolution of point sources. *Proceedings of SPIE*, 3459:179–190, July 1998.

[141] D. Stewart, D. Blacknell, A. Blake, R. Cook, and C. Oliver. Optimal approach to SAR image segmentation and classification. *IEE Proceedings – Radar, Sonar and Navigation*, 147(3):134–142, June 2000.

[142] P. Tait. *Introduction to Radar Target Recognition*. IEE Radar, Sonar and Navigation series 18. The Institution of Electrical Engineers, 2005.

[143] J. Tamminen and E. Kyrölä. Bayesian solution for nonlinear and non-gaussian inverse problems by Markov chain Monte Carlo method. *Journal of Geophysical Research*, 106(D13):14377–14390, July 2001.

[144] C. Theys, A. Ferrari, and M. Vieira. Marginal Bayesian analysis of polynomial-phase signals. *Signal Processing*, 81(1):69–82, January 2001.

[145] C. Theys, M. Vieira, and G. Alengrin. A reversible jump sampler for polynomial-phase signals. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1833–1836, March 1999.

[146] C. Theys, M. Vieira, and A. Ferrari. Bayesian estimation of the parameters of a polynomial phase signal using MCMC methods. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3553–3556, April 1997.

[147] J. Tian and K.-K. Ma. A MCMC approach for Bayesian super-resolution image reconstruction. *IEEE International Conference on Image Processing*, 1:45–48, September 2005.

[148] J. Tian and K.-K. Ma. Markov chain Monte Carlo super-resolution image reconstruction with artifacts suppression. *IEEE Asia Pacific Conference on Circuits and Systems*, pages 940–943, December 2006.

[149] J. Tsao and B. D. Steinberg. Reduction of sidelobe and speckle artifacts in microwave imaging: the CLEAN technique. *IEEE Transactions on Antennas and Propagation*, 36(4):543–556, April 1988.

[150] A. van den Bos. Alternative interpretation of maximum entropy spectral analysis. *IEEE Transactions on Information Theory*, IT-17:493–494, July 1971.

[151] D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz. Phase gradient autofocus - a robust tool for high resolution SAR phase correction. *IEEE Transactions on Aerospace and Electronic Systems*, 30(3):827–835, July 1994.

[152] K. D. Ward, R. J. A. Tough, and S. Watts. *Sea Clutter: Scattering, the K Distribution and Radar Performance*. The Institution of Engineering and Technology, 2006.

[153] M. Wax and T. Kailath. Efficient inversion of Toeplitz-block Toeplitz matrix. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-31(5):1218–1221, October 1983.

[154] A. R. Webb. Gamma mixture models for target recognition. *Pattern Recognition*, 33(12):2045–2054, December 2000.

[155] A. R. Webb. *Statistical Pattern Recognition*. John Wiley & Sons, Chichester, 2nd edition, August 2002.

[156] S. Werness, W. Carrara, L. Joyce, and D. Franczak. Moving target imaging algorithm for SAR data. *IEEE Transactions on Aerospace and Electronic Systems*, 26(1):57–67, January 1990.

[157] S. Wolfram. *The Mathematica Book*. Wolfram Media/CUP, 4th edition, 1999.

[158] P. M. Woodward. *Probability and Information Theory with Applications to Radar*. Pergamon Press, 1955.

[159] R. Wu and J. Li. Autofocus and super-resolution synthetic aperture radar image formation. *IEE Proceedings – Radar, Sonar and Navigation*, 147(5):217–223, October 2000.

[160] J. Xu, Y. Peng, and X.-G. Xia. Parametric autofocus of SAR imaging–inherent accuracy limitations and realization. *IEEE Transactions on Geoscience and Remote Sensing*, 42(11):2397–2411, November 2004.

[161] K. Zarb Adami. *Bayesian inference and deconvolution*. PhD thesis, University of Cambridge, UK, December 2003.

[162] X. Zhang, R. Shen, and G. Guo. Automatic HRR target recognition based on Prony model wavelet and probability neural network. *CIE International Conference of Radar*, pages 143–146, October 1996.

[163] X. Zhang, Z. Zhuang, and G. Guo. Automatic HRR target recognition based on matrix pencil method and multiresolution neural network. *Proceedings of SPIE*, 3069:518–525, June 1997.

[164] B. M. Zrnic, A. J. Zejak, and I. S. Simic. Target detection enhancement for the chirp radar in the eclipsing zone. *IEEE 6th International symposium on spread spectrum techniques and applications*, 1:269–273, September 2000.

[165] J. P. Zwart, B. Kröse, and S. Gelsema. Aircraft classification from estimated models of radar scattering. *Technical Report IAS-UVA-03-02, Informatics Institute, University of Amsterdam, The Netherlands*, January 2003.