Estimation of K-Distribution Parameters with Application to Target Detection

Mohammad Hamiruce Marhaban

Submitted for the Degree of Doctor of Philosophy from the University of Surrey

Centre for Vision, Speech and Signal Processing School of Electronics and Physical Sciences University of Surrey Guildford, Surrey GU2 7XH, U.K.

March 2003

© Mohammad Hamiruce Marhaban 2003
Summary

Probabilistic models have been used extensively in the past to underpin classification algorithms in statistical pattern recognition. The most widely used model is the Gaussian distribution. However, signals of impulsive nature usually deviate from Gaussian and it is necessary to work with more realistic models.

K-distribution is one of the long-tailed density which is known in the signal processing community for fitting the radar sea clutter accurately. The work presented in this thesis reflects the efforts made to model the background features, extracted from the sea images, by using a K-distribution.

A novel approach for estimating the parameter of K-distribution is presented. The method utilises the empirical characteristic function, and is proven to perform better than any existing estimation technique. A classifier is then developed from the empirical characteristic function. This technique is applied to a problem of automatic target recognition with promising results.

**Key words:** K-distribution, parameter estimation, empirical characteristic function, automatic target recognition
Acknowledgements

In the name of Allah, the Most gracious, the Most merciful.

I would first like to convey my most sincere thanks to both my supervisors, Prof. Josef Kittler and Dr. George Szajnowski, for their great support and guidance during the completion of this thesis. I am truly grateful for the freedom I have been granted to explore new ideas and the opportunity to live an enriching experience.

I would also like to thank my sponsor, namely, Universiti Putra Malaysia, for their financial support during my study.

My special thanks to the members of the Centre for Vision, Speech and Signal Processing for their help, support, interesting discussion and friendship throughout my research. In particular, thanks to Kieron Messer, for his help and collaboration at several stages of my research.

My greatest thanks go to my parents, who always believe in me. Thanks for always ensuring the best education for me, and I wholeheartedly appreciate all the sacrifices you made. This thesis is dedicated to both of you. Thanks to all my brothers and sisters, and the entire family.

My final thanks is for my wife, and my son, the source of my inspiration.

"Khas buat ayahanda dan bonda"
Contents

1 Introduction ................................................................. 1
  1.1 Automatic Target Recognition ............................................. 3
    1.1.1 Overview ........................................................................... 3
    1.1.2 Pattern Recognition Approach for ATR ......................... 5
    1.1.3 Preprocessing ...................................................................... 6
    1.1.4 Target detection ............................................................... 7
    1.1.5 Segmentation ...................................................................... 8
    1.1.6 Feature Computation .......................................................... 10
    1.1.7 Classification and Identification ....................................... 11
  1.2 Research Interests, Aims and Objectives .............................. 12
  1.3 Outline of the Thesis and Contributions .............................. 13

2 Statistical Modelling and Characteristic Function .................. 17
  2.1 Gaussian Distribution as a Probabilistic Models .................. 18
  2.2 Radar Sea Clutter ............................................................... 19
    2.2.1 K-Distribution ................................................................. 20
    2.2.2 Quadrature Components .................................................. 21
  2.3 Characteristic Function ....................................................... 24
  2.4 Empirical Characteristic Function ....................................... 27
  2.5 Conclusion ........................................................................... 28
## 3 K-Distribution Parameter Estimation; A Review

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Maximum Likelihood Methods</td>
<td>30</td>
</tr>
<tr>
<td>3.2</td>
<td>Method of Moments</td>
<td>31</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Higher Order Moments</td>
<td>32</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Fractional Lower Order Moments</td>
<td>33</td>
</tr>
<tr>
<td>3.3</td>
<td>Texture Measure Based Method</td>
<td>34</td>
</tr>
<tr>
<td>3.4</td>
<td>Density Approximation Method</td>
<td>36</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Gamma Density Approximation</td>
<td>37</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Generalised Bessel K function Density</td>
<td>39</td>
</tr>
<tr>
<td>3.5</td>
<td>Characteristic Function Method</td>
<td>41</td>
</tr>
<tr>
<td>3.6</td>
<td>Neural Networks Based Method</td>
<td>43</td>
</tr>
<tr>
<td>3.7</td>
<td>Result and Discussion</td>
<td>43</td>
</tr>
<tr>
<td>3.8</td>
<td>Conclusion</td>
<td>47</td>
</tr>
</tbody>
</table>

## 4 A Novel Approach to K-Distribution Parameter Estimation

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Motivation: A Case of Known Scale Parameter</td>
<td>50</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Simulation Results and Discussion</td>
<td>52</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Conclusion</td>
<td>55</td>
</tr>
<tr>
<td>4.2</td>
<td>Log of K-Distribution</td>
<td>56</td>
</tr>
<tr>
<td>4.3</td>
<td>Scale Invariant Random Variables</td>
<td>60</td>
</tr>
<tr>
<td>4.4</td>
<td>Parameter Estimation</td>
<td>63</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Using Second Order Moments</td>
<td>63</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Parameter Estimation by Empirical Characteristic Function</td>
<td>64</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Parameter Estimation using Kernel Characteristic Function Estimator</td>
<td>66</td>
</tr>
<tr>
<td>4.5</td>
<td>Methods for Scale Parameter Elimination</td>
<td>68</td>
</tr>
<tr>
<td>4.6</td>
<td>Application of the Proposed Methods to Other Densities</td>
<td>70</td>
</tr>
<tr>
<td>4.7</td>
<td>Simulation Result</td>
<td>72</td>
</tr>
<tr>
<td>4.7.1</td>
<td>A Comparison on the Scale Parameter Elimination Method</td>
<td>72</td>
</tr>
</tbody>
</table>
## Contents

4.7.2 Comparison of the Characteristic Function Based Method 76
4.8 Discussion .......................................................... 79
4.9 Conclusion ......................................................... 82

5 Distance Function Via Characteristic Function 85
5.1 One Dimensional Statistical Distance .................. 87
5.2 Multidimensional Statistical Distance ................. 92
5.3 Conclusion ......................................................... 94

6 Application of K-Distribution Model in Target Detection 95
6.1 Preprocessing ..................................................... 97
6.2 Model Generation .............................................. 98
6.3 Model Optimisation .......................................... 99
6.4 Target Detection ............................................... 102
6.5 Temporal Tracking ............................................. 102
6.6 Simulation Result .............................................. 103
6.7 Conclusion ......................................................... 111

7 Conclusion 113
7.1 Summary .......................................................... 113
7.2 Future Work ..................................................... 115

A Mathematical Derivation 117
A.1 Scale Invariant K-Distributed RV ...................... 117
A.2 Scale Invariant Weibull Distributed RV ............. 118

B Original Formulation of Previous Method 121

C Matlab Source Code 123
List of Figures

1.1 Block diagram of a typical ATR system. .................................................. 6
1.2 FLIR Images after segmentation. Left-the binary segmented image. Right-the grey value image. ................................................................. 9

2.1 The amplitude PDF of K-distribution. ..................................................... 22
2.2 The PDF of univariate K-distribution. .................................................... 22
2.3 A comparison of univariate K-distribution (t) process with $\nu = 0.1$ and Gaussian $y(t)$ process with equivalent variance. .................. 23
2.4 A comparison of K-distribution $x(t)$ process with $\nu = 0.1$ and Gaussian envelope (Rayleigh) $y(t)$ process with equivalent variance. ................................................................. 23

3.1 A Comparison of the K-distribution (solid line) and Gamma (dashed line) densities for $\nu = 0.1, \beta = 0.24$ (top left),$\nu = 1, \beta = 1.61$ (top right),$\nu = 2, \beta = 2.27$ (bottom left), and $\nu = 3, \beta = 2.61$ (bottom right). ................................................................. 38
3.2 Comparison of the theoretical and empirical characteristic function of K-distribution for $\alpha = 1$ and $\nu = 1$ and its studentised version. ................................................................. 42
3.3 Estimated bias of the estimates of $\nu$ for $N = 128$. .......................... 45
3.4 Estimated bias of the estimates of $\nu$ for $N = 256$. .......................... 45
3.5 Estimated bias of the estimates of $\nu$ for $N = 512$. .......................... 45
3.6 Variance of the estimators for $N = 128$. .......................... 46
3.7 Variance of the estimators for $N = 256$. .......................... 46
3.8 Variance of the estimators for $N = 512$. .......................... 46
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>The Log of Characteristic Function of K-distribution (quadrature components with $a = 1$ and $\nu = 1$). (-) Theory, (·) Empirical. 52</td>
</tr>
<tr>
<td>4.2</td>
<td>Estimated Bias of the Estimates of $\nu$ for $N = 100$. 53</td>
</tr>
<tr>
<td>4.3</td>
<td>Estimated Bias of the Estimates of $\nu$ for $N = 200$. 53</td>
</tr>
<tr>
<td>4.4</td>
<td>Standard Deviation, $\sigma$ of the Estimators for $N = 100$. 54</td>
</tr>
<tr>
<td>4.5</td>
<td>Standard Deviation, $\sigma$ of the Estimators of $\nu$ for $N = 200$. 54</td>
</tr>
<tr>
<td>4.6</td>
<td>Block diagram of LOG/CFAR receiver. 56</td>
</tr>
<tr>
<td>4.7</td>
<td>A Comparison of theoretical PDF and sample histogram of K-distributed RVs for a fixed $\nu$ and variable scale parameter. (H) Histogram (T) Theoretical. 57</td>
</tr>
<tr>
<td>4.8</td>
<td>Theoretical PDF and histograms of log K-distributed RV for a fixed $\nu$ and variable scale parameter. (H) Histogram (T) Theoretical. 59</td>
</tr>
<tr>
<td>4.9</td>
<td>Empirical characteristic function of RV $Y$. 59</td>
</tr>
<tr>
<td>4.10</td>
<td>Theoretical PDF and histogram of two K-distributed RVs after transformation 4.16. 61</td>
</tr>
<tr>
<td>4.11</td>
<td>Plot of variance of $Z$ against $\nu$. 62</td>
</tr>
<tr>
<td>4.12</td>
<td>Realisation graph of ECF and fixed KCFE. Thick lines denote the theoretical CF. 67</td>
</tr>
<tr>
<td>4.13</td>
<td>Comparison of variance of the estimator. 69</td>
</tr>
<tr>
<td>4.14</td>
<td>The mean square error of the estimates of $\nu$ for $N = 128$. 75</td>
</tr>
<tr>
<td>4.15</td>
<td>The mean square error of the estimates of $\nu$ for $N = 256$. 75</td>
</tr>
<tr>
<td>4.16</td>
<td>A comparison of the estimates of $\nu$ for $N = 128$. 77</td>
</tr>
<tr>
<td>4.17</td>
<td>A comparison of the estimates of $\nu$ for $N = 256$. 77</td>
</tr>
<tr>
<td>4.18</td>
<td>The Estimated Variance of the estimates of $\nu$ for $N = 128$. 78</td>
</tr>
<tr>
<td>4.19</td>
<td>The Estimated Variance of the estimates of $\nu$ for $N = 256$. 78</td>
</tr>
<tr>
<td>4.20</td>
<td>The mean square error of the estimates of $\nu$ for $N = 128$. 80</td>
</tr>
<tr>
<td>4.21</td>
<td>The mean square error of the estimates of $\nu$ for $N = 256$. 80</td>
</tr>
<tr>
<td>4.22</td>
<td>Independent trials of the estimates of $\nu$ for $N = 32.\Delta$-FLOM, o-Raghavan's and x-KCFE. 81</td>
</tr>
<tr>
<td>4.23</td>
<td>Independent trials of the estimates of $\nu$ for $N = 64.\Delta$-FLOM, o-Raghavan's and x-KCFE. 82</td>
</tr>
</tbody>
</table>
5.1 Plot of 2-dimensional feature vectors and curves of equal Euclidean distance and equal Mahalanobis distance. 87
5.2 Comparison of two PDF having same mean and different variance. 88
5.3 The operation of empirical characteristic function. 89
5.4 A comparison of distance between outlier and ECF for $\mathcal{N}(1,1)$ and $\mathcal{N}(1,4)$. 90
5.5 Contour of equal distance for bivariate data. Proposed method is on the left and Mahalanobis is on the right. 93
6.1 The first six frame of the original image with its corresponding groundtruth. 96
6.2 The segmented image of the first frame. 97
6.3 Generation of a single 4-dimensional data sample $x$ from a 2D image patch, where $r = 2$ and $c = 2$. 99
6.4 Histogram of the feature vectors (-) from frame=1 and band=1. Comparison is made with K-distribution (-) and Gaussian(---). 104
6.5 The empirical characteristic function of the feature vectors (-). Comparison with theoretical K-distribution (-) and Gaussian(---). 105
6.6 The results for the first 8 frames obtained using the proposed method (5 frames temporal average). 107
6.7 The results for the first 8 frames obtained using Mahalanobis distance(5 frames temporal average). 108
6.8 The results for the first 8 frames obtained using the proposed method (3 frames temporal average). 109
6.9 The results for the first 8 frames obtained using Mahalanobis distance (3 frames temporal average). 110
6.10 Plots of false positive. 111
List of Tables

4.1 Sample mean and sample variance of the estimates of $\nu$ for $N = 128$ by using moments method. ................................................................. 73

4.2 Sample mean and sample variance of the estimates of $\nu$ for $N = 256$ by using moments method. ................................................................. 74
Glossary

Abbreviation

ANN  Artificial Neural Networks
ATD  Automatic Target Detection
ATR  Automatic Target Recognition
CF   Characteristic Function
CFAR Constant False Alarm Rate
DERA Defence Evaluation and Research Agency
ECF  Empirical Characteristic Function
EM   Expectation Maximisation
FLIR Forward Looking Infrared Radar
FLOM Fractional Lower Order Moment
GBK Generalised Bessel K Function
IID  Independent and Identically Distributed
KCFE Kernel Characteristic Function Estimator
LANTIRN Low Altitude Navigation and Targeting Infrared-for-Night
MGF  Moment Generating Function
ML   Maximum Likelihood
MOM  Method of Moments
MSE  Mean Square Error
NM   Numerical Maximisation
PCA  Principal Component Analysis
PDF  Probability Density Function
PR   Pattern Recognition
ROI  Region of Interest
RV/RVs Random Variable(s)
SAR  Synthetic Aperture Radar
SFFS Sequential Floating Forward Search
SFS  Sequential Forward Search
Symbol

\( p(x) \) — probability of \( x \)
\( m \) — number of dimension
\( \mu \) — mean
\( \Sigma \) — covariance matrix
\( \exp(\cdot) \) — exponential; \( e^x \)
\( \mu_k \) — moments of order \( k \)
\( T \) — matrix transpose
\( p(x|y) \) — conditional probability of \( x \) occurring given \( y \)
\( x \) — \( K \)-distributed RV (unless otherwise stated)
\( \ln \) — natural logarithm; \( \ln(e^x) = x \)
\( a \) — scale parameter of \( K \)-distribution
\( \nu \) — shape parameter of \( K \)-distribution
\( \Gamma(\cdot) \) — gamma function
\( K_\lambda(\cdot) \) — modified Bessel function
\( \kappa \) — \( K \)-distributed quadrature component
\( \kappa_\perp \) — orthogonal component of \( \kappa \)
\( \phi(t) \) — characteristic function
\( \langle \cdot \rangle \) — expectation operator
\( \mathbb{R} \) — the set of real number
\( j \) — \( \sqrt{-1} \)
\( \varepsilon \) — member
\( F(x) \) — distribution function
\( \mu_k \) — \( k^{th} \) order moment
\( \tilde{\phi}(t) \) — empirical characteristic function
\( N \) — number of sample
\( \psi(\cdot)^{(\lambda)} \) — polygamma function of order \( \lambda \)
\( \gamma \) — Euler's constant: 0.5772...
\( \mu_a \) — arithmetic mean
\( \mu_g \) — geometric mean
\( \psi(\cdot) \) — trigamma function
\( y \) — log of \( K \)-distribution (unless otherwise stated)
\( F(a, b, c; x) \) — Gauss hypergeometric function
\( z \) — proposed scale invariant log of \( K \)-distributed RV
\( \tau \) — \( 1/\nu \)
\( \Psi(t) \) — second characteristic function; \( (\log \phi(t)) \)
\( \Im \) — imaginary part of complex number
\( \Re \) — real part of complex number
\( \mathcal{K}(a, \nu) \) — \( K \)-distributed RV with parameter \( a \) and \( \nu \)
\( \mathbf{H} \) — hadamard matrix
\( \mathbf{I} \) — identity matrix
Chapter 1

Introduction

Pattern recognition (PR) is an evolving technology since its birth in the mid twentieth century. Since the beginning of the new millennium, the success of this technique is seen in many applications. They include human speech and character recognition by computers, and identification of people from fingerprints, hand shape and size, retinal scans, voice characteristics, typing patterns and handwriting.

In geology, pattern recognition has been applied in the classification of seismic signals for oil and minerals exploration, and earthquake prediction. With the improvement of the satellite technology, pattern recognition commenced providing an automatic tool to determine the type and condition of agricultural crops, weather prediction and water reserves.

Automatic inspection of parts on an assembly line and automatic grading of plywood, steel and other sheet material are some examples of pattern recognition applications in industries. In medicine, pattern recognition techniques have been used in classification of electrocardiograms into diagnostic categories of heart disease, automated analysis of medical images obtained from various sources and all other medical waveform analysis.
Chapter 1. Introduction

As an important element of machine intelligence, pattern recognition is concerned with the automatic description or classification of measurements, extracted from the objects of interest, which are generally called patterns. The measurements used to classify the patterns are called features which could be symbolic, for example colour, or numerical such as length.

In some cases, features can be a combination of both, whether it be a low-level data or extracted from the input data by applying feature extraction algorithms. As features are intended to be fewer in number than the observations, they are subject to a process called feature selection in order to reduce the dimensionality of the problem data without discarding valuable information.

One of the approaches to pattern recognition that have been used extensively is the statistical one. As its name implies, statistical pattern recognition uses a probabilistic model to underpin classification algorithms, where the most widely used model is the Gaussian distribution.

Usually, a Gaussian assumption is made without a reason to believe that a given feature is approximately normally distributed, but merely due to its convenience as it often leads to analytically tractable results. Some even argue that since features often result from the combination of a large number of independent observations, the Central Limit Theorem applies and therefore Gaussianity. However, the real world never fits the model exactly and in some natural environments, Gaussian model may not be at all appropriate.

Therefore it is necessary to look beyond the oversimplified Gaussian assumption and work with a more realistic non-Gaussian models. In the instance where a normal distribution does not fit well, densities that have heavier tails may be useful. A number of models have been proposed for such impulsive phenomena, either by fitting experimental data or based on physical grounds.
1.1. Automatic Target Recognition

Although it may come at a price of more complicated algorithms, a well fitted model may improve the performance of a system.

1.1 Automatic Target Recognition

1.1.1 Overview

Automatic target recognition (ATR) is one important application of pattern recognition. ATR or sometimes referred to as automatic target detection (ATD), is a process to detect, track and/or recognise targets from a cluttered background. With the current development of sensors which make possible to view the battlefield at night or even during severe weather conditions, soldiers are overloaded with a large amount of data and a demand to make rapid decisions.

Generally accredited to both autonomous and aided recognition, ATR is aimed at reducing the workload and fatigue of human operators. Moreover, target detection and classification problem increases in difficulty when the targets are small while the surveillance area is large. In aided recognition (or cueing), human interacts with the system and makes some of the decisions. Usually, the acquisition is done by the targeting system but ultimately recognition is done by a person [4].

This is unsuitable to be implemented in a battlefield situation where human's survivability is low, as a system with human operator in loop is generally slow, unreliable and vulnerable. Such system if implemented, may limit the overall performance or mission in real situation [93]. Thus, methods are sought to perform ATR autonomously. Acceptable autonomous operation is still an unattainable goal, which when accomplished, would be one of the key components of future unmanned vehicle mission.
Chapter 1. Introduction

In an ATR system, target acquisition and classification are done by a computer processing of data, usually in a form of image sequence. The sequence of images are extracted from a wide variety of sensors and platforms. Forward looking infrared (FLIR), synthetic aperture radar (SAR), laser radar, millimetre wave radar, multi-spectral or hyperspectral sensors, low-light television and video are some examples of sensors currently being used. These sensors are normally attached from different platforms such as ground vehicle, aircraft, missile, ship and satellite.

Extracted images are then subject to numerous processing steps and algorithms, involving a wide range of tools such as image processing, pattern recognition and artificial intelligence. The generic problem of an ATR system is to take information from one or more sensors, and combine it with a priori information. A decision is then made about the type of targets present in the scene, usually prioritised by their tactical importance [4]. Recently, the knowledge-based approach has been included in the processing steps to solve ATR adaptively [3].

It has been established that ATR is multidisciplinary field which requires diverse technology and expertise in sensors, processing algorithms, architectures, and evaluation of hardware and software systems [4]. An ideal system should keep the false alarm rate to its minimum, while maintaining high detection of a true target. Another important challenge for ATR is to insert an unforeseen target rapidly and retrain itself when necessary. The algorithms should be invariant to a natural or man-made clutter and complex varying background. Nonetheless, the ultimate challenge for an ATR system is to evaluate the algorithms in real time.

ATR technology is applicable to a wide range of applications. They include assessment of battlefield situations, low altitude navigation and targeting infrared-for-night (LANTIRN) system, fire-and-forget and lock-on-after-launch
missile and target surveillance over land, sea and air. Although much work in this field had been focusing on defence applications (since military is the origin of the ATR concept), it can also be applied for many non-military purposes. Another area that is worth considering are automatic fingerprint and face recognition, photo interpretation, robot vision and police surveillance.

1.1.2 Pattern Recognition Approach for ATR

As already indicated, pattern recognition techniques are an important component of an ATR system and are used for both data preprocessing and decision making. The method is based on the hypothesis that target features lie in an easily separable regions of a multidimensional feature space from its background. Statistical and syntactic approaches have been used in the past, although there is growing interest in applying neural network methods, especially when there is lack of suitable statistical or structural model.

A processing flow chart of a typical ATR implementation is shown in Figure 1.1. First, the scene is captured by some sensor(s) and converted into a signal to be processed. Depending on the type of sensors, whether it's a digital or an analogue type, an appropriate processor is used to process the data. Typically, a digital electronic processor is implemented, although, an analogue signal from an optical sensor can be easily converted into digital by an analogue to digital converter.

This signal is then subjected to various pattern recognition processing steps which is shown in the figure by the blue background. The processor may also receive an auxiliary data such as the range from a laser range finder, position coordinates from Global Positioning System, and weather data [70].
1.1.3 Preprocessing

Preprocessing is an image enhancement process, where depending on the captured scene, it may or may not be required. In this stage, target contrast is improved and noise and/or clutter present in the image is reduced. Functions chosen to enhance the image are usually specific for a particular application, as a method that is useful for one application may not be suitable for another.

The process of image enhancement is divided into two major categories; a spatial domain method which is based on a direct manipulation of pixels in an image plane, and, a frequency domain approach which is based on modifying the Fourier transform of an image.

Examples of spatial domain methods include grey level transformations, histogram processing, image subtraction or averaging, and many methods derived from spatial filtering. Image sharpening by using highpass filter is an example.
of frequency domain preprocessing. At the end of this process, the separability between the target and the background is expected to be increased so that the effectiveness of subsequent processing steps is enhanced.

1.1.4 Target detection

The target detection stage is a localisation process of those areas in the image, where a potential target is likely to be present. At this stage, we are interested in determining and analysing features of target signatures, which are separable from the background. Typically, target detection is based on some sort of threshold, determined by the contrast of an object compared to the local background in an arbitrary box drawn around the object. The detection problem is then solved by accurately modelling the background as a stationary random process.

The probability of detection at this stage is usually set to be as close to 100% as possible to ensure that targets are not missed. This is because, if a target is missed in this process, it will be missed altogether. Consequently, the probability of false alarm is allowed to be comparatively high. Most of the techniques for the target detection can be adapted to detect either light or dark targets, where in FLIR images, targets which are hotter than the background generally appear as bright contrasting objects [3]. In some techniques such as 'superslice' [51], localisation and segmentation are inseparable.

A few examples of methods that have been used in the past include double window filtering [13, 65], statistical modelling and Probability Density Function (PDF) estimation [78], spoke filtering [52], intensity, edge and range information analysis [5], intensity and texture measure extraction [53], and linear discriminant analysis [76]. The extension of the mode seeker technique [57] called 'superspike' has been reported to produce good results.
Several performance measures have been proposed in [3], which include:

1. Probability of target detection - a target is said to have been detected if its centroid lies within a small window centred at the centroid of the true target.

2. Computational efficiency - ratio of localised area to the image area.

3. Probability of False alarm - the number of localised areas to the actual number of targets present in the image

1.1.5 Segmentation

The false alarm rate after the target detection stage will be unacceptably high within any real world ATR. Thus it will be necessary to perform segmentation in order to reduce the false alarm rate before any target classification can be made. Even in the most homogeneous terrain, such as grassland, a number of false alarms will occur due to image noise and speckle. False alarm also arises from manmade clutter such as vehicles or buildings.

The segmentation operation extracts the target from the background as accurately as possible after it has been detected. Typically, the first step in the segmentation process is performed by using a basic edge-finding operators on the Region of Interest (ROI). The edge segments are then logically connected and gaps are filled to form a continuous line around the presumptive target. This is achieved by selecting the threshold such that the coincidence between the thinned edge and the border of the connected components is maximised. Finally the region is converted to a binary image by assigning a high-bit value to all pixels inside the line.

In addition to edge magnitude, Minor and Sklansky [52] proposed a method that utilises the edge direction. Brown and Frei [12] use a sequential region
1.1. Automatic Target Recognition

Figure 1.2: FLIR Images after segmentation. Left-the binary segmented image. Right-the grey value image.

growing technique which uses size and rate of growth to reject nontargets, Resenfeld et al [73] and Bhanu et al [3] use relaxation methods while Chen and Yen [14] use Fisher's linear discriminant to segment by pixel classification.

Segmentation is one of the key components in determining ATR's success. For example, blob-like shapes in FLIR images with intensity gradient across targets (partly hot, partly cold), are not extracted as a single blob. This would cause distortion of shape features and would result in high false alarm rate, particularly in cluttered scenes. Figure 1.2 shows an example of a segmented image of a ship from a FLIR sensor. To evaluate the segmentation operations, the following quantitative measures have been proposed [3]

1. Target pixel misclassification.

2. Correlation coefficient between the true and extracted target.

3. Mean square error between the true and extracted target.

4. Shape difference between the true and extracted targets.

5. Object to background contrast, intensity difference and Bhattacharyya distance between the true target and clutter objects.
1.1.6 Feature Computation

After segmentation, a set of features is computed for each object. Firstly, features are extracted from the segmented target to reduce the processing load in the decision making step. The feature space is usually made up of a set of basis vectors that correspond to certain defined measurable quantities on the candidate targets [3].

Various shape, grey scale and projection features are commonly used while semantic features such as geographical and temporal context have been used only to a very limited extent [3]. In addition, feature extraction algorithms such as Principal Component Analysis (PCA) had been utilised where the target is then represented in the feature space as a vector whose components are a linear combination of the values of the defined quantities that have been measured.

The second part of the step is feature selection which is a key to good discrimination. Feature selection is an optimisation technique that reduces memory requirements of the classifier. The primary goal of feature selection is to obtain features which maximise the similarity of objects in the same class, while maximising the dissimilarity of objects in different classes [3]. Features are usually assumed to be uncorrelated, since statistical dependency may exclude the best feature from the best subset. In the context of the ATR problem, feature selection has been performed by histogram examination [51], Bhattacharyya measure [3], Kolmogorov-Smirnov test [75], F-statistic, exhaustive scheme [82], physical reasoning and linear regression technique [80].

There are many different automatic feature subsets selection methods. Since the feature spaces of an ATR problem is usually large, traditional optimal search methods are impractical to be implemented, due to their prohibitive computational cost and the nesting effect. Probably, the most effective fea-
feature selection technique is the sequential floating search methods. This non-exhaustive method has been a popular choice even though the technique is not optimal.

1.1.7 Classification and Identification

After features have been extracted and selected, the next step is classification. At this stage, a target signature is distinguished from those resulting from the clutter or non-target objects, by means of a classifier. Examples of classification methods that have been used in the past are Bayesian, K-nearest neighbour algorithm, Parzen, linear and quadratic classifiers, structural classifier, tree-based classifier and clustering technique. The performance of a classifier is measured by the probability of successful classification and the false alarm rate [3].

Classification is subsequently followed by a target recognition and identification stage. This stage is concerned with the precise identification of the target, where, depending on the differences between various target signatures, a type of target can be distinguished from one to another. However, target can be hard to recognise and identify due to the variability of target signatures arising from a number of factors such as target location and orientation. Since generally these factors are not known a priori, the system must be prepared to detect and recognise targets at all positions and orientations. Thus determining, measuring and storing all the variations in target signatures and data fusion techniques are required.

After targets have been identified, the targets of interest are annotated on the final image to the operator in terms of their location and description. A decision is then made either by humans or a computer, depending on the type of mission or vehicle.
Chapter 1. Introduction

1.2 Research Interests, Aims and Objectives

Pattern recognition has played an important role in ATR systems. Even in the context of the recently developed knowledge-based approaches, the classical PR method is still needed before Artificial Intelligence techniques can be applied for symbolic representation and reasoning. Due to the lack of structural information in ATR systems, its statistical counterpart has been a preferred method.

Perhaps the most vital stages in statistical pattern recognition are the feature computation and target detection, which are of the main interest in our research. Numerous statistical methods have been proposed for these stages, but in most problems, proposed solutions rely on a Gaussian assumption. Gaussian models have been chosen not only due to their mathematical tractability, but also because PR community lack readily available methods for parameter estimation and distance function computation, if a non-Gaussian assumption is made. The assumption of non-Gaussian models also leads to extra computational costs which is another reason why it is not attractive.

The aim of this research is to use a non-Gaussian distribution in order to develop statistical pattern recognition techniques. K-distribution is one of the long-tailed densities and is known in the signal processing community for fitting the radar sea clutter accurately. The thesis presents the efforts made to define the background features, extracted from the sea images, by using a K-distribution model. This technique is then applied to the automatic detection of targets.

Density functions, described by mathematical formulas, always involve a set of parameters whose values determine the size and shape of the density. In many problems in pattern recognition, the form of the densities may be known or in most cases assumed. However, the values of the associated parameters
are usually not known. For the traditional Gaussian density, the parameters are the mean vector and the covariance matrix, which is certainly not the case for K-distribution. The first step for a successful application is thus to propose a new method for parameter estimation of K-distribution which is the backbone of this thesis.

It is frequently useful to define the distance between two points in a vector space, or a normal distance between a point and a surface which in pattern recognition is known as the discriminant function. Usually, as a Gaussian assumption is made, Mahalanobis distance is used in the classification procedure as a distance measurement. The probability density function of a K-distribution is however of a rather complicated form. A method is thus derived by using the ECF, since K-distribution is more conveniently described in terms of its characteristic function.

1.3 Outline of the Thesis and Contributions

This thesis is made up of seven chapters.

The first chapter introduced the motivation behind the research, where the terms and applications that are usually associated with pattern recognition were briefly described. In the second part of the chapter, an introduction to Automatic Target Recognition was presented. The various processing steps in the pattern recognition approach were portrayed together with representative methods that had been proposed in the past for each step. The research interest, aims and objectives were set out along with the proposal of applying a non-Gaussian technique in Automatic Target Recognition, a field which relies heavily on the success of pattern recognition.

The next chapter, entitled 'Statistical Modelling and Characteristic Func-
Chapter 1. Introduction

The chapter is intended as an overview of the approach that we are seek to develop. The chapter starts with a brief introduction to the traditional approach based on Gaussian models and how it leads to a measure called Mahalanobis distance. The chapter follows with a description on how the impulsive nature at sea leads to a distribution model which does not adhere to the Normal assumption. The K-distribution which has been used to model the sea clutter is introduced. The characteristic function which is the Fourier transform of a probability density function is then presented together with its basic properties. The empirical characteristic function which is its sample counterpart is also presented in this chapter.

Chapter three reviews the estimation techniques that has been proposed in the past to estimate the parameters of the K-distribution. Since the maximum likelihood solution of the K-distribution has to be solved numerically, many alternative methods have been proposed. It is important to note that the formulae in this chapter may differ from the original published papers, so as to agree with the K-distribution density derived in Chapter three. The unified formalism this chapter provides, can be a very useful reference for anyone wishing to pursue a research in this field. The original formulation of the previous methods is depicted in Appendix B. In order to provide a fair evaluation, the methods described in this chapter have been simulated using the Monte Carlo approach, and the results are presented.

A novel approach to K-distribution parameter estimation is presented in Chapter four. The chapter starts with a motivation behind the development of a method proposed to estimate the shape parameter when the scale parameter is known a priori. A novel approach to eliminate the scale parameter is then presented, resulting in two new approaches to estimate the shape parameter. Simulation results of the proposed estimators are then presented and a comparison is made with the previously published methods. The comparison
shows that the proposed method performs better in terms of bias and variance of the estimator.

The fifth chapter starts with a brief introduction to the Euclidean and Mahalanobis distance metric. A comparison is made between the two metric and the problems arising are highlighted. A novel distance function which is derived from the characteristic function is then discussed. The method is first developed for the univariate case and then extended to the multidimensional case. The proposed distance function not only takes into account the population mean and variance (or covariance in the multidimensional case) but also the characteristic function of the model distribution.

In Chapter six, the proposed method for estimating the K-distribution is applied to target detection problem. A comparison with the Gaussian model shows that the proposed method results in a lower false alarm rate and a higher target detection rate. In a long run, the proposed method also reduces the processing time require by the conventional distance function. The number of frames used in the temporal averaging stage can also be reduced, thus producing the results faster.

Finally, the thesis is concluded in Chapter seven. The capabilities and limitations of the proposed method are presented. Further, some perspectives for future work are suggested.
Chapter 1. Introduction
Chapter 2

Statistical Modelling and Characteristic Function

A statistical approach provides a systematic framework for integrating prior knowledge about the scene and targets with observational models. This technique is based on the probability models usually containing adjustable parameters that must be estimated from the data. In principle, it is possible to compute optimal solution to the problem of detection, classification and parameter estimation once accurate statistical models have been identified. A decision strategy or classifier is then designed to integrate all available problem information, such as measurements and a prior probabilities.

This chapter starts with brief introduction to the conventional assumption of Gaussian distribution and how it leads to a measure called Mahalanobis distance. Since impulsive nature usually leads to a distribution with a heavier tail, K-distribution which have been used to model the sea clutter is presented next. The theory of characteristic function and its sample counterpart are also presented, as we exploit their properties during the derivation of our new parameter estimation technique and distance function.
2.1 Gaussian Distribution as a Probabilistic Models

In feature selection and classification, the assumption of Gaussian distribution of the data is generally made, even though it may not have a multivariate Gaussian distribution. A multidimensional Gaussian distribution is given by:

\[
p(x) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp \left[ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right]
\]  

(2.1)

where \( x \) is \( m \)-dimensional with mean vector \( \mu \) and covariance matrix \( \Sigma \). The class dependence is obtained parametrically via class specific mean vectors, \( \mu_i \) and covariance matrices \( \Sigma_i \).

In statistical PR, we concentrate on developing a decision rule which may be formulated in several interrelated ways. Usually, Bayes theorem is used in formulating the classifier where this is done by converting a priori class probability \( P(w_i) \) into a posteriori probability \( P(w_i|x) \). Classification in this manner is based on finding the discriminant function \( g_i(x) \) for which \( p(x|w_i) \) is the largest. In the Gaussian case, with the class dependence introduced in 2.1 via \( \mu_i \) and \( \Sigma \), the discriminant function is given by

\[
g_i(x) = -\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) - \left( \frac{m}{2} \right) \ln(2\pi) - \frac{1}{2} \ln|\Sigma|.
\]  

(2.2)

The second and third terms of equation 2.2 are class-independent constant biases and may be eliminated, thus classification through this procedure is influenced by the squared distance of feature vector \( x \) from the mean vector \( \mu_i \), weighted by the inverse of the covariance matrix \( \Sigma_i^{-1} \). The quantity

\[
d_i^2 = (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)
\]  

(2.3)
is known in the PR community as the Mahalanobis distance, and has been used in many areas as a measure of similarity.

The use of the Mahalanobis distance removes several of the limitations of the Euclidean metric, where it is correct for correlation between the different features, automatically accounts for the scaling of the coordinate axes and it can provide curved as well as linear decision boundaries. In the special case where the features are uncorrelated and the variances in all directions are the same, the Mahalanobis distance becomes equivalent to the Euclidean distance.

2.2 Radar Sea Clutter

The detection performance of maritime radars is often limited by the unwanted sea echo or clutter. Sea echo is the vector sum of scattering from the sea surface within the illuminated area. The movement of the scatterers (waves, ripples, etc.) causes a change in the relative phases of their separate echoes and a resulting change in the total echo. The echo is often noise-like, as would be expected from a collection of randomly moving scatterers.

Sometimes the echo appears to resolve itself into distinct point target-like echoes with regions of very low signal occurring between these echoes. For low resolution and high grazing angles, the clutter amplitude can be modelled by Rayleigh distribution. However, as the radar resolution increases and for smaller grazing angles, the clutter returns are often described as becoming 'spiky', thus the amplitude distribution developing a longer 'tail'. The temporal and spatial correlation characteristics of the clutter also change.

The amplitude statistics have been found to fit other distributions namely Lognormal, Weibull, Contaminated-Normal, Log-Weibull and K-distribution.
2.2.1 K-Distribution

It has been found from practical measurements that the sea for high resolution radar and low grazing angles can be well modelled by two components [94]. The first component is a spatially varying mean level $y$ that results from a bunching of scatterers associated with the sea swell structure. The second component, termed the 'speckle' component, occurs due to the multiple scatterer nature of the clutter in any range cell and has fast fluctuation [94]. Based on the two components, the overall amplitude sea clutter distribution $p(x)$ is derived by averaging the speckle component over all possible values of the local mean level [95]:

$$p(x) = \int_0^\infty p(x|y)p(y)dy, \quad 0 \leq x \leq \infty \quad (2.4)$$

where the speckle component $p(x|y)$ is Rayleigh distributed

$$p(x|y) = \frac{\pi x}{2y^2} \exp \left( -\frac{\pi x^2}{4y^2} \right), \quad (2.5)$$

and $p(y)$ is the PDF of the clutter mean level, which has been found to be a good fit to the Chi family of amplitude distributions

$$p(y) = \frac{2b^{2\nu}y^{2\nu-1}}{\Gamma(\nu)} \exp \left( -b^2y^2 \right) \quad (2.6)$$

where $b$ is the scale parameter and $\nu$ is a shape parameter. Substituting 2.5 and 2.6 into 2.4 yields the K-distribution [95]

$$p(x) = \frac{2}{a\Gamma(\nu)} \left( \frac{x}{2a} \right)^\nu K_{\nu-1} \left( \frac{x}{a} \right) \quad x > 0, \quad (2.7)$$
2.2. Radar Sea Clutter

where $K_{\nu}(\cdot)$ is the modified Bessel function of order $\lambda$, $a = \sqrt{\pi}/2b$ is a scale parameter and $\nu$ is the same shape parameter as in equation 2.6. A few sample plots of the PDF for unit second moment ($a = 1/(2\sqrt{\nu})$) and for various $\nu$ are shown in Figure 2.1. For high resolution sea clutter, values of $\nu$ are generally observed in the region $0.1 \leq \nu \leq \infty$, where $\nu \sim 0.1$ represents very spiky clutter and $\nu = \infty$ represents thermal noise. As $\nu \to \infty$ the density in 2.7 tends to a Rayleigh density.

2.2.2 Quadrature Components

Two orthogonal components $x(t)$ and $x(t)$ of $K$-distributed clutter $x(t)$ can be represented by a compound complex Gaussian random process as [85]

$$x(t) = x(t) + jx(t) = \xi(t)[g(t) + jg(t)] \quad (2.8)$$

where $j$ denotes $\sqrt{-1}$, and $g(t)$ and $g(t)$ are independent Gaussian processes with zero mean and unit variance. The characteristic function of the quadrature component $x$ (or that of $x(t)$) can be determined from [83]

$$\phi_\nu(t) = \int_0^\infty \phi_\nu(t\xi)p(\xi) d\xi = \left(\frac{1}{1 + \alpha^2t^2}\right)^\nu \quad (2.9)$$

where $\phi_\nu(t) = \exp(-t^2/2)$ is the characteristic function of a standard Gaussian distribution, and the modulating process, $\xi(t)$ has the PDF of the form [85]

$$p(\xi) = \frac{2\alpha^\nu\xi^{2\nu-1}}{\Gamma(\nu)} \exp\left(-\alpha^2\xi^2\right) \quad \xi > 0, \quad \alpha > 0 \quad (2.10)$$

The inverse Fourier transform of $\phi_\nu(t)$ yields the univariate $K$-distribution PDF

$$p(\nu) = \frac{1}{a\pi\Gamma(\nu)} \left(\frac{|\nu|}{2a}\right)^{\nu-\frac{3}{2}} K_{\nu-\frac{1}{2}} \left(\frac{|\nu|}{a}\right). \quad (2.11)$$
Figure 2.1: The amplitude PDF of K-distribution.

Figure 2.2: The PDF of univariate K-distribution.
2.2. Radar Sea Clutter

Figure 2.3: A comparison of univariate K-distribution $x(t)$ process with $\nu = 0.1$ and Gaussian $y(t)$ process with equivalent variance.

Figure 2.4: A comparison of K-distribution $x(t)$ process with $\nu = 0.1$ and Gaussian envelope (Rayleigh) $y(t)$ process with equivalent variance.
Figure 2.2 shows the plots of the PDF 2.11 for unit second moment and different values of $\nu$. It is interesting to note that when $\nu = 1/2$, the PDF 2.11 is simply the PDF of the product of two independent Gaussian random variables (RVs) and when $\nu = 1$, $p(x)$ has a Laplacian distribution. Figures 2.3 and 2.4 show the comparison of K-distribution and Gaussian distribution random process. For the same value of mean and variance, the random process of K-distribution is more spiky than the Gaussian. This is described by the PDF 2.7 and 2.11 which have a longer tail than Gaussian. The following conditions are known to result in echo PDFs that have longer tails [95]:

- Smaller areas of the resolution cell (higher resolution)
- Lower grazing angles of incidence typically less than 3°
- Rougher sea states
- Horizontal polarisation of transmitted and received signals

2.3 Characteristic Function

Recently, there is growing interest in applying methods using Characteristic Function (CF) among the signal processing processing community. The interest stems from the need to apply signal models more complex than the Gaussian Model. A non-Gaussian model such as $\alpha$-stable distributions does not have a closed-form density function, except for the special case of Gaussian, Cauchy and Inverse-Gaussian distribution.

In other cases, signals and noise do not have PDF in easily tractable forms but the same signals are often conveniently characterised through the characteristic function. This suggests that such properties should or could be tested.
2.3. Characteristic Function

either more conveniently, or solely through the use of sample or empirical characteristic functions rather than sample distributions or densities.

Let $X$ be a real-valued random variable with distribution function $F(x)$. The characteristic function $\phi(t)$ of the distribution function $F(x)$ (or of the RV $X$) is by definition

$$\phi(t) \triangleq \int_{-\infty}^{\infty} e^{itx} dF(x) = \langle e^{itX} \rangle, \quad t \in \mathbb{R} \quad (2.12)$$

where $\langle \cdot \rangle$ is the expectation operator. If $F(x)$ is absolutely continuous with density $p(x)$, then $\phi(t)$ is the Fourier transform of $p(x)$:

$$\phi(t) = \int_{-\infty}^{\infty} p(x) \cdot e^{itx} \, dx. \quad (2.13)$$

The theorem of the characteristic function can be summarised as follows

1. Any CF $\phi(t)$ is uniformly continuous and satisfies the following conditions:
   
   (a) $\phi(0) = 1$
   
   (b) $|\phi(t)| \leq 1$ for all real $t$ (boundedness theorem)
   
   (c) $\phi(-t) = \overline{\phi(t)}$ where the horizontal bar denotes the complex conjugate.

2. Two distribution functions are identical if and only if their CFs are identical. From the property of the Fourier transform, the CF and the distribution correspond one to one (uniqueness theorem).

3. A characteristic function is real if and only if it is even.
4. A distribution function is symmetric if and only if its CF is real and even. A distribution function \( F(x) \) is called symmetric if \( F(x) = 1 - F(-x) \).

5. The CF of the convolution of distribution function is the product of the corresponding CFs. Suppose \( X \) and \( Y \) are two independent RVs with CFs \( \phi_X(t) \) and \( \phi_Y(t) \), the CF of \( X + Y \) is \( \phi_X(t)\phi_Y(t) \).

The uniqueness and boundedness are specific properties of the CF. The moment generating function (MGF) which is the Laplace transform of the PDF is similar to the characteristic function, but it is not bounded, and there are some pathological examples of distributions which for instance, have all degrees of moments, but do not have the MGFs [90].

Particularly the boundedness is an advantage for stability of numerical calculations. The equation 2.13 may be expanded in power series as

\[
\phi(t) = \int_{-\infty}^{\infty} p(x) \, dx + j t \int_{-\infty}^{\infty} x \, p(x) \, dx + \frac{1}{2} (jt)^2 \int_{-\infty}^{\infty} x^2 \, p(x) \, dx + ... \\
= \sum_{k=0}^{\infty} \frac{(jt)^k}{k!} \mu'_k \\
= 1 + j t \mu'_1 - \frac{1}{2} t^2 \mu'_2 - \frac{1}{3!} j t^3 \mu'_3 + ..., \quad (2.14)
\]

where \( \mu'_k \) is the \( k \)th moment about zero and \( \mu'_0 \equiv 1 \). The characteristic function can therefore be used to generate moments about zero through

\[
\phi^{(k)}(0) \equiv \left[ \frac{d^k \phi(0)}{dt^k} \right]_{t=0} = j^k \langle x^k \rangle \quad (2.15)
\]

The properties of the Fourier transform also allow CF to behave well under shifts, scale changes and summation of RVs, thus it has been used extensively in areas such as testing for goodness-of fit, testing for independence, and for parameter estimation.
2.4 Empirical Characteristic Function

The simplest estimator of the CF is the sample or empirical characteristic function (ECF) which is defined as

\[ \hat{\phi}(t) \triangleq \frac{1}{N} \sum_{i=1}^{N} \exp(jtx_i) \]  

(2.16)

where \( x_i; i = 1, 2, ..., N \) represent independent and identically distributed (IID) RVs with CF \( \phi(t) \). Obviously, the ECF is directly calculated from the empirical distribution and \( \hat{\phi}(t) \) is computable for all values of \( t \in \mathbb{R} \). At a given \( t \), \( \hat{\phi}(t) \) is an RV and \( \hat{\phi}(t), -\infty < t < \infty \) is a stochastic process. Some of the convergence properties of \( \hat{\phi}(t) \) are established by Feuerverger and Mureika [24], where for \( T < \infty \),

\[ P \left\{ \lim_{N \to \infty} \sup_{|t| \leq T} |\hat{\phi}(t) - \phi(t)| = 0 \right\} = 1 \]  

(2.17)

holds. Also, let \( Y_N(t) \) be a stochastic process that is a residual of the ECF and CF:

\[ Y_N(t) = \{ \hat{\phi}(t) - \phi(t) \} \sqrt{N}, \quad (-T \leq t \leq T). \]  

(2.18)

As \( N \to \infty \), \( Y_N(t) \) converges to a zero mean complex Gaussian process \( Y(t) \) satisfying \( Y(t) = \overline{Y(-t)} \) thus

\[ \langle Y(t)Y(s) \rangle = \phi(t + s) - \phi(t)\phi(s). \]  

(2.19)

ECF is an unbiased estimator of the corresponding CF and satisfies the following condition

1. \( \hat{\phi}(0) = 1 \)
2. \( |\hat{\phi}(t)| \leq 1 \)
2.5 Conclusion

In this chapter, the classical approach of Gaussian assumption for statistical modelling in pattern recognition has been narrated. In the design of the classifier, this approach leads to the derivation of Mahalanobis distance which has been used in many applications as a measure of similarity.

There has been a growing interest in applying a non-Gaussian model in the signal processing community to model noise or clutter present in the signal. Since some of these distributions are more conveniently characterised through its characteristic function, many methods such as parameter estimation have been developed from it.

However, the application of non-Gaussian model and methods derived from the characteristic function in pattern recognition are still in their infancy. This is because the assumption of Gaussian distribution always leads to many readily available methods. Furthermore, the assumption of non-Gaussian model to a method purposely built for Gaussian model will degrade the performance significantly.

Similar to the sample mean, empirical characteristic function is relatively easy to implement. Since ECF is the sample type estimator for the CF but on the other hand it is a model-free estimator, its application in pattern recognition could be very beneficial. Feature vectors could be modelled with a more complex and accurate density function, but still be fast computationally.
Chapter 3

K-Distribution Parameter Estimation: A Review

Estimating the parameters of a statistical distribution from measured sample values forms an essential part of many signal processing tasks. Accurate estimation of the order parameter $\nu$ is of great importance for a variety of radar detection and estimation problems in a K-distributed clutter environment, including CFAR target detection, coherent target detection, texture segmentation and clutter classification in SAR images [44].

This chapter reviews estimation techniques that have been proposed in the past to estimate the parameters of K-distribution. At the moment, an accurate estimate can only be achieved by using a very computationally intensive method, whereas the performance for a simpler method is unreliable especially for a small sample case. It is important to note that the formulae in this chapter may differ from the original published papers, so as to agree with the K-distribution density derived in Chapter three. All the methods described have been simulated using the Monte Carlo approach, and the results are presented and discussed.
3.1 Maximum Likelihood Methods

The most common method of deriving estimators is by using the method of Maximum Likelihood (ML). The essential feature of the principle of maximum likelihood, as it applies to estimation theory, is that one requires to choose, as an estimate of a parameter, that value for which the probability of obtaining an actually observed sample is as large as possible [54].

Suppose \(x_1, x_2, \ldots, x_N\), be \(N\) independent variables, where each is distributed according to 2.7. The log-likelihood function of the K-distribution is given by [34]

\[
L = -N \ln [(2a)^{\nu+1} \Gamma(\nu)] + \nu \sum_{i=1}^{N} \ln(x_i) + \sum_{i=1}^{N} \ln \left[ K_{\nu-1} \left( \frac{x_i}{a} \right) \right].
\] (3.1)

The partial derivatives of the log-likelihood function are given by

\[
\frac{\partial L}{\partial \theta} = \frac{(\nu + 1)N}{a} - \frac{1}{2a} \sum_{i=1}^{N} \frac{K_{\nu-2} \left( \frac{x_i}{a} \right)}{K_{\nu-1} \left( \frac{x_i}{a} \right)} \left( \frac{x_i}{a} \right)
\] (3.2)

\[
\frac{\partial L}{\partial \nu} = -N [\ln(2a) + \psi(\nu)] + \sum_{i=1}^{N} \ln(x_i) + \sum_{i=1}^{N} \frac{\partial}{\partial \nu} \left[ \frac{K_{\nu-1} \left( \frac{x_i}{a} \right)}{K_{\nu-1} \left( \frac{x_i}{a} \right)} \right]
\] (3.3)

where \(\psi(x)\) is the Digamma function which is defined as the derivative of \(\log(\Gamma(x))\) with respect to \(x\).

The ML estimates of the parameters \(\nu\) and \(a\) of the K-distribution can be found by equating 3.2 and 3.3 to zero. This however does not lead to a closed-form expressions, even in the case where one of the parameters is known. Since a closed-form solution for the maximum is unobtainable, the maximum has to be solved numerically where a two-dimensional (2-D) numerical maximisation (NM) has been reported in [39].
3.2. Method of Moments

Recently, Expectation Maximisation (EM) algorithm has been used to derive ML estimates for the K-distribution. The EM algorithm consists of two primary steps: The expectation step where the unknown underlying variables are obtained using the current estimate of the parameters and conditioned upon the observations [54]. The maximisation step provides a new estimate of the parameters. The two steps are iterated where the likelihood of estimates produced by successive iterations increases until a stationary point is reached.

EM method demonstrates computational advantage compared with 2-D numerical maximisation. This is because the numerical maximisation method attempts to find a maximal point, whereas the EM algorithm will find a stationary point only, which require less computation. However, the number of iterations required for the convergence of EM increases proportionally with the shape parameter $\nu$. Although ML based methods yield asymptotically efficient estimates, the amount of computation required makes this method impractical to be implemented in real-time systems. The ML method is only suitable for applications involving large amounts of data when the importance of highly accurate estimates is worth the extra computation required.

3.2 Method of Moments

An alternative approach that has been used extensively in the past is based on the method of moments (MOM). Estimates of the parameters may be accomplished by equating theoretical expressions and calculated values for two moments of the distribution and then solving for the two parameters. The moments of the K-distribution are given by

$$\mu_k = \langle X^k \rangle = \frac{\Gamma(0.5k + 1)\Gamma(\nu + 0.5k)}{\Gamma(\nu)} (2\alpha)^k$$  (3.4)
and the sample moments

\[ \hat{\mu}_k = \frac{1}{N} \sum_{i=1}^{N} x_i^k, \quad k \geq 0 \]  

(3.5)

where \( \{x_i; i = 1, \ldots, N\} \) is a set of realizations of \( N \) statistically independent RVs.

### 3.2.1 Higher Order Moments

The simplest choice for the two moments to be used in 3.4 is the mean and variance [8], given by

\[ \langle X \rangle = \frac{\Gamma(1.5)\Gamma(\nu + 0.5)}{\Gamma(\nu)} 2^a \]  

(3.6)

\[ \langle X^2 \rangle = 4a^2\nu - \langle X \rangle^2. \]  

(3.7)

This approach is numerically inefficient in that the derivation of \( \nu \) or \( a \) from 3.6 and 3.7 requires solving a tedious nonlinear equation, which needs to be performed numerically.

A more general expression was given by Raghavan in [69] where the ratio

\[ \alpha_m = \frac{\mu_{2m}}{\mu_{m}^2}, \quad m = 1, 2, \ldots \]  

(3.8)

is independent of parameter \( a \) and may be used for estimating \( \nu \). Since the parameters of K-distribution in radar systems are usually defined in terms of the second and fourth order moments, the approach that has been used extensively in the past is based on these two moments. Substituting 3.4 into 3.8 for \( m = 2 \), the parameter \( \nu \) can be obtained from

\[ \nu = \frac{2}{\frac{\mu_4}{\mu_2^2} - 2}, \quad \text{provided } \frac{\mu_4}{\mu_2^2} \neq 2 \]  

(3.9)
3.2. Method of Moments

Once the parameter $\nu$ is calculated from 3.9, the parameter $a$ can be obtained by simply using one of the moments, for example, the first order moment of $X$, which yields

$$a = \frac{\mu_1 \Gamma(\nu)}{\sqrt{\pi} \Gamma(\nu + 0.5)} \quad (3.10)$$

Estimates of $\nu$ and $a$ are then obtained from 3.9 and 3.10 after replacing the unknown moments by their estimates. This method works well when the number of samples is large, for example, more than a thousand. Since this method utilises higher order moments, such a good performance cannot be demonstrated if the number of samples available is small.

3.2.2 Fractional Lower Order Moments

Taking into consideration the fact that estimates based on higher order moments show large variability, Iskander et al [33] have proposed an estimation method based on lower order moments. As an alternative to 3.8, they proposed a ratio which is also independent of $a$ and is given by

$$\alpha_{p,q} = \frac{\mu_{p+2q}}{\mu_p \mu_{2q}} \quad p > 0, q = 1, 2, ... \quad (3.11)$$

Setting $q = 1$, the parameter $\nu$ is then given by

$$\nu = \frac{p(p+2)}{4[\beta_p - (0.5p + 1)]}, \quad p > 0 \quad (3.12)$$

where,

$$\beta_p = \alpha_{p,1} = \frac{\mu_{p+2}}{\mu_p \mu_2} \quad (3.13)$$
For \( p = 2 \), the ratio \( \beta_2 \) reduces to the one used in 3.8. To reduce the variance of the moment estimates, \( p \) is set in the range \( 0 < p \leq 2 \) where such interval will enable one to use fractional lower order moments (FLOM). All moments of order other than positive integer are called FLOM's. They were introduce to deal with distributions where higher order moments do not exist such as \( \alpha \)-stable distribution [97]. It has been shown in [33] that the best estimator using this method is achieved by setting \( p = 1/10 \), where \( \nu \) can be obtained from

\[
\nu = \frac{21/400}{\mu_{21/10}} - \frac{21}{\mu_{1/10} \mu_2} \quad (3.14)
\]

provided \( \frac{\mu_{21/10}}{\mu_{1/10} \mu_2} \neq \frac{21}{20} \). The use of fractional moments led to a lower variance of the parameter estimates when compared with the standard method based on the second and fourth order moments.

### 3.3 Texture Measure Based Method

Oliver [44] has discussed three estimators for the shape parameter which is based on three texture measures namely, normalised log \( (U) \), contrast \( (V) \) and variance of log \( (W) \). Their respective expectation values are related to the order parameter through

\[
U = \langle \ln x^2 \rangle - \ln \langle x^2 \rangle = \psi(\nu) - \ln \nu - \gamma \quad (3.15)
\]

\[
V = \frac{\langle x^4 \rangle}{\langle x^2 \rangle^2} - 1 = 1 + \frac{2}{\nu} \quad (3.16)
\]

\[
W = \langle \ln(x^2)^2 \rangle - \ln(x^2)^2 = \vartheta(\nu) + \frac{\pi^2}{6} \quad (3.17)
\]

where \( \vartheta(\nu) \) is the Trigamma function and \( \gamma \approx 0.5772 \) is Euler's constant. Over a finite number of samples \( N \), estimates of the expectation value of a random
variable $x$ (represented by $\hat{x}$ instead of $\langle x \rangle$) are obtained using the operator \ref{equation:3.3} with $k = 1$. The estimated values of the three texture measures $\hat{U}, \hat{V}, \hat{W}$ are then expressed in terms of these quantities.

Drawing on the previous observation that an improved accuracy results when the log of the data is used, Blacknell et al. \cite{10} has proposed a new method based on the moments of the form $[x^r \log(x)]$. From the compound form of the K-distribution \cite{94}

$$\langle x^r \rangle = \frac{\Gamma(r + 1)\Gamma(r + \nu)}{\Gamma(\nu)}$$ (3.18)

where $b = \nu/\mu$ and $\mu$ is the mean. Differentiation with respect to $r$, and setting $r = 1$, the statistics

$$X = \frac{\langle x \log x \rangle}{\langle x \rangle} - \langle \log x \rangle = 1 + \frac{1}{\nu}$$ (3.19)

can be used to estimate the shape parameter. The variances of the four statistics were derived using the higher-order asymptotic expansions, where the third-order expansions are given by

$$\sigma_U^2 = \frac{1}{N} \left\{ \psi(1)(\nu) + \frac{\pi^2}{6} - 1 \right\} - \frac{1}{2N^2} \left\{ (2r + 1)(4r + 1) \right\}$$ (3.20)

$$+ \frac{1}{6N^3} \left\{ 176r^3 + 194r^2 + 30r - 1 \right\}$$

$$\sigma_V^2 = \frac{1}{N^2} \left\{ 32(r + 1)(183r^3 + 123r^2 + 22r + 1) \right\}$$ (3.21)

$$+ \frac{1}{11N^3} \left\{ 4(r + 1)(103968r^4 + 88344r^3 + 22556r^2 + 1919r + 39) \right\}$$

$$\sigma_W^2 = \frac{1}{N} \left\{ \psi(3)(\nu) + 2\psi(1)(\nu)^2 + \frac{11\pi^4}{90} + \frac{2\pi^2}{3} \psi(1)(\nu) \right\}$$ (3.22)

$$- \frac{1}{N^2} \left\{ 2\psi(3)(\nu) + 2\psi(1)(\nu)^2 + \frac{17\pi^4}{90} + \frac{2\pi^2}{3} \psi(1)(\nu) \right\}$$

$$+ \frac{1}{N^3} \left\{ \psi(3)(\nu) + \frac{\pi^4}{15} \right\}$$
\[
\sigma_X^2 = \frac{1}{N} \left\{ -2\tau^3 + \frac{(2\tau + 1)\pi^2}{6} + (2\tau + 1)\psi^{(1)}(\nu) \right\} + \frac{1}{3N^2} \left\{ 36\tau^4 + 42\tau^3 - 3(27 + 2\pi^2)\tau^2 \right. \\
+ 6(7\pi^2)\tau + (3 - \pi^2) - 6(6\tau^2 + 7\tau + 1)\psi^{(1)}(\nu) \left. \right\} \left\{ -504\tau^5 - 684\tau^4 + 6(379 + 14\pi^2)\tau^3 \right. \\
+ 6(199 + 19\pi^2)\tau^2 \\
+ (24 + 31\pi^2)\tau - (6 - \pi^2) + 6(84\tau^3 + 114\tau^2 + 31\tau + 1)\psi^{(1)}(\nu) \right\}
\]

where \(\psi^{(\lambda)}\) is the polygamma function of order \(\lambda\) and \(\tau = 1/\nu\). The variance of the estimated \(\nu\) from any estimator \(S\) is then given by

\[
\sigma^2 \approx \frac{\sigma^2_S}{|S'(\nu)|^2}.
\]

Since this relationship is most accurate when \(S(\cdot)\) is a linear function of the parameter, \(\tau\) is preferred in the calculation [10]. Although the identities \(V\) and \(X\) given by equation 3.17 and 3.19, respectively, can be easily inverted to obtain an estimate of the shape parameter, the quality (in terms of variance) of both estimators is poor compared with the normalised log estimator. A comparative study by Blacknell [8] also concluded that estimation based on U-statistics provided the best performance. However, problems in estimating \(\nu\) by using statistics \(U\) arises due to the nonlinearity of the inversion.

### 3.4 Density Approximation Method

The difficulty of the ML solution for the K-distribution may be overcome by approximating the density 2.7 with a more convenient expression for which the ML estimator is more easily determined and implemented.
3.4. Density Approximation Method

3.4.1 Gamma Density Approximation

One of the candidates for the approximation is the Gamma density, the PDF of which is given by [69]

\[ p(x) = \frac{x^{\beta-1}}{b^\beta \Gamma(\beta)} \exp\left(-\frac{x}{b}\right). \]  

(3.25)

If we let the quantity \( \beta \) represent the shape parameter of a Gamma density, it is seen that a K-distribution with parameter \( \nu = 0.5 \) is identical to a Gamma distribution with shape parameter \( \beta = 1 \). In addition, for values of \( \nu \) equal to a positive integer plus half (\( \nu = 1.5, 2.5, 3.5, \ldots \)) the PDF in 2.7 is exactly expressed as a mixture of Gamma densities with different shape parameters.

By equating the first and second moments of the RVs, the parameters \( \nu \) and \( a \) may be expressed as functions of the gamma density parameters \( \beta \) and \( b \). Specifically, the parameter \( \beta \) of the equivalent gamma density may be expressed in terms of the parameter \( \nu \) by

\[ \beta = \left[ \frac{4(\nu)\Gamma^2(\nu)}{\pi\Gamma^2(\nu + 0.5)} - 1 \right]^{-1}. \]  

(3.26)

A comparison of the two PDFs for several values of \( \nu \) are made in Figure 3.1. It can be seen that as the parameter \( \nu \) increases, Gamma density provides a poor approximation. A comparison of higher order moments of the two densities in [69] shows that the above approach yields good approximations for values of \( \nu \) in the range \( 0.2 < \nu \leq 2 \). For larger values of \( \nu \) a weighted sum of two or more Gamma density functions is a better approximation than a single Gamma density. This generalisation is unnecessary because estimates of \( \nu \) with greater accuracy are more important for smaller values of \( \nu \) (\( \nu < 0.2 \)) which represent a very spiky clutter.
Given $N$ independent realizations from a Gamma density (parameters $\beta$ and $\gamma$), the ML estimate of the parameters ($\hat{\beta}_{ml}$ and $\hat{\gamma}_{ml}$) can be obtained as

$$\rho_N = \hat{\beta}_{ml} \exp \left[ -\psi(\hat{\beta}_{ml}) \right] = \frac{\mu_a}{\mu_g}, \quad \text{and} \quad (3.27)$$

$$\hat{\gamma}_{ml} = \frac{\mu_a}{\hat{\beta}_{ml}}, \quad (3.28)$$

where the geometric mean $\mu_g$ is defined as

$$\mu_g = \left( \prod_{i=1}^{N} x_i \right)^{\frac{1}{N}} \quad (3.29)$$

and the arithmetic mean, $\mu_a$ can be obtained by setting $k = 1$ in equation
3.4. Density Approximation Method

3.5. Parameter \( \nu \) can be estimated from relation 3.26 and 3.27 by using two-dimensional numerical search. From the expressions for the first order moment of \( X \) (equation 3.4) the parameter \( a \) may be obtained from the estimate \( \hat{\nu} \) and \( \mu_a \) as

\[
\hat{a} = \frac{\mu_a}{2} \frac{\Gamma(\hat{\nu})}{\Gamma(\hat{\nu} + 0.5)\Gamma(1.5)}
\]

(3.30)

Since \( \mu_a, \mu_\theta \) and \( \rho_N \) are evaluated from a finite number of samples, these quantities are RVs and their statistics depends upon the PDF of the samples \( x_i \), and the number of the samples used in the computation [69]. The parameters \( \nu \) and \( a \) may be determined from the knowledge of the first two moments of \( \rho_n \) and \( m_a \), and variance of this estimator is given by

\[
\langle \rho_N^2 \rangle = \frac{1}{N} \left( \nu - \frac{1}{N} \right) \left( 1 - \frac{1}{N} \right) \left[ \frac{\Gamma(\nu - N^{-1})\Gamma(1 - N^{-1})}{\Gamma(\nu)} \right]^N
\]

\[
+ \left( 1 - \frac{1}{N} \right) \left[ \frac{\Gamma(\nu + 0.5 - N^{-1})\Gamma(1.5 - N^{-1})}{\Gamma(\nu)} \right]^2
\]

\[
\times \left[ \frac{\Gamma(\nu - N^{-1})\Gamma(1 - N^{-1})}{\Gamma(\nu)} \right]^{N-2}, \quad N > \nu^{-1}
\]

(3.31)

It is verified in [69] that as \( N \rightarrow \infty \) the variance of \( \rho \) approaches zero.

3.4.2 Generalised Bessel K function Density

Another candidate for the approximation is the Generalised Bessel K function (GBK) distribution where PDF of a GBK-distributed RV is given by

\[
f(x) = \frac{2c}{\beta \Gamma(\alpha_1) \Gamma(\alpha_2)} \left( \frac{x}{\beta} \right)^{\frac{\alpha_1 + \alpha_2}{2} - 1} K_{(\alpha_2 - \alpha_1)} \left[ 2 \left( \frac{x}{\beta} \right)^{\frac{\alpha_1}{2}} \right]
\]

(3.32)

where \( \alpha_1 > 0, \alpha_2 > 0, \beta > 0, \) and \( c > 0 \) are the distribution parameters. The
GBK-distribution includes the K-distribution as a special case for the set of the parameters \((\alpha_1, \alpha_2, \beta, c) = (1, \nu, 2a, 2)\). Equating then the partial derivatives of the log-likelihood function for the GBK distribution to zero we have,

\[
c = \frac{\psi(\alpha_1) + \psi(\alpha_2)}{\frac{1}{N} \sum_{i=1}^{N} \ln \left( \frac{x_i}{\beta} \right)} \quad (3.33)
\]

\[
\frac{1}{N} \sum_{i=1}^{N} \mathcal{K}(x_i) = \alpha_1 + \alpha_2 \quad (3.34)
\]

\[
\frac{1}{N} \sum_{i=1}^{N} \mathcal{K}(x_i) \ln \left( \frac{x_i}{\beta} \right) = \frac{2}{c} + (\alpha_1 + \alpha_2) \frac{1}{N} \sum_{i=1}^{N} \ln \left( \frac{x_i}{\beta} \right) \quad (3.35)
\]

The maximum likelihood estimate of the K-distribution can be found by setting \(\alpha_1 = 1, \alpha_2 = \nu, \beta = 2a\) and \(c = 2\) and solving for the unknowns

\[
\hat{\alpha} = \frac{1}{2} \exp \left( \frac{\gamma - \psi(\nu)}{2} + \frac{1}{N} \sum_{i=1}^{N} \ln(x_i) \right) \quad (3.36)
\]

and

\[
\sum_{i=1}^{N} \mathcal{K}(x_i) \ln \left( \frac{x_i}{2\hat{\alpha}} \right) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{K}(x_i) \sum_{i=1}^{N} \ln \left( \frac{x_i}{2\hat{\alpha}} \right) + N \quad (3.37)
\]

where,

\[
\mathcal{K}(x_i) = \frac{K_{\nu-2} \left( \frac{x_i}{a} \right) + K_{\nu} \left( \frac{x_i}{a} \right)}{K_{\nu-1} \left( \frac{x_i}{a} \right)} \quad (3.38)
\]

Parameter \(\nu\) can be estimated by substituting \(\hat{\alpha}\) from 3.36 into 3.37, where such procedure needs to be performed numerically. The result given in 3.36 is useful in some radar applications where the values of \(\nu\) can be expressed as functions of the grazing angles, cross-range resolution, polarisation and aspect angle [99]. The \(k\)-th order moments given by 3.4 can be used to provide an explicit formula for an estimate of the shape parameter \(\nu\). Replacing these moments by its sample counterpart, \(\nu\) can be estimated as

\[
\hat{\nu} = g_k^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} \ln(x_i^k) - \ln \left[ \frac{1}{N} \sum_{i=1}^{N} x_i^k \right] + \frac{k\gamma}{2} \ln \left[ \Gamma \left( 1 + \frac{k}{2} \right) \right] \right) \quad (3.39)
\]
where $g_k^{-1}(\cdot)$ is the inverse function of $g_k(\cdot)$,

$$g_k(\tilde{\nu}) = \ln \left[ \frac{\Gamma(\tilde{\nu})}{\Gamma\left(\tilde{\nu} + \frac{k}{2}\right)} \right] + \frac{k \psi(\tilde{\nu})}{2}$$

(3.40)

and function $g_k(\cdot)$ is strictly monotonically increasing. A proof for monotonicity is given in [99]. It is interesting to note that fractional moments can also be used in equation 3.39. Although this method outperforms the higher order and fractional moment based techniques, the estimator is inferior with respect to Raghavan's method [69] for $\nu > 0.3$.

### 3.5 Characteristic Function Method

The methods described so far have been derived from the bivariate version of the K-distribution, namely a circularly symmetric random vector with the envelope having the PDF in 2.7. A K-distributed RV $x$ can be transformed into its two quadrature components by

$$x = x \cos(2\pi U)$$

(3.41)

$$x_\perp = x \sin(2\pi U)$$

(3.42)

where $U$ is uniformly distributed RV on the interval $(0,1)$. A comparison of the theoretical and an empirical CF of a univariate K-distribution is depicted in Figure 3.2. 1000 K-distributed RV were generated for $a = 1$ and $\nu = 1$. It is worth pointing out that the PDF given in 2.11 is symmetric. Thus the CF of a univariate K-distribution is real and even.
Figure 3.2: Comparison of the theoretical and empirical characteristic function of K-distribution for $a = 1$ and $\nu = 1$ and its studentised version.

Because the univariate K-distributed RV has a finite second moment, Ilow et al [31] have proposed a method of estimating the parameter $\nu$ by normalising the data by the estimate of the standard deviation. By doing this, the ECF-based estimation methods remain invariant under scale transformations. This process is known in the statistics literature as 'ECF studentisation' [56]. The theoretical and sample plot of the 'studentised' version is depicted in Figure 3.2. For the studentised data, the CF at point $t$ is given by

$$\phi(t) = \left(\frac{1}{1 + \frac{t^2}{2\nu}}\right)^\nu$$

(3.43)

thus $\nu$ can be obtained as a solution of a nonlinear equation through an iterative procedure. It has been verified in [31], theoretically and by simulation, that the value of $t$ close to 0.5 results in a good estimator performance for $\nu \in (0, 1)$. 
3.6 Neural Networks Based Method

Recently, a new method to estimate the shape parameter of K-distribution by using Artificial Neural Networks (ANN) [92] has been proposed. The estimation was formulated as a function approximation problem, to which ANNs are particularly well suited. Multilayer perceptrons trained with the backpropagation learning algorithm were implemented for their fast output generation in contrast to kernel-based networks.

The third and fourth standardised central moments, skewness ($\gamma_1$) and kurtosis ($\gamma_2$), respectively were used to characterise the shape of a distribution. As functions of $\nu$, the transformations $\gamma_1^{-2}$ and $\gamma_2^{-2}$ were found to have a high dynamic range, and were used as the input to the ANN [92]. The parameter estimation problem was then formulated as

$$\hat{\nu} = g(\gamma_1^{-2}, \gamma_2^{-2}),$$

and the goal of the ANN was to learn the mapping $g(\cdot, \cdot)$. For specific data and problem under consideration, an ANN with two hidden layers having ten and five neurons (sigmoidal activation function) was found as the simplest network with good training and testing results [81]. All networks were trained for 200 epochs. The results produced in [92] suggest that the greatest benefit of the neural approach is in estimating parameters for large $\nu$.

3.7 Result and Discussion

An estimation method based on the maximum likelihood principle is not implemented here, due to its computational complexity. Therefore, a comparison was only made by using the best estimator of each non-ML method. In order to provide a fair evaluation, methods described in this chapter have been
simulated simultaneously, by using the Monte Carlo approach. Independent realisations of K-distributed variables $X$ with parameters $a$ and $\nu$ were generated by

\[ x = 2a\sqrt{G_1 G_\nu} \]  

(3.45)

where $G_\rho$ denotes gamma RV with parameter $\beta = \rho$. The gamma RVs were generated using the function `gamrnd(\rho,1,M,N)` available in MATLAB Statistical Toolbox. Throughout the simulation, $a$ is assumed to be known and equal to unity. K-distributed data was generated for 15 different values of $\nu$ ($\nu = 0.1 : 0.1 : 1.5$).

The number of data samples was chosen to be 128, 256 and 512. The estimation was performed over 10,000 independent trials where in each case the averages were obtained. Figure 3.3, 3.4 and 3.5 show the estimated bias of the estimators for $N = 128$, $N = 256$ and $N = 512$ respectively. The estimated variance of the estimators are depicted in Figure 3.6, 3.7 and 3.8.

A large variability of the results obtained using the method derived from higher order moments is evident in all six figures. As for the case of Ilow's CF method, 'ECF studentisation' has proven to add more uncertainty to the data. Transformation 3.41 also contributes to the randomness making this method impractical as it is also computationally expensive.

The estimation method based on statistics $U$ produced the best results in terms of bias and variance compared with other texture measure based methods. Although the newly develop method based on statistics $X$ is simpler than the normalised log estimator, its bias and variance are larger. It can be seen that the method based on contrast ($V$-statistics) shows a large variability as it also utilises higher order moments.

The fractional lower order moment method can be regarded as the best low cost method currently available. However, it can be seen that the variance of
3.7. Result and Discussion

Figure 3.3: Estimated bias of the estimates of $\nu$ for $N = 128$.

Figure 3.4: Estimated bias of the estimates of $\nu$ for $N = 256$.

Figure 3.5: Estimated bias of the estimates of $\nu$ for $N = 512$.
Chapter 3. K-Distribution Parameter Estimation; A Review

Figure 3.6: Variance of the estimators for $N = 128$.

Figure 3.7: Variance of the estimators for $N = 256$.

Figure 3.8: Variance of the estimators for $N = 512$. 
this estimator is not as good as that of the methods employing a density approximation technique. The performance of the GBK approximation method is similar to the performance of Raghavan's estimator. For value of $\nu < 0.2$ the GBK method outperforms Raghavan's method as gamma density does not provide a good approximation in this region. On the other hand, the performance degrades with respect to Raghavan's method as the shape parameter increases.

3.8 Conclusion

Existing methods for estimating the parameters of a K-distribution have been depicted in this chapter. The maximum likelihood estimates of the parameters of a K-distribution are difficult to obtain and require computationally expensive numerical methods. Over the past decades, alternative means of estimating the parameters have been proposed and they are generally much faster.

However, the simplest of the methods namely the higher order moment based method produces the worst results, showing largest bias and variance. The fractional lower order moment method was found to be dramatically better.

In terms of accuracy and variance, the method proposed by Raghavan and the ML/MOM method are currently the best. The reason for this is that both methods are the solution of the maximum likelihood estimation formulation. Since both method are based on density approximation, the performance degrades as the dissimilarity of the approximated density and K-distribution increases.

It can be seen that the variance in $\nu$ increases with $\nu$. This is because the
K-distribution is less sensitive to the parameter $\nu$ at larger values of $\nu$. Thus it does not necessarily mean that the estimates are less accurate but the reduced sensitivity of the distribution to errors in $\tilde{\nu}$ compensates for the increased variability of $\tilde{\nu}$. For better performance, one has to choose more complex methods such as those that involve a two-dimensional numerical search.
Chapter 4

A Novel Approach to
K-Distribution Parameter
Estimation

The K-distribution is a two parameter model which uses the scale parameter, $a$, together with the shape parameter, $\nu$ to provide a complete description of the single point statistics of a K-distributed process. The shape parameter is a measure of the degree of the spikiness of the data. Small values of $\nu$ are associated with a spiky clutter, which represent sea echo for high resolution radar.

The use of such statistical models to describe the clutter data necessitates the task of estimating the parameter values from data. The accuracy of the parameter estimate has direct consequences for the performance of radar detection and estimation techniques. Armstrong and Griffiths [2] showed that incorrect estimates of $\nu$ give rise to considerable performance degradation in the constant false alarm rate (CFAR) target detection. Another radar detection technique which greatly relies on the accuracy of the $\nu$ estimates is the
classification of clutter in SAR images which makes use of the \( \nu \) value of the clutter region.

ML estimation offers the optimum parameter estimate when the form of the distribution to be estimated is known. However, the K-distribution lacks a closed form analytical solution for the ML parameter estimate. To maximise the performance of target detection and estimation, it is desirable to have a single parameter distribution, but still describing the sea clutter efficiently. The nuisance parameter should be eliminated and the scale parameter is the likely candidate.

In this chapter, we are going to see a new approach to estimate the parameter of K-distribution. In the first part, we are going to consider a case where the scale parameter is known a priori. The good performance shown with this method motivate us to investigate further two aspects; elimination of the scale parameter, and derivation of an estimation technique based on empirical characteristic function. These two aspects are shown in the subsequent section.

4.1 Motivation: A Case of Known Scale Parameter

Let us consider the case when the scale parameter can be assumed to be known priori. A K-distributed random variable is transformed into its two quadrature components by transformation 3.41 and 3.42. Since the PDF of the quadrature components is an even function, the imaginary component of its characteristic function is zero. Therefore, its sample counterpart, \( \hat{\phi}(t) \) is given by
4.1. Motivation: A Case of Known Scale Parameter

\[ \hat{\phi}_x(t) = \frac{1}{N} \sum_{i=1}^{N} \cos(t \alpha_i). \]  

(4.1)

where \( \alpha_i \) is the quadrature component of K-distributed RV and \( \hat{\phi}_x(t) \) is computable for all values of \( t \in \mathbb{R} \). Taking the logarithm of equation 2.9 and changing sign, we obtain

\[ -\ln (\phi_x(t)) = \nu \ln \left(1 + a^2 t^2\right) \]

(4.2)

where the logarithm of CF is now a linear function \( Y = mX + c \) with

\[
\begin{align*}
Y &= -\ln (\phi_x(t)), \\
m &= \nu, \\
X &= \ln \left(1 + a^2 t^2\right), \text{ and} \\
c &= 0.
\end{align*}
\]

(4.3)

Figure 4.1 shows the plot of the characteristics function with transformed axis as defined in equation 4.3. As we can see, the graph is now a straight line with gradient \( \nu \). This method can be used to estimate the shape parameter \( \nu \) for a case where scale parameter \( a \) is known.

The gradient of the slope can be estimated by using a least square estimator of the regression parameters. Suppose that the responses \( Y_i \) corresponding to input values \( X_i, i = 1, ..., N \) are to be observed and used to estimate the gradient \( m \), the shape parameter \( \nu \) can be estimated by

\[
\hat{\nu} = \frac{N \sum_{i=1}^{N} X_i Y_i - \bar{X} \sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i^2 - N \bar{X}^2}
\]

(4.4)
Chapter 4. A Novel Approach to K-Distribution Parameter Estimation

4.1.1 Simulation Results and Discussion

For computational simplicity and to avoid fluctuation at the tail of the characteristic function (in case of a small number of samples), it is proposed that the value $t$ should be in the range $0 \leq t \leq 2$. Throughout the simulation, $a$ is considered to be known ($a = 1$). K-distributed data were generated for seven different values of the shape parameter $\nu : 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5$.

ECF of the data was calculated for $t = [0 : 0.1 : 1]$. The number of data samples was chosen to be $N = 100$ and $N = 200$ and the estimation was performed over 1000 independent trials where in each case the averages were obtained. $Y$ is obtained by taking the logarithm of the estimated ECF $\hat{\phi}(t)$. For the same value of $t$, $X$ is calculated where $X$ according to equation 4.3.

Figure 4.2 and 4.3 show the estimated bias of the estimators. As we can

![Figure 4.1: The Log of Characteristic Function of K-distribution (quadrature components with $a = 1$ and $\nu = 1$). (-) Theory, (• •) Empirical.](image)
Figure 4.2: Estimated Bias of the Estimates of $\nu$ for $N = 100$.

Figure 4.3: Estimated Bias of the Estimates of $\nu$ for $N = 200$. 
Chapter 4. A Novel Approach to K-Distribution Parameter Estimation

Figure 4.4: Standard Deviation, $\sigma$ of the Estimators for $N = 100$.

Figure 4.5: Standard Deviation, $\sigma$ of the Estimators of $\nu$ for $N = 200$. 
see, the bias of the proposed method is small compared with all other Non-ML methods, where the maximum difference between the mean of the estimated \( \nu \) and the theoretical value for the proposed method was recorded as 0.01. From the two plots, we can conclude that the proposed method based on ECF is unbiased even for small number of samples.

The plotted values of the standard deviation of the estimates for \( N = 100 \) and \( N = 200 \) are shown in Figure 4.4 and 4.5 respectively. For \( \nu < 0.5 \), the performance of the ML/MOM method and the fractional moments method is slightly better than that of the proposed method by a margin of less than 0.015. However, as \( \nu \) increases the ECF method outperforms all the other methods.

It can be seen that the proposed method can offer an estimate of accuracy less than 0.2 as measured by the standard deviation, for \( \nu \) in the range \((0, 1.5)\). The accuracy of the estimator increases as the sample size increased. For \( \nu = 1.5 \) and \( N = 200 \), \( \nu \) could be estimated by ML/MOM method as 1.9, while the proposed method standard deviation is only 0.14.

### 4.1.2 Conclusion

It has been shown that the shape parameter of K-distribution can be estimated via the empirical characteristic function. The simulation results show that the method is unbiased even for a small size sample. Although, the variance of the proposed method is larger for small \( \nu \), the difference is insignificant. As \( \nu \) increases, the performance of the proposed method is a lot better compared to the other methods. In addition, the proposed method is computationally efficient, as it does not require solving non-linear equations or complicated functions. The only drawback is that the method is only applicable if the scale parameter \( \alpha \) is known a priori.
4.2 Log of K-Distribution

To improve target detection ability over sea clutter, anti-clutter techniques have been utilised, where a dual beam antenna is used. A significant improvement has been obtained with regards to the signal to clutter ratio, although due to a large dynamic range, the receiver output can be saturated, rendering target detection impossible. To overcome such difficulties, a LOG/CFAR technique is employed, which reduces the clutter output to about the receiver noise level by means of a logarithmic amplifier and a CFAR circuit. The block diagram of a LOG/CFAR receiver is depicted in Figure 4.6.

Let \{X_i; i = 1, 2, ..., N\} be \(N\) independent and identically distributed RVs with PDF 2.7. A K-distributed clutter \(X\) can be represented by

\[ X = a \cdot X_0 \]  \hspace{1cm} (4.5)

where \(a\) is the scale parameter and \(X_0\) is the K-distributed RV with \(a = 1\). A
sample plot of K-distributed RVs with the same shape parameter and different values of the scale parameter is shown in Figure 4.7. The histograms were obtained from 5000 sample of K-distributed RVs.

Assuming the input of the logarithmic amplifier is distributed according to 2.7, a sample obtained at the output, $Y$, is given by

$$Y = \ln X = \ln \alpha + \ln X_0.$$  \hspace{1cm} (4.6)

Let $F_X(x)$ and $F_Y(y)$ denote the CDF's for $X$ and $Y$, respectively. Then

$$F_Y(y) = P(Y \leq y) = P(\ln X \leq y) = P(X \leq e^y)$$

$$= \int_{-\infty}^{e^y} P_X(x) \, dx = F_X(e^y).$$  \hspace{1cm} (4.7)
Differentiating 4.7 with respect to $y$, we obtain the PDF of $Y$ in terms of the PDF of $X$ in the form

$$p_Y(y) = e^y p_X(e^y)$$  \hspace{1cm} (4.8)

thus, the PDF of $Y$ is given by

$$p_Y(y) = \frac{2e^y}{a\Gamma(\nu)} \left( \frac{e^y}{2a} \right)^\nu K_{\nu-1} \left( \frac{e^y}{a} \right).$$  \hspace{1cm} (4.9)

A comparison of the theoretical PDF in 4.9 and histogram of RV transformation 4.6 is made in Figure 4.8. As expected from equation 4.6, both densities have a similar shape but different location. The characteristic function of $Y$ can be obtained from

$$\phi_Y(t) = \langle e^{ist} \rangle = \langle x^i t \rangle = \int_0^\infty x^i t \cdot p(x)dx$$  \hspace{1cm} (4.10)

where $p(x)$ is distributed according to 2.7. Thus

$$\phi_Y(t) = \frac{(2a)^i t}{\Gamma(\nu)} \Gamma \left( \nu + \frac{jt}{2} \right) \Gamma \left( 1 + \frac{jt}{2} \right)$$  \hspace{1cm} (4.11)

The plot of the empirical characteristic function of the log K-distribution is depicted in Figure 4.9. It is interesting to note that the term $(2a)^i t$ forms a circle in a complex plane, with magnitude equal to 1 for all values of $a$ and $t$ ($|(2a)^i t| = 1$). Therefore, the absolute value of the CF is independent of the scale/shift parameter and can be use to estimate $\nu$. For value of $\nu = 1$, the magnitude of the characteristic function, $Y$, can be simplified to

$$|\phi_{Y,\nu=1}(t)| = \begin{cases} 1 & \text{if } t = 0 \\ \frac{\pi t}{2 \sinh \frac{\pi t}{2}} & \text{if } t \neq 0 \end{cases}$$  \hspace{1cm} (4.12)
4.2. Log of K-Distribution

Figure 4.8: Theoretical PDF and histograms of log K-distributed RV for a fixed $\nu$ and variable scale parameter. (H)-Histogram (T)-Theoretical

Figure 4.9: Empirical characteristic function of RV $Y$
From relation 2.15, the mean and variance of $Y$ is given by

$$\langle Y \rangle = \ln 2 \alpha + \frac{\psi(\nu)}{2} - \frac{\gamma}{2}$$  \hspace{0.5cm} (4.13)$$

$$\langle Y^2 \rangle = \frac{\psi'(\nu)}{4} + \frac{\pi^2}{24}.$$  \hspace{0.5cm} (4.14)

From the moments equation above, it is confirmed that the scale parameter of the K-distribution is transformed into a shift parameter which controls the location of the PDF of $Y$. The variance of the log K-distribution is scale/shift invariant. Thus it could be used to estimate the shape parameter. It can be verified that as $\nu \rightarrow \infty$, the log K-distribution tends to log Rayleigh, where

$$\lim_{\nu \rightarrow \infty} \langle Y^2 \rangle = \frac{\pi^2}{24}$$  \hspace{0.5cm} (4.15)

which is a variance of the log Rayleigh distribution.

### 4.3 Scale Invariant Random Variables

Let us now define a new RV $Z$ as

$$Z = Y_\alpha - Y_\beta$$  \hspace{0.5cm} (4.16)

where $Y_\alpha$ and $Y_\beta$ are two independent and identical random variable, distributed according to 4.9. Substituting equation 4.6 into 4.16, shows that the scale parameter of the K-distribution can be eliminated, thus RV $Z$ is scale invariant with the same shape parameter $\nu$ remaining unaffected. Since RV $Y_\alpha$ and $Y_\beta$ are independent, then the density of their sum $Z = Y_\alpha + (-Y_\beta)$ equals the convolution of their respective densities

$$f(z) = \int_{-\infty}^{\infty} f_{y_\beta}(y_\alpha - z) f_{y_\alpha}(y_\alpha) \, dy_\alpha.$$  \hspace{0.5cm} (4.17)
4.3. Scale Invariant Random Variables

Figure 4.10: Theoretical PDF and histogram of two K-distributed RVs after transformation 4.16.

Thus, the PDF of $Z$ is given by

$$
p(z) = 2e^{2y}y^2 \frac{\Gamma(2\nu)}{\Gamma(2\nu + 2)} \left( \frac{2}{1 + e^{2z}} \right)^{2\nu} \cdot F(\nu, \nu + 0.5, \nu + 1.5; Q) \quad (4.18)
$$

where $Q = \left( \frac{1 - e^{2z}}{1 + e^{2z}} \right)^2$ and $F(a, b, c; Q)$ is the Gauss Hypergeometric function. Figure 4.10 shows the plot of PDF 4.18 for $\nu = 1$ and a histogram of two K-distributed RV (same shape parameter but different in scale) after transformation 4.16. First and second moment of $Z$ is related to moments of $Y$ as

$$
\langle Z \rangle = \langle Y_\alpha - Y_\beta \rangle = 0 \quad \text{and} \quad \langle Z^2 \rangle = 2 \langle Y^2 \rangle = \frac{1}{2} \psi'(\nu) + \frac{\pi^2}{12}. \quad (4.19)
$$
From the moment generating function of $Z$ (see Appendix A), it is verified that the above equation holds, and the fourth order moment of $Z$ is given by

$$\mu_4 = \frac{1}{8}\psi^{(3)}(\nu) + \frac{3}{4}\psi'(\nu)^2 + \frac{\pi^2}{4}\psi'(\nu) + \frac{7\pi^2}{240}$$

(4.21)

where $\psi^{(\lambda)}$ is the polygamma function of order $\lambda$.

A plot of function 4.20 against $\nu$ is depicted in Figure 4.11. Since the function 4.20 is monotonic, it can be used to estimate the shape parameter. As $\nu \to \infty$, the variance of $Z$ reduces to the variance of the sum of two log-Rayleigh distributions.
4.4 Parameter Estimation

4.4.1 Using Second Order Moments

The shape parameter $\nu$ can be estimated by using the second order moment of the transformed RV. This is done by equating theoretical and sample variance of RV $Z$. Thus the estimate of $\nu$ is given by

$$\hat{\nu} = \vartheta^{-1} \left( 2\hat{\mu}_2 - \pi^2/6 \right) \quad (4.22)$$

where $\vartheta^{-1}(\cdot)$ denotes the inverse trigamma function, and $\hat{\mu}_2$ is the estimated variance. The theoretical variance of the above estimator can be found by considering a statistic $\hat{\mu}_2$ which is related to a parameter $\nu$ through a function $f$, where

$$\hat{\mu}_2 = f(\hat{\nu}) = \frac{1}{2} \vartheta(\hat{\nu}) + \frac{\pi^2}{12}. \quad (4.23)$$

An estimate for $\nu$, namely $\hat{\nu}$ can be defined such that the measured statistic is related to the parameter estimate as $\hat{\mu}_2 = f(\hat{\nu})$. In the asymptotic limit the measured statistic will take its expectation value and in this limit the inverse function of the statistic will be the true value of the parameter, thus the parameter estimate is asymptotically unbiased. From a Taylor series expansion

$$f(\hat{\nu}) - f(\langle \hat{\nu} \rangle) = (\hat{\nu} - \langle \hat{\nu} \rangle) \frac{df(\langle \hat{\nu} \rangle)}{d\hat{\nu}} + \frac{(\hat{\nu} - \langle \hat{\nu} \rangle)^2}{2} \frac{d^2f(\langle \hat{\nu} \rangle)}{d\hat{\nu}^2} + ... \quad (4.24)$$

where $\langle \cdot \rangle$ represents the expectation operator. Substituting for the statistic $\hat{\mu}_2$ in 4.24 for the asymptotic limit $\langle \hat{\nu} \rangle \rightarrow \nu$, we obtain (after taking the square of both sides and then taking the expectation values)
\[
\mu_2 - \mu_2 = \int_0^\infty (\hat{\nu} - \langle \hat{\nu} \rangle)^2 \left( \frac{df(\langle \hat{\nu} \rangle)}{d\hat{\nu}} \right)^2 P_\nu(\hat{\nu}) \, d\hat{\nu} + \int_0^\infty (\hat{\nu} - \langle \hat{\nu} \rangle)^3 \frac{d^3 f(\langle \hat{\nu} \rangle)}{d\hat{\nu}^3} P_\nu(\hat{\nu}) \, d\hat{\nu} + \ldots \quad (4.25)
\]

Assuming that \( \langle (\hat{\nu} - \langle \hat{\nu} \rangle)^N \rangle \ll \sigma_\nu^2 \) where \( N \geq 3 \), then the variance in \( \mu_2 \) is related to the variance in the estimated quantity \( \hat{\nu} \) as [36]

\[
\sigma_{\mu_2}^2 = \sigma_\nu^2 \left( \frac{df(\langle \hat{\nu} \rangle)}{d\hat{\nu}} \right)^2. \quad (4.26)
\]

Since the variance of the second moment is given by

\[
\sigma_{\mu_2}^2 = \frac{1}{N} (\mu_4 - \mu_2^2) \quad (4.27)
\]

where \( N \) is the number of sample, and \( \mu_4 \) and \( \mu_2 \) are given in 4.21 and 4.20 respectively, the theoretical variance of the estimator 4.22 is given by

\[
\sigma_\nu^2 = \frac{4(\mu_4 - \mu_2^2)}{N \psi^{(3)}(\nu)^2} \quad (4.28)
\]

### 4.4.2 Parameter Estimation by Empirical Characteristic Function

The characteristic function of the convolution of distribution function is the product of the corresponding CFs. Thus from the convolution theorem, the characteristic function of RV \( Z \) is given by

\[
\phi_Z(t) = |\phi_Y(t)|^2 = \frac{\Gamma(\nu + \frac{\alpha t}{2}) \Gamma(\nu - \frac{\alpha t}{2}) \Gamma(1 + \frac{\beta t}{2}) \Gamma(1 - \frac{\beta t}{2})}{\Gamma(\nu)^2} \quad (4.29)
\]
4.4. Parameter Estimation

In the product term, this is given by

\[ \phi_Z(t) = \prod_{n=0}^{\infty} \left[ 1 + \frac{t^2}{4(n + \nu)^2} \right]^{-1} \cdot \frac{\pi t}{2 \sinh \frac{\pi t}{2}} \]  

(4.30)

It is frequently convenient to use the logarithm of the characteristic function, where the function \( \Psi(t) = \ln \phi(t) \) is known in the statistical literature as the second characteristic function of the RV [62]. Thus, by taking logarithm of equation 4.30 and rearranging the equation, we have

\[ \Psi_Z(t) - \ln \left( \frac{\pi t}{2 \sinh \frac{\pi t}{2}} \right) = \sum_{n=0}^{\infty} -\ln \left( 1 + \frac{t^2}{4(n + \nu)^2} \right) \]  

(4.31)

In many methods employing the empirical characteristic function, it has been recommended to use small value of \( t \) in order to achieve good estimates. Thus, without loss of generality and assuming \( t \) is small, the right hand side of equation 4.31, becomes

\[ \sum_{n=0}^{\infty} -\ln \left( 1 + \frac{t^2}{4(n + \nu)^2} \right) \approx -\frac{t^2}{4} \sum_{n=0}^{\infty} \frac{1}{(\nu + n)^2} \]  

(4.32)

where the summation term is the Trigamma function [1]. By letting

\[ \kappa = \ln \frac{\pi t}{2 \sinh \frac{\pi t}{2}} \]  

(4.33)

the shape parameter can be estimated as,

\[ \hat{\nu} = \nu^{-1} \left[ -4t^{-2}(\Psi_Z(t) - \kappa) \right] \]  

(4.34)

where \( \Psi_Z(\cdot) \) is the logarithm of the empirical characteristic function.
4.4.3 Parameter Estimation using Kernel Characteristic Function Estimator

To improve the performance of the estimator derived from the CF, we propose the use of Kernel Characteristic Function Estimator (KCFE). Primarily, KCFE has been introduced by Zoubir and Arnold for testing Gaussianity in IID case. The general formulation for the KCFE of the characteristic function of an IID random vector $X = [X_1, X_2, ..., X_N]$ is given by [98]

$$
\hat{\phi}_X(t; \varphi_X) = \frac{1}{N} \sum_{i=0}^{N} \varphi_X(t) \exp(jX_it) \tag{4.35}
$$

where $\varphi_X(t)$ is known as the characteristic function domain kernel function, which in general may depend on the data and the number of samples $N$. Here, the convolutional smoothing operation of the kernel density estimation has been replaced by a multiplicative smoothing operation in the characteristic function domain. The most general form of KCFE which is amenable to statistical analysis is the fixed KCFE whose kernel function $\varphi(t)$ is a fixed function of $t$ [98]. For a zero-mean symmetrical process the optimum solutions for the kernel function are given by

$$
\Re\varphi_R(t) = \Re\varphi_\Theta(t) = \Im\varphi_\Theta(t) = 0, \tag{4.36}
$$

leaving us to select $\Re\varphi_R(t)$ [98]. One choice of kernel function for estimation of a K-distribution CF is Gaussian which it has been reported in [98] to have a wide range of applicability amongst non-Gaussian process. The kernel function of a Gaussian is given by

$$
\varphi^R(t) = \exp(-\sigma^2 \sigma^2 t^2 / 2) \tag{4.37}
$$
where $\sigma_R$ is the kernel function (inverse) width. The first order approximation to the solution of the optimum width for a Gaussian kernel function for a Gaussian process is given by $\sigma_R = \sigma(2/3N)^{1/5}$, and a more accurate numerical analysis with respect to the dependence of $\sigma_R$ on $N$ gives $\sigma_R = 0.97\sigma N^{-0.2}$ [98]. Since RV $Z$ from 4.16 is a zero mean symmetrical process, the results from 4.36 and 4.37 can be applied. Figure 4.12 shows multiple realisations of the magnitude of ECF and a fixed KCFE for RV $Z_{n=64} \sim \mathcal{K}(1,1)^1$. The comparison clearly displays the improved variance performance. The parameter $\nu$ can then be estimated by replacing the empirical characteristic function by its KCFE in equation 4.34.

1$\mathcal{K}(\alpha, \nu)$ denotes the K-distributed RV with parameter $\alpha$ and $\nu$
4.5 Methods for Scale Parameter Elimination

Let \( Y = [Y_1, Y_2, \ldots, Y_{N-1}, Y_N] \) be \( N \) log-K-distributed RVs acquired from the output of the logarithmic amplifier in Figure 4.6. The simplest way to obtain the RV \( Z \) is by splitting \( Y \) into two independent sets \( Y_\alpha \) and \( Y_\beta \), where from

\[
Y_\alpha = [Y_1, Y_2, \ldots, Y_{N/2}] \\
Y_\beta = [Y_{N/2+1}, Y_{N/2+2}, \ldots, Y_N]
\]

\( Z \) is given by equation 4.16. Acquiring RV \( Z \) in this manner however will reduce the number of samples by one half of its original amount. Figure 4.13 shows the theoretical variance of this estimator for \( N = 64 \) and 128. A Comparison is made with the simulation results from Raghavan’s and FLOM method for \( N = 128 \). It can be seen that, as the number of sample reduces to one half, the performance degrades significantly.

Another method that can be used is to subtract RV \( Y \) successively in sequence where

\[
Z_i = Y_{i+1} - Y_i \quad i = 1, 2, \ldots, N.
\]

Even though this procedure will produce \( N - 1 \) samples of \( Z \), a simulation showed that it can not match the performance given by equation 4.28. This is because acquiring \( Z \) in this way will produce highly correlated variables with different other statistical and structural characteristics.

In order to overcome the drawbacks arising from the two methods described earlier, Hadamard transform can be used to acquire the uncorrelated sample of RV \( Z \). A Hadamard transform is based upon the Hadamard matrix where an \( N \times N \) matrix \( H_N \) is a Hadamard matrix of order \( N \) if the entries of \( H_N \)
are either +1 or -1 and such that

\[ H_N H_N^T = N \mathbf{I} \]  \hspace{1cm} (4.41)

where T denotes the transpose and I is the identity matrix of order \( N \). The matrix also must satisfy the following:

- \( N \) must be 1, 2, or \( 4k \), where \( k \) is an integer.
- The scalar product of any two distinct rows (or columns) is zero.

These matrices were first considered as Hadamard determinants as they satisfy equality in Hadamard's determinant theorem. The theorem states that, if \( X = x_{ij} \) is a matrix of order \( N \) where \( |x_{ij}| \leq 1 \) for all \( i \) and \( j \), then

\[ |\det X| \leq N^{N/2} \]  \hspace{1cm} (4.42)
It is apparent that if the rows and columns of a Hadamard matrix are permuted, the matrix remains Hadamard. It is also true that if any row or column is multiplied by -1, the Hadamard property is retained. Thus it is possible to arrange to have the first row and first column of a Hadamard matrix to contain only +1 entries. A Hadamard matrix in this form is said to be normalised. For \( N=4 \), it is given by

\[
H_4 = \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1
\end{bmatrix}
\]  (4.43)

Let us now consider the logarithm of K-distributed RV, \( Y = [Y_1, Y_2, Y_3, Y_4] \), and let \( Z = H_4 Y^T \) be the 4-point Hadamard transform of \( Y \),

\[
\begin{bmatrix}
Z_1 \\
Z_2 \\
Z_3 \\
Z_4
\end{bmatrix} = \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1
\end{bmatrix} \begin{bmatrix}
Y_1 \\
Y_2 \\
Y_3 \\
Y_4
\end{bmatrix}
\]

\[
= \begin{bmatrix}
Y_1 + Y_2 + Y_3 + Y_4 \\
Y_1 - Y_2 + Y_3 - Y_4 \\
Y_1 + Y_2 - Y_3 - Y_4 \\
Y_1 - Y_2 - Y_3 + Y_4
\end{bmatrix}
\]  (4.45)

Since the first row of the transformation contains only summation, \( Z_1 \) is discarded. The resultant \( N-1 \) samples of scale invariant RV \( Z_i; i = 2, 3, ..., N \) must be normalised by multiplying each array by \( \sqrt{2/N} \).

4.6 Application of the Proposed Methods to Other Densities

The method outlined in the previous sections is also applicable for estimating the parameters of other densities. For example, Weibull density is also useful
4.6. Application of the Proposed Methods to Other Densities

for modelling radar clutter. The PDF of a Weibull distribution is given by

\[ p(x) = \frac{c}{b} \left( \frac{x}{b} \right)^{c-1} \exp \left( - \left( \frac{x}{b} \right)^{c} \right) \]  

(4.46)

where \( b \) is the scale parameter and \( c \) is the shape parameter. The moments of the Weibull distribution are given by

\[ \mathbb{E}[X^k] = b^k \Gamma \left( \frac{k}{c} + 1 \right) \].  

(4.47)

Let \( Z = \ln Y - \ln X \) where \( X \) and \( Y \) are two IID Weibull RV. The moment generating function of the scale invariant RV \( Z \) is given by

\[ m_Z(t) = \Gamma \left( \frac{c+t}{c} \right) \Gamma \left( \frac{c-t}{c} \right) \].  

(4.48)

After some algebra, \( c \) can be estimated using its second order moment by

\[ \hat{c} = \sqrt{\frac{\pi^2}{3\hat{\mu}_2}} \]  

(4.49)

or alternatively by using ECF or KCFE

\[ \hat{c} = \frac{\pi t}{\sinh(\pi t/\hat{c}) \cdot \phi_Z(t; \varphi_Z)} \]  

(4.50)

where \( \varphi_Z(t) = 1 \) if ECF is used. The derivation of 4.48, 4.49 and 4.50 is shown in Appendix A.
4.7 Simulation Result

The proposed estimation methods have been simulated simultaneously, by using the Monte Carlo approach. Throughout the simulation, \( \alpha \) is assumed to be known and equal to unity. K-distributed data was generated for 15 different values of \( \nu \) (\( \nu = 0.1 : 0.1 : 1.5 \)). The number of data samples was chosen to be 128 and 256. The estimation was performed over 10,000 independent trials where in each case the averages were obtained.

The results of the proposed estimation technique are compared with the Fractional Lower Order Moments and Raghavan’s methods only. The FLOM method is chosen because it is the second most efficient computationally, while Raghavan’s provides the best overall performance in term of variances. Although the method derived from the second and fourth order moments is the fastest, the variance of the estimator is too large to include in the comparison.

4.7.1 A Comparison on the Scale Parameter Elimination Method

In this section, the results from three methods of scale parameter elimination are compared. The estimator described by equation 4.22 which is based on moments is used. Although the inverse trigamma function is unavailable, the numerical solution is easy to implement. First, the function \( \mu_2 \) is calculated for a range of \( \nu \) through function 4.20. From the estimate \( \hat{\mu}_2 \) and function \( \mu_2 \), we can obtain the estimate of \( \nu \) using for example cubic spline interpolation, i.e.;

\[
\nu = \text{spline}(\mu_2, \nu, \hat{\mu}_2)
\]

available in Matlab. The sample mean and the sample variance of the estimates are depicted in Table 4.1 and 4.2 for \( N = 128 \) and \( N = 256 \), respectively.

The results are based on the scale elimination method as follows:
4.7. Simulation Result

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$\bar{\nu}_1$</th>
<th>$\text{Var}[\bar{\nu}_1]$</th>
<th>$\bar{\nu}_2$</th>
<th>$\text{Var}[\bar{\nu}_2]$</th>
<th>$\bar{\nu}_3$</th>
<th>$\text{Var}[\bar{\nu}_3]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1023</td>
<td>6.4843e-5</td>
<td>0.1018</td>
<td>4.9020e-5</td>
<td>0.1014</td>
<td>3.6732e-5</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2051</td>
<td>9.3838e-4</td>
<td>0.2048</td>
<td>8.1767e-4</td>
<td>0.2027</td>
<td>6.7007e-4</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3101</td>
<td>2.4065e-3</td>
<td>0.3078</td>
<td>1.9901e-3</td>
<td>0.3058</td>
<td>1.6385e-3</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4158</td>
<td>5.0154e-3</td>
<td>0.4129</td>
<td>4.0659e-3</td>
<td>0.4086</td>
<td>3.1975e-3</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5233</td>
<td>1.0013e-2</td>
<td>0.5186</td>
<td>7.9751e-3</td>
<td>0.5119</td>
<td>5.7716e-3</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6350</td>
<td>1.9558e-2</td>
<td>0.6306</td>
<td>1.4400e-2</td>
<td>0.6173</td>
<td>1.0078e-2</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7508</td>
<td>3.9758e-2</td>
<td>0.7387</td>
<td>2.6282e-2</td>
<td>0.7231</td>
<td>1.7179e-2</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8738</td>
<td>7.6253e-2</td>
<td>0.8593</td>
<td>4.5369e-2</td>
<td>0.8333</td>
<td>2.8793e-2</td>
</tr>
<tr>
<td>0.9</td>
<td>1.0044</td>
<td>1.5065e-1</td>
<td>0.9771</td>
<td>7.8985e-2</td>
<td>0.9445</td>
<td>4.8702e-2</td>
</tr>
<tr>
<td>1.0</td>
<td>1.1463</td>
<td>2.6226e-1</td>
<td>1.1046</td>
<td>1.2735e-1</td>
<td>1.0622</td>
<td>8.0712e-2</td>
</tr>
<tr>
<td>1.1</td>
<td>1.3026</td>
<td>4.5684e-1</td>
<td>1.2372</td>
<td>1.9930e-1</td>
<td>1.1854</td>
<td>1.3212e-1</td>
</tr>
<tr>
<td>1.2</td>
<td>1.4647</td>
<td>6.9241e-1</td>
<td>1.3614</td>
<td>2.8552e-1</td>
<td>1.3142</td>
<td>2.0653e-1</td>
</tr>
<tr>
<td>1.3</td>
<td>1.6318</td>
<td>9.8731e-1</td>
<td>1.4887</td>
<td>4.3685e-1</td>
<td>1.4477</td>
<td>3.1231e-1</td>
</tr>
<tr>
<td>1.4</td>
<td>1.7984</td>
<td>1.3234e+0</td>
<td>1.6110</td>
<td>5.4939e-1</td>
<td>1.5856</td>
<td>4.1473e-1</td>
</tr>
<tr>
<td>1.5</td>
<td>1.9714</td>
<td>1.7023e+0</td>
<td>1.7391</td>
<td>6.9151e-1</td>
<td>1.7243</td>
<td>5.2119e-1</td>
</tr>
</tbody>
</table>

Table 4.1: Sample mean and sample variance of the estimates of $\nu$ for $N = 128$ by using moments method.

- **Method 1** Estimates $\bar{\nu}_1$ by $Z = Y_\alpha - Y_\beta$.
- **Method 2** Estimates $\bar{\nu}_2$ by $Z = Y_{i+1} - Y_i$.
- **Method 3** Estimates $\bar{\nu}_3$ by $Z = H_N Y^T$.

As indicated in Table 4.1 and Table 4.2, the parameter estimation method based on the Hadamard transform recorded the best performance in terms of bias and variance. To make two independent sets, the scale elimination based on method 1 reduces the number of samples to one half. The result can be seen to reduce the performance of the estimator significantly. However, this method is by far better in terms of the quality and processing time, than the method based on second and fourth moments of K-distribution.
Chapter 4. A Novel Approach to K-Distribution Parameter Estimation

When the processing time is more important than the quality of the estimates, it is suggested to use method 2 instead. With roughly the same amount of computation, method 2 performed much better than method 1. The unattractive feature of this method is that it will produce a highly correlated RV, the correlation of which depends on the sequence of the RVs.

The Mean Square Error (MSE) of the estimates are shown in Figure 4.14 and 4.15. It can be seen that the performance of method 2 and method 3 was very close. However, the difference between method 1 and the other two becomes larger as the shape parameter increases. When the quality of the estimator is important, the scale elimination by method 3 is worth the extra computation.

### Table 4.2: Sample mean and sample variance of the estimates of $\nu$ for $N = 256$ by using moments method.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$E[\hat{\nu}_1]$</th>
<th>$\text{Var}[\hat{\nu}_1]$</th>
<th>$E[\hat{\nu}_2]$</th>
<th>$\text{Var}[\hat{\nu}_2]$</th>
<th>$E[\hat{\nu}_3]$</th>
<th>$\text{Var}[\hat{\nu}_3]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1009</td>
<td>1.8223e-05</td>
<td>0.1008</td>
<td>1.4821e-05</td>
<td>0.1006</td>
<td>1.1367e-05</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2024</td>
<td>4.3176e-04</td>
<td>0.2024</td>
<td>3.6604e-04</td>
<td>0.2013</td>
<td>3.1898e-04</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3048</td>
<td>1.1174e-03</td>
<td>0.3044</td>
<td>9.7412e-04</td>
<td>0.3027</td>
<td>7.9455e-04</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4076</td>
<td>2.3028e-03</td>
<td>0.4065</td>
<td>1.9091e-03</td>
<td>0.4042</td>
<td>1.5564e-03</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5112</td>
<td>4.3479e-03</td>
<td>0.5090</td>
<td>3.5577e-03</td>
<td>0.5058</td>
<td>2.7486e-03</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6162</td>
<td>7.7169e-03</td>
<td>0.6134</td>
<td>6.2035e-03</td>
<td>0.6080</td>
<td>4.6163e-03</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7235</td>
<td>1.3567e-02</td>
<td>0.7181</td>
<td>1.0193e-2</td>
<td>0.7118</td>
<td>7.6403e-03</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8324</td>
<td>2.3083e-02</td>
<td>0.8253</td>
<td>1.7122e-2</td>
<td>0.8155</td>
<td>1.2107e-02</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9445</td>
<td>3.8716e-02</td>
<td>0.9344</td>
<td>2.6758e-2</td>
<td>0.9209</td>
<td>1.9026e-02</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0612</td>
<td>6.4605e-02</td>
<td>1.0462</td>
<td>4.3714e-2</td>
<td>1.0286</td>
<td>2.9396e-02</td>
</tr>
<tr>
<td>1.1</td>
<td>1.1805</td>
<td>1.0446e-01</td>
<td>1.1596</td>
<td>6.6076e-2</td>
<td>1.1377</td>
<td>4.5080e-02</td>
</tr>
<tr>
<td>1.2</td>
<td>1.3084</td>
<td>1.6691e-01</td>
<td>1.2696</td>
<td>9.3484e-2</td>
<td>1.2504</td>
<td>6.8176e-02</td>
</tr>
<tr>
<td>1.3</td>
<td>1.4388</td>
<td>2.4889e-01</td>
<td>1.3925</td>
<td>1.2820e-1</td>
<td>1.3657</td>
<td>1.0209e-01</td>
</tr>
<tr>
<td>1.4</td>
<td>1.5740</td>
<td>3.5510e-01</td>
<td>1.5088</td>
<td>1.6947e-1</td>
<td>1.4836</td>
<td>1.4352e-01</td>
</tr>
<tr>
<td>1.5</td>
<td>1.7107</td>
<td>4.7706e-01</td>
<td>1.6244</td>
<td>2.1353e-1</td>
<td>1.6043</td>
<td>2.0267e-01</td>
</tr>
</tbody>
</table>
4.7. Simulation Result

Figure 4.14: The mean square error of the estimates of $v$ for $N = 128$.

Figure 4.15: The mean square error of the estimates of $v$ for $N = 256$. 
4.7.2 Comparison of the Characteristic Function Based Method

In this section, the performance comparison of the method based on CF is carried out. The inverse trigamma function again has to be solved numerically. In the simulation, \( \phi_Z(t) \) is first calculated for a range of \( \nu \) and selected \( t \). The estimates is then achieved by \( \hat{\nu} = \text{spline}(\phi_Z(t), \nu, \xi) \), where \( \xi \) could be ECF or KCFE. In all cases, the Hadamard transform was used in the scale elimination procedure.

It is interesting to note that the PDF of RV \( Z \) is even and symmetric thus its CF is real and even, therefore the empirical characteristic function can be simplified to equation 4.1. The KCFE can be easily obtained by multiplying the calculated ECF with a Gaussian kernel function, thus

\[
\hat{\phi}_Z(t; \varphi_{Z,i}) = \hat{\phi}(t) \cdot \exp(-\sigma^2 t^2 / 2) \tag{4.51}
\]

where the term \( \sigma^2 \) is calculated from the sample. To make the new estimator operational, we need to select the constant \( t \). Selecting \( t \) entails a tradeoff between the asymptotic variance, its robustness, as measured by the breakdown point, as well as for the assumption 4.32 to be valid. In order for the approximation to be accurate for \( \nu \in (0.1, \infty) \) to double precision, \( t \) has to be \( \leq 0.02 \). Values of \( t \) in this region provide us with small asymptotic variance and a higher breakdown point. From Monte Carlo simulations, the best estimator for the shape parameter \( \nu \) is achieved when \( t = 0.01 \).

The estimates of the two estimators is compared in Figure 4.16 and 4.17 for \( N = 128 \) and \( N = 256 \), respectively. As we can see, the estimated bias of the estimates from the method employing KCFE is smaller than those by ECF. A comparison of the estimated variance in Figure 4.18 and 4.19 also shows that
Figure 4.16: A comparison of the estimates of $\nu$ for $N = 128$.

Figure 4.17: A comparison of the estimates of $\nu$ for $N = 256$. 
Figure 4.18: The Estimated Variance of the estimates of $\nu$ for $N = 128$.

Figure 4.19: The Estimated Variance of the estimates of $\nu$ for $N = 256$. 
KCFE outperforms the ECF-based method. This shows that KCFE is a better estimator than the ECF, especially in a small sample case. The application of Gaussian kernel in the KCFE method also does not seem to deteriorate results.

4.8 Discussion

Here, comparison is made between the proposed method and the previously published method. Figure 4.20 and 4.21 shows the mean square error of the estimates of $\nu$ for $N = 128$ and $N = 256$, respectively. As expected, KCFE method performs better than the proposed moment-based method (denoted in the figure as $\mu_2$), especially for the small sample case. It can be seen that the MSE of both of the proposed methods outperforms all other existing methods for $\nu < 0.8$. This is very important in radar application as it represents a very spiky clutter.

However, as the shape parameter increases, FLOM method is slightly better than $\mu_2$, for $N = 256$ and $\nu > 1.3$ and Raghavan’s is slightly better than KCFE, for $N=256$ and $\nu > 1.2$. The MSE of Raghavan’s method also exceeds the FLOM method for $\nu < 0.2$ for $N = 128$ and $\nu < 0.1$ for $N = 256$. Although the variance of the estimates obtained by Raghavan’s method is smaller than FLOM, as depicted in Figure 3.6 and 3.7, the estimated bias is greater.

It is suggested to choose method $\mu_2$ when the number of samples is large. This is because the method is very fast computationally and the accuracy of the estimates is adequate for any application. However, if the sample available is small which is a normal case in radar applications, the KCFE method is to be used instead. Although it is more complicated, the estimator can give a reliable estimates, even at a sample size of 20.
Mean Square Error

Chapter 4. A Novel Approach to K-Distribution Parameter Estimation

4.20: The mean square error of the estimates of $\nu$ for $N = 128$.

Figure 4.20: The mean square error of the estimates of $\nu$ for $N = 128$.

4.21: The mean square error of the estimates of $\nu$ for $N = 256$.

Figure 4.21: The mean square error of the estimates of $\nu$ for $N = 256$. 
4.8. Discussion

Figure 4.22 and 4.23 show the independent trials of estimates of parameter $\nu$ for $N = 32, 64$, respectively. The estimates were obtained for 50 independent trials. In each case the K-distributed RV is generated as $\mathcal{K}(1, 0.1)$. A comparison with FLOM and Raghavan’s method shows that the estimates obtained from the KCFE method are always better than those of any other method. It can also be seen that the FLOM method shows large variability at this sample size where, in one case out of 50 ($N = 32$), $\nu$ is estimated as 0.9.

The Matlab source code for all of the non-ML methods described in this thesis are shown in Appendix C. The source code can be used as a guide on the computational complexity of each method. It is easy to see that the proposed moment method is the fastest computationally. Although the KCFE method is faster than Raghavan’s or the ML/MOM method but it is slower than FLOM method.
A new approach to estimate the parameters of the K-distribution has been proposed. The methods are derived from the newly developed RV whose PDF and CF are scale invariant. The scale parameter can be eliminated either by subtracting the logarithm of the K-distribution successively or by using the Hadamard transform. Both techniques can be used in different applications depending on whether the processing time or accuracy is more important.

Once the scale parameter has been eliminated, two methods have been proposed to estimate the shape parameter. In the general case, the method employing the variance of the transformed K-distributed clutter can be used. The method is very fast computationally and the performance is better than any existing moment-type method. If however a more accurate estimate is needed, or when the sample size is very small, the KCFE method can be used.
4.9. Conclusion

Although it is computationally slower, the accuracy of the estimates is worth the extra computation.

After the shape parameter has been estimated, the detection can be done by using the newly developed characteristic function. Using this method will greatly improve the performance of radar detection, as the new CF does not require one to estimate the scale parameter.
Chapter 5

Distance Function Via Characteristic Function

A useful way to think about similarity is by using the notion of distance between two data points. If the two points are similar, the distance between them is small. As their similarity decreases, the distance between them gets larger. It is frequently useful to define the distance between two points in a vector space. The usual way to define it is as the magnitude of the vector difference between the points. If \( x \) and \( y \) are vectors representing points in the same \( m \) dimensional vector space, the distance between the two points is defined as

\[
d_E(x, y) = |x - y| = \left( \sum_{i=1}^{m} (x_i - y_i)^2 \right)^{\frac{1}{2}}
\]  

(5.1)

where \( d_E(x, y) \) is known as the Euclidean distance.

This measure of similarity is only appropriate for measuring the distance between points in a case where each dimension has a similar range and variance. Unfortunately, it does not take into account the fact that dimensions with
larger scales and variances will dominate the distance measure. To overcome this, we need to normalise the ranges and variances of the different dimensions, where, one way to do this is to use a scaled Euclidean distance. This is done by dividing the squared difference in each dimension by the variance of that dimension, i.e.

\[
d_M(x, y) = \left[ \sum_{i=1}^{m} \frac{(x_i - y_i)^2}{\sigma_i^2} \right]^{\frac{1}{2}}
\]  

where \(d_M(x, y)\) is similar to equation 2.3 which is known as the Mahalanobis distance. Intuitively, this distance is a statistical measure that weights the difference in a given dimension according to how significant that difference is based on the estimated variance of the data in that dimension. When the covariance matrix of the data set which specifies the correlation between variables is invertible, then we can bring the Mahalanobis distance measure to bear on the problem.

Let us consider a 2 dimensional feature vector \(x\). A plot of such vectors is given in Figure 5.1, showing the equidistant curves for the Euclidean and Mahalanobis metric. The figure demonstrates the difference between the Euclidean and Mahalanobis distance, whereby the set of equidistant points as measured by the Euclidean metric forms a circle (or hyperspheres in the general case), while after taking into account the covariance matrix of the data, the Mahalanobis distance curve is an ellipse.

Unfortunately, the Mahalanobis distance requires one to assume that the feature vectors are Normally distributed, which is not the case for many applications. Also, with high-dimensional heterogeneous data the covariance matrix can become singular. Thus the Mahalanobis distance is not possible to compute. It is desirable to have a distance function that not only takes into account its range and covariance matrix, but also can be described by its probability
5.1 One Dimensional Statistical Distance

To understand further a statistical distance in the one dimensional case, let us consider a case where two RVs $X$ and $Y$ have the same PDF and mean but different variance, with $\text{Var}[X] < \text{Var}[Y]$. In the Euclidean space, the distance from point $x_1$ and $y_1$ (for example $x_1 = y_1 = 10$) to its mean are the same. However, the distance $J_x$ from point $x_1$ to its mean should statistically be greater than the distance $J_y$ from point $y_1$ to its mean. This is because, statistical distance should take into account the variance of the data where in this case, data $Y$ is further spread than data $X$. Figure 5.2 shows a comparison of two PDFs having the same mean and different variance.

The empirical characteristic function of an RV $X$ can be represented by the
Figure 5.2: Comparison of two PDF having same mean and different variance.

The operation of the estimator ECF is illustrated in Figure 5.3. For data \( x_1, x_2, \ldots, x_n \) and a given \( t \), the points \( (c_i, s_i) = (\cos tx_i, \sin tx_i) \) fall on the unit circle. The ECF of an RV is thus the mean of these points which is denoted in Figure 5.3 as \((\bar{c}, \bar{s})\). The modulus of the ECF \(|\hat{\phi}(t)|\) is simply its distance from the origin, also known as the resultant length. From the figure, it can be easily seen that at \( t = 0 \) the coordinate \((\bar{c}, \bar{s}) = (1, 0)\) and at \(-\infty < t < \infty\), \(|\hat{\phi}(t)| \leq 1\).

Without loss of generality, let us select a small value of \( t \) and assume \( X \) and \( Y \) are Normally distributed as \( \mathcal{N}(\mu, \sigma) \) where the characteristic function of the Gaussian distribution is given by

\[
\phi(t) = \exp \left( j\mu t - \frac{\sigma^2 t^2}{2} \right)
\]  

(5.4)
Next, two sets of independent Normally distributed RV with the same mean and different variance are generated, more precisely, $X \sim \mathcal{N}(1,1)$ and $Y \sim \mathcal{N}(1,4)$, and assigned an outlier to the data, for example $x_1 = y_1 = 10$. The plot of point $c_i$ against $s_i$ for these data is shown in Figure 5.4.

A small value of $t$ ($t < 1$) was selected to maximise the spread between the outliers and the background of the data in the complex plane. As expected, the outliers of the data (which are denoted by $\times$ in the figure) fall at the 'tail' of the plot, while its background clusters together. It is interesting to note that the Euclidean distance $J_x$ between the outlier and its ECF for the first data is larger than the Euclidean distance $J_y$ between the outlier and its ECF for the second data. This method thus can be used as a distance measure, where
a distance from any random variable \( x_i \) to its background can be calculated using the Pythagoras theorem,

\[
J_i^2 = (\bar{c} - c_i)^2 + (\bar{s} - s_i)^2
\]  (5.5)

where \( c_i = \cos tx_i \) and \( s_i = \sin tx_i \) are calculated from the sample. Since ECF is an unbiased estimator of the corresponding CF, the point \((\bar{c}, \bar{s})\) could either be the ECF or theoretical CF. When the underlying distribution of the RV is known, the distance can be calculated as

\[
J_i^2 = (\Re \phi(t) - c_i)^2 + (\Im \phi(t) - s_i)^2
\]  (5.6)

where for a Gaussian, \( \phi(t) \) is given in 5.4.

In many practical situations a designer is confronted with features whose value lies within different dynamic ranges. Thus, features with large values
may have a larger influence in the cost function than features with small values. The problem is overcome by normalising the features so that their values lie within similar ranges. A simple way of normalising a data is by using the respective estimates of the mean and variance, where after normalisation, features will have zero mean and unit variance. Thus by using only the real part of CF (or ECF), the distance function can then be simplified to

$$J_i = \Phi - \cos \frac{t(x_i - \bar{x})}{\sigma}$$

(5.7)

where $\bar{x}$ and $\sigma$ are the mean and standard deviation of the data, and $\Phi$ is the real part of CF or ECF. There are three criteria that must be fulfilled for any distance function to be valid. A distance function $d(x, y)$ is any scalar-valued function that satisfies the following conditions

- **Positive semi-definiteness**: $d(x, y) > 0$ if $x \neq y$ and $d(x, y) = 0$ if $x = y$.

- **Symmetry**: $d(x, y) = d(y, x)$

- **Triangular inequality**: $d(x, y) + d(y, z) \geq d(x, z)$

Since the proposed distance functions is simply an Euclidean distance in the characteristic function space, it is easy to see that the conditions are satisfied. The last condition which is known as the triangular inequality is a particularly strong constraint, where, functions meeting only the first two conditions may be useful in many analyses, although they are not true distance functions. The triangular inequality in the Euclidean plane is simply the assertion that going from $x$ to $z$ cannot be shortened by going first to point $y$, where equality corresponds to point $y$ being located on the line connecting $x$ and $z$.
5.2 Multidimensional Statistical Distance

In the multidimensional case, the definition of the empirical characteristic function and its basic properties are similar to those in the univariate case. Let $X_1, X_2, ..., X_n$ be independent $m$-dimensional random vectors with common distribution function $F(x)$, $x \in \mathbb{R}^m$. Denote the characteristic function of $F(x)$ by $\phi(t)$, and empirical distribution function associated with the sample $X_1, X_2, ..., X_n$ by $F_n(x)$. The empirical characteristic function associated with the sample $X_1, X_2, ..., X_n$ is defined as

$$\hat{\phi}(t) = \int_{\mathbb{R}^m} e^{jt \cdot x} dF_n(x) = \frac{1}{n} \sum_{i=1}^{n} e^{jt \cdot X_i}$$

(5.8)

As in the univariate case, the multidimensional ECF is an unbiased estimator of the corresponding characteristic function. At each fixed point it converges to the latter in the mean square sense [90]. If a sample under consideration consists of multidimensional random vectors, and it is known that their components are independent, then it is reasonable to use the product of one-dimensional marginal empirical characteristic functions

$$\prod_{i=1}^{m} \phi(0, ..., 0, t_i, 0, ... 0)$$

(5.9)

instead of the multivariate empirical characteristic function $\hat{\phi}(t)$. It thus follows from the univariate case that a distance from $X_1, X_2, ..., X_n$ to its background in $m$-dimensional space is given by

$$J_i = \sum_{i=1}^{m} (\Phi - c_i)$$

(5.10)

where, $c_i = \cos t \hat{X}_i$ and $\hat{X}_i$ is the normalised $m$-dimensional vector which is
5.2. Multidimensional Statistical Distance

The quantity $\Phi$ is the real component of the empirical or theoretical multidimensional characteristic function. In many cases, features can be assumed independent, thus instead of using the multidimensional CF, the vector of one-dimensional characteristic function can be used. Figure 5.5 shows the contour of equal distance of the proposed distance function for a 2-dimensional case. The bivariate data was generated for $X_{n=3000} \sim \mathcal{N}(0,1)$. The comparison given by

$$\hat{X}_i = (X_i - \mu)\Sigma^{-1}.$$  \hfill (5.11)
Chapter 5. Distance Function Via Characteristic Function

shows that the proposed method performs similarly to the Mahalanobis distance for bivariate normal data.

5.3 Conclusion

A new statistical distance function to measure the relationship of a data point to its background has been proposed. The method not only utilises the mean and variance of the data, but also takes into account the characteristic function of the data. A comparison of global distance for bivariate normal data shows that the proposed method performs similarly to the Mahalanobis distance.

The advantage of the proposed method is that it does not require one to assume that the data is normally distributed. One can choose to use the theoretical characteristic function, when the distribution of the data is known a priori. In other cases, the empirical characteristic function can be used which provides a distribution-free distance function.
Chapter 6

Application of K-Distribution Model in Target Detection

It is now time to look beyond the oversimplified Gaussian assumption and work with a more realistic non-Gaussian models. In this chapter, the techniques applied in the ATR processing steps were advocated in the method originally proposed by Messer et.al published in [49]. The novelty of our approach is the use of a K-distribution as a model instead of the Gaussian. A few modifications have been made to the technique presented in [49] so as to improve performance.

The ATR problem that we are facing is to detect a multiple target on the sea surface. The data was made available by DERA Farnborough. It consists of a sequence containing about twenty frames which have been artificially generated using a standard ray-tracing package. The first six frames of this sequence are depicted in Figure 6.1, together with their corresponding ground truth. It represents the scenario of a sensor attached to a ship looking out over the sea with five targets inserted into this sequence. The locations of these targets are given by the ground truth.
These targets are very small (typically one pixel) and represent missiles moving towards the observer. Finding the targets in this sequence is extremely hard, even for human observers. The target detection problem is viewed as an outlier detection problem, where anything that does not normally occur in the background can be considered as a potential target. The three basic steps proposed in [49] for the target detection algorithm are

- **Model generation** The homogeneous background is described using a
6.1 Preprocessing

As the sensor attached to the ship was looking out over the sea at some angle, the granularity of the texture in the image changes gradually from top to bottom. It seems advisable thus to pre-segment the image roughly into bands of basic statistical model.

- **Model optimisation** The model and model size are optimised using training data.

- **Target Detection** Outliers are found by deciding per pixel, whether it is accurately described by the model.

### 6.1 Preprocessing

As the sensor attached to the ship was looking out over the sea at some angle, the granularity of the texture in the image changes gradually from top to bottom. It seems advisable thus to pre-segment the image roughly into bands of...
homogeneous texture, where each horizontal band is then processed individually. Figure 6.2 shows the split image of the first frame. Although this is done manually in the experiment, one could find the region locations automatically, for example, by locally measuring the texture granularity and clustering the data observed.

6.2 Model Generation

Instead of using pre-defined features such as wavelets, the background can be described adaptively by using Principal Component Analysis (PCA). PCA, which is also known as the Karhunen-Loeve transform, finds a linear \((r \cdot c)\)-dimensional base to describe the dataset and the axes that retain the maximum amount of variance in the data. A PCA base is constructed by taking \(N\) random rectangles of size \(r \times c\) from a set of training images. These rectangles are then packed into a \((r \cdot c)\)-dimensional vector \(X_i\), in a row-by-row fashion, which results in a matrix \(X\) of \((r \cdot c)\) rows and \(N\) columns. Figure 6.3 shows the generation of a single 4-dimensional data sample \(x\) from a 2D image patch, where \(r = 2\) and \(c = 2\). The principal components are the eigenvectors of the covariance matrix of \(X\), which are the columns of the matrix \(E\), satisfying

\[
EDE^{-1} = XX^T
\]  

where \(D\) is a diagonal matrix containing the eigenvalues corresponding to the eigenvectors in \(E\). The set of 2D filters is then generated by unpacking each row of \(E^T\) into a filter of size \(r \times c\), where the PCA filters can be expressed as

\[
W = D^{-\frac{1}{2}}E^T.  
\]
6.3. Model Optimisation

In the feature extraction stage, PCA generates many different texture filters which have different responses as each has been tuned to the different image features present in the training frame. There are cases where one filter is sensitive to horizontal edges while another is sensitive to vertical edges. It was also found out that the first few filters with the corresponding highest eigenvalues capture the maximum intensity spread within the image. As a result, the targets lie in the middle of the intensity histogram, making target detection unreliable.

A problem is thus to find the features that maximise the distance between the background and target. Given a set \( Y \) of \( N \) measurements, \( Y_N = [y_1, y_2, \ldots, y_N] \) we wish to select a subset \( X_k = [x_1, x_2, \ldots, x_k] \) of \( k \) features, \( k < N \), such that each feature \( x_i \) is identical to a distinct measurement \( y_i \). We wish to find a set \( \chi_k \) which is optimal with respect to a criterion function \( J(\chi_k) \).
defined over all possible sets of $k$ out of $N$ measurements,

$$J(X_k) = \max_{X_k} (\chi_k)$$  \hspace{1cm} (6.3)

where $\chi_k$ denotes a candidate set of features, $\chi_k = \{\zeta_i | i = 1, ..., k, \zeta_i \in Y\}$. For this application, the distance function proposed in Chapter 6 is used as the criterion function which is evaluated as follows

1. Select a subset of features, $\chi_k$ for which the performance has to be evaluated.

2. Normalise the feature vector with respect to its mean and covariance matrix.

3. Estimate the parameter $\nu$ and $\omega$ of the background in this feature space.

4. Calculate the distance of all the target pixel features to the mean of the background pixel features by using the method described in Chapter 6.

5. The maximum distance over all the target pixels is the criterion value $J(\chi_k)$.

The maximum distance was chosen to ensure that the selected features maximise the distance between target and background.

Having defined a measure for the classification effectiveness of the feature vectors, the problem is now to search for the optimal subset from among all possible feature subsets. To optimise the distance function proposed in Chapter 6, the sequential floating forward search (SFFS) was implemented for this system. Although the algorithm does not guarantee finding all the best feature subsets, the SFFS method gives near optimal performance but with low computational cost. Since the algorithm is initialised by running the sequential forward selection algorithms (SFS), the parameters and the
characteristic function estimated in the early stage can be reused in the next hierarchy, thus avoiding the need to estimate the parameter in each stage. More formally, the algorithm implemented can be described as follows

**Input:**

\[ Y = \{y_i| i = 1, ..., N\} // \text{available measurements} // \]

**Output:**

\[ X_k = \{x_i| i = 1, ..., k, X_i \in Y\}, k = 0, 1, ..., N \]

**Initialisation:**

\[ X_0 := \emptyset; \quad k := 0 \]

Apply SFS twice and estimate the parameters of K-distribution in this space.

**Termination:**

Stop when \( k \) equals the number of features required.

**Step 1 (Inclusion)**

\[ x^+ := \arg \max_{x \in Y - X_k} J(X_k + x) \]

\[ X_k + 1 := X_k + x^+; \quad k := k + 1 \]

**Step 2 (Conditional Exclusion)**

\[ x^- := \arg \max_{x \in X_k} J(X_k - x) \]

if \( J(X_k - \{x^\}) > J(X_{k-1}) \)

\[ X_{k+1} := X_k - x^-; \quad k := k - 1 \]

else

\[ \text{go to Step 2} \]

\[ \text{go to Step 1} \]

where \( x^+ \) and \( x^- \) correspond respectively, to the most and the least significant feature in \( X_k \). Once the best features are identified for each feature set cardinality, a graph which plots the criterion function value as a function of feature subset can be used to select an optimal feature subset.
6.4 Target Detection

For every pixel in the image, we thus obtain a feature vector \( f_i = [y_0, y_1, \ldots, y_n] \) where \( 1 \leq i \leq N \) and \( N \) is the total number of pixels in the background area. Each \( y_k \) represents a measurement obtained by the \( k^{th} \) filter where \( 1 \leq k \leq N \). The target detection problem is viewed as an outlier detection problem, where the distribution of the \( N \) feature vectors is assumed to be K-distributed for the homogeneous textured region of the image. Since the PDF of a K-distribution is in a rather complicated form, its characteristic function can be used to model the background. Assuming that each feature vector is independent, the multidimensional CF of K-distribution is given by

\[
\phi(t) = \prod_{i=1}^{m} \left( \frac{1}{1 + \alpha_i^2 t^2} \right)^{\nu_i} \quad (6.4)
\]

where \( m \) is the number of dimension. Each feature vector, \( f_{test} \) is tested in turn to see whether it belongs to the same distribution as the background or is an outlier (i.e. possible target). This is done by using the distance measure developed in Chapter 6, where the parameters of K-distribution are estimated by using the proposed method, as described in Chapter 5.

6.5 Temporal Tracking

As the target will stay in approximately the same place from one frame to the next, a real target will be persistent across a sequence of frames while genuine noise is random and should disappear after just a few frames. To reward targets that appear in approximately the same location across a series of consecutive frames, each thresholded target image is dilated using a \( 3 \times 3 \) mask. This will allow a small movement of the target from one frame to the
6.6 Simulation Result

The geometric mean of \( t \) consecutive dilated images is then calculated, which is given by

\[
\mu_g = \left[ \prod_{i=1}^{t} x_i \right]^\frac{1}{t}
\]

(6.5)

where \( x_i \) is the dilated image in \( i \) frame. By doing this, the resultant image is already linearly scaled between 0 and 255. This resulting image displays the likelihood of certain target where the brighter the blob the more certain that the target found is real. Low intensity blobs can be dismissed as background noise, thus a second threshold is needed in this stage.

6.6 Simulation Result

For each regions of the segmented image, a different window size and set of filters were constructed. Starting from top to bottom (of the segmented image), the size of the random rectangles were set according to the granularity of the texture, which is as follow,

- **First band**, \( r = 3 \ c = 2 \)
- **Second band**, \( r = 5 \ c = 3 \)
- **Third band**, \( r = 7 \ c = 5 \)

This random rectangles were selected uniformly within the image.

Before we can proceed further, we need to established that the feature vectors extracted by using PCA filter can be modelled by K-distribution. Figure 6.4 shows the histogram of the feature vectors obtained from the first band of the first frame. For the sake of comparison, the 6-dimensional vectors are treated as a 6 univariate RV, and the histograms are compared with the Gaussian and K-distribution.
Figure 6.4: Histogram of the feature vectors (-) from frame=1 and band=1 . Comparison is made with K-distribution (---) and Gaussian(---).
Figure 6.5: The empirical characteristic function of the feature vectors (·). Comparison with theoretical K-distribution (---) and Gaussian(--).
The parameters of the theoretical PDF were estimated from a sample, where, for K-distribution, the method described in Chapter 4 was used. It can be seen that, K-distribution is a better model compared with the Gaussian distribution. A comparison of the theoretical and empirical characteristic function in Figure 6.5 also shows that K-distribution is a better fit.

The probability of detection was set to 0.99 to ensure that targets are not missed in the feature selection stage. Consequently, the probability of false alarm is comparatively high. This is acceptable as the false alarm will be reduced in the temporal averaging stage, where an average of five frames was used in both cases.

Figure 6.6 shows the temporally averaged results by using the proposed method. The corresponding results using the Mahalanobis distance are given in Figure 6.7. It can be seen that all five targets have been detected by using both methods, as the probability of detection was set close to 100%. However, the probability of false alarm for Mahalanobis distance is much higher than for the proposed method. The average false-positive rates for Mahalanobis distance is 10.36 which decreases dramatically to 1.45 under the K-distribution assumption. A plot of false positive versus frame number is given in Figure 6.10(a).

The good result achieved by using the proposed method motivate us to use a smaller number of frames for temporal averaging. This greatly improves the response time, thus providing us with the earliest possible warning. In the next experiment, three frames were used in the temporal averaging stage with the results shown in Figure 6.8 and 6.9 for the proposed method and Mahalanobis distance, respectively. It can be seen that the proposed method performed better. The average false-positive rate for the proposed method is 2.54 while for the Mahalanobis distance, the average is 61.82. A plot of false positive versus frame number is given in Figure 6.10(b).
Figure 6.6: The results for the first 8 frames obtained using the proposed method (5 frames temporal average).
Figure 6.7: The results for the first 8 frames obtained using Mahalanobis distance (5 frames temporal average).
Figure 6.8: The results for the first 8 frames obtained using the proposed method (3 frames temporal average).
Figure 6.9: The results for the first 8 frames obtained using Mahalanobis distance (3 frames temporal average).
6.7 Conclusion

In this chapter, we have demonstrated that K-distribution can be used to model the features extracted from the sea image by using PCA. Using histograms and empirical characteristic functions it was shown that K-distribution provides a better fit to the data, compared with the Gaussian model. As the probability of detection was set to 99%, all of the five targets have been detected using both method. However, the false alarm rate for the Gaussian model is larger compared with the K-distribution assumption. With a similar parameter setting, a significantly lower false-positive rate was observed when using the proposed method.

It is interesting to note that the proposed method performs similarly (in some cases better) to the 3D-PCA method described in [49] under the Gaussian assumption. In the 3D-PCA method, $d$ consecutive frames were used to construct a three dimensional filter in the feature extraction stage. Therefore, in a real life situation, one has to wait until three frames are captured be-
fore any feature processing step can be made. With the right model chosen, such method is unnecessary, where the 2D-PCA filter described in the previous section can be used, yet the results are similar.

With the great performance improvement shown by the proposed method, the number of frames used in the temporal averaging stage can be reduced. With a reduced number of frames, the movement of the targets from one frame to another may not be too significant, thus avoiding the need to dilate the thresholded image. This will not only assist the computation, but also the target detection using the original ‘undilated’ image will reduce the false alarm rate significantly. It has been found that some of the non-targets appear in approximately the same location in the consecutive frames, thus dilation will not only reward the real targets, but also the false ones as well. Furthermore, a fewer frame will mean a quicker detection (albeit computational complexity) thus providing an earliest possible warning.
Chapter 7

Conclusion

7.1 Summary

K-Distribution is a member of the 'long tail' density family, which has been shown to fit the sea clutter efficiently. Although it has been used to model the sea clutter in radar systems, the application of the distribution in other areas is still limited. With the newly developed technique to estimate the parameters and simplified characteristic function, the attitude towards the distribution is expected to change. As the proposed transformation successfully eliminates the scale parameter, the detection can be done by using the newly derived characteristic function, which is scale invariant, and thus avoiding the need to estimate the scale parameter.

The proposed estimator has been shown to offer a great performance improvement in terms of accuracy and variance. The method utilising the variance of the transformed RV is very cheap to implement, and can be used in an unmanned vehicle where human survivability is low. Typically, the log of K-distribution is readily available in the radar system employing a LOG/CFAR receiver. Thus, one only needs to do the transformation and estimate the vari-
ance of this transformed RV. The estimates can then be obtained by using a look-up table.

If the sample available is too small, which is a typical case in radar applications, methods employing the characteristic function can be used. Depending on the level of accuracy needed in different applications, one can choose whether to use ECF or KCFE. It is interesting to note that the magnitude of the characteristic function of the log of K-distribution is scale invariant and could also be used to estimate the shape parameter. Although the characteristic function method is computationally more intensive compared to the proposed moments method, the accuracy of the estimates is worth the extra computation.

The proposed scale elimination technique could also be used in estimating the parameters of different density functions. The expression for Weibull density is provided in this thesis as an example. Once the scale parameter has been eliminated, the estimation problem is down to a single parameter. Moreover, the detection can be made by using the characteristic function directly. Although the novel parameter estimation technique in this thesis was originally proposed to be used in radar systems, the application of the method in other areas could also improve performance.

As Gaussian assumption is made, the simplest way to design a classifier is by using the Mahalanobis distance. With the newly proposed distance function, a non-Gaussian assumption can be made. The distance function not only takes into account the mean and covariance matrix of the data, but also its characteristic function. If one chooses to use a distribution-free statistic, the characteristic function can be replaced by its sample counterpart.

The proposed parameter estimation technique and distance function have been applied to an automatic target recognition problem. The application of the method shows a great improvement in terms of higher detection rate and
lower false alarm, compared with the traditional Gaussian assumption. Since significant improvement is achieved, the number of frames used in the temporal averaging stage can be reduced thus allowing for the earliest possible warning.

7.2 Future Work

In this section, some research topics related to our study are presented for further development.

ECF-based Detection

The detection of known signals in the presence of additive noise has been traditionally approached as distinguishing between the goodness of fit of data to different PDFs corresponding to different hypotheses. Recently, Ilow et. al. [31] have introduced a new detector based on the empirical characteristic function, where the detection problem is viewed as a binary hypothesis testing.

By using his approach and the result presented in this thesis, the characteristic function of the transformed RV can be used to design a new detector. Since the characteristic function of the transformed RV is scale invariant, such detector will greatly improve the performance of a radar system.

ECF-based Distance Function

The proposed distance function takes into account not only the mean and covariance matrix of the data but also its characteristic function. Early results have shown a great performance improvement compared with the Mahalanobis distance. However, similar to Mahalanobis distance, for high-dimensional heterogeneous data the covariance matrix can become singular, making its inverse...
impossible to compute. It is thus desirable to remove the covariance matrix from the calculation and devise a distance function that uses only the characteristic function to measure the spread of the data. This could greatly improve the processing time, as the calculation of the covariance matrix is not required.

\textbf{\(\alpha\)-Stable distribution as a Model}

A broad and increasingly important class of non-Gaussian phenomena encountered in practice can be characterised by their impulsive nature. \(\alpha\)-stable distribution is an important model in non-Gaussian statistical signal processing and this distribution is characterised by its CF given by

\[ \phi(t) = \exp(jat - \gamma|t|^\alpha[1 + j\beta \text{sgn}(t)\omega(t, \alpha)]) \quad (7.1) \]

where

\[ \omega(t, \alpha) = \begin{cases} \tan \frac{\alpha \pi}{2} & \text{if } \alpha \neq 1 \\ \frac{2}{\pi} \ln|t| & \text{if } \alpha = 1 \end{cases} \]

A stable characteristic function is completely determined by the four parameters of the distribution: \(\alpha\) - characteristic exponent, \(\beta\) - symmetry parameter (symmetric if \(\beta = 0\)), \(\gamma\) - dispersion and \(a\) - location parameter. Unlike most statistical models, the \(\alpha\)-stable distributions do not have closed-form PDF except for the Gaussian (\(\alpha = 2, \beta = 0\)), Cauchy (\(\alpha = 1, \beta = 0\)) and Pearson (\(\alpha = 0.5, \beta = 1\)) distributions.

Stable distribution has been used in many applications, and the use of this model in pattern recognition community could lead to improve of performance. The attractive feature of this distribution is the inclusion of Gaussian distribution as a limiting case.
Appendix A

Mathematical Derivation

A.1 Scale Invariant K-Distributed RV

Let $X$ and $Y$ be two IID K-distributed RV, where the PDF of $X$ (or $Y$) is given by

$$p(x) = \frac{2}{a\Gamma(\nu)} \left(\frac{x}{2a}\right)^\nu K_{\nu-1} \left(\frac{x}{a}\right) \quad x > 0. \quad (A.1)$$

Let transformation $Z = \ln Y - \ln X = \ln \left(\frac{Y}{X}\right)$, the moment generating function of RV $Z$ is given by

$$m_Z(t) = \langle e^{zt} \rangle = \left\langle \left(\frac{Y}{X}\right)^t \right\rangle \quad (A.2)$$

$$= \int_0^\infty \int_0^\infty \left(\frac{y}{x}\right)^t \frac{2}{a\Gamma(\nu)} \left(\frac{x}{2a}\right)^\nu K_{\nu-1} \left(\frac{x}{a}\right) \cdot \frac{2}{a\Gamma(\nu)} \left(\frac{y}{2a}\right)^\nu K_{\nu-1} \left(\frac{y}{a}\right) \, dx \, dy$$

$$= \frac{2^{2-2\nu}}{a^{2+2\nu}\Gamma(\nu)^2} \int_0^\infty x^{\nu-t} K_{\nu-1} \left(\frac{x}{a}\right) \, dx \int_0^\infty y^{\nu+t} K_{\nu-1} \left(\frac{y}{a}\right) \, dy$$

Since [28]

$$\int_0^\infty x^\mu K_\nu(ax) \, dx = 2^{\mu-1}a^{-\mu-1}\Gamma \left(\frac{1+\mu+\nu}{2}\right) \Gamma \left(\frac{1+\mu-\nu}{2}\right) \quad (A.3)$$
for $\Re (1 + \mu \pm \nu) > 0$ and $\Re a > 0$,

$$m_Z(t) = \frac{\Gamma(\nu + \frac{1}{2})\Gamma(1 + \frac{1}{2})\Gamma(\nu - \frac{1}{2})\Gamma(1 - \frac{1}{2})}{\Gamma(\nu)^2}$$ \hspace{1cm} (A.4)

Since $m_Z(t)$ is differentiable at zero, then the first, second and fourth moments about the origin are given by

$$m_Z'(0) = 0$$ \hspace{1cm} (A.5)
$$m_Z''(0) = \frac{\psi'(\nu)}{2} + \frac{\pi^2}{12}$$ \hspace{1cm} (A.6)
$$m_Z^4(0) = \frac{1}{8}\psi^{(3)}(\nu) + \frac{3}{4}\psi'(\nu)^2 + \frac{\pi^2}{4}\psi'(\nu) + \frac{7\pi^2}{240}$$ \hspace{1cm} (A.7)

### A.2 Scale Invariant Weibull Distributed RV

The PDF of Weibull distributed RV is given by

$$p(x) = \frac{c}{b} \left(\frac{x}{b}\right)^{c-1} \exp\left(-\left(\frac{x}{b}\right)^c\right)$$ \hspace{1cm} (A.8)

where $b$ is the scale parameter and $c$ is the shape parameter. From transformation $Z = \ln Y - \ln X = \ln (Y/X)$, the moment generating function of RV $Z$ is given by

$$m_Z(t) = \left\langle e^{zt} \right\rangle = \left\langle \left(\frac{Y}{X}\right)^t \right\rangle$$ \hspace{1cm} (A.9)

$$= \int_0^\infty \int_0^\infty \left(\frac{y}{x}\right)^t p(x)p(y) \, dx \, dy$$

$$= \frac{c^2}{b^2} \int_0^\infty x^{c-1} \exp\left(-\frac{x^c}{b}\right) \, dx \int_0^\infty y^{c-1} \exp\left(-\frac{y^c}{b}\right) \, dy$$
A.2. Scale Invariant Weibull Distributed RV

Since [28]
\[
\int_0^\infty x^{\nu-1} \exp(-\mu x^p) \, dx = \frac{1}{\mu^p} \Gamma\left(\frac{\nu}{p}\right)
\]
(A.10)

for $\Re \mu > 0$, $\Re \nu > 0$ and $\Re p > 0$, the moment generating function of the scale invariant RV $Z$ is given by

\[
m_Z(t) = \Gamma\left(\frac{c + t}{c}\right) \Gamma\left(\frac{c - t}{c}\right).
\]
(A.11)

The first and second moments of $Z$ are given by

\[
m'_Z(0) = 0
\]
(A.12)
\[
m''_Z(0) = \frac{\pi^2}{3c^2}.
\]
(A.13)

Therefore $c$ can be estimated using its second order moment by

\[
c = \sqrt{\frac{\pi^2}{3\mu_2}}.
\]
(A.14)

From moment generating function, the CF of RV $Z$ is given by

\[
\Phi_Z(t) = \Gamma\left(\frac{c + jt}{c}\right) \Gamma\left(\frac{c - jt}{c}\right)
\]
\[
= \frac{\pi t}{c \sinh \frac{\pi t}{c}}.
\]
(A.15)

Thus, $c$ can be estimated by using ECF or KCFE

\[
c = \frac{\pi t}{\sinh(\pi t/c) \cdot \hat{\varphi}_Z(t; \varphi_Z)}
\]
(A.17)

where $\varphi_Z(t) = 1$ if ECF is used.
Appendix B

Original Formulation of Previous Method

In the original publication, a few authors use a different starting point for parameter $\nu$ in the PDF of K-distribution. Authors such as Raghavan and Iskander used the K-distribution which is described by PDF

$$p(x) = \frac{2}{a\Gamma(\nu + 1)} \left( \frac{x}{2a} \right)^{\nu+1} K_\nu \left( \frac{x}{a} \right) \quad x > 0, \quad (B.1)$$

where $\nu > -1$. In this thesis the first derived expression for K-distribution by Ward [94] is used. Here, the originally proposed algorithms by previous authors are presented as reference. From PDF B.1, the $k^{th}$ order moment of the K-distribution is given by

$$\mu_k = \langle X^k \rangle = \frac{\Gamma(0.5k + 1)\Gamma(\nu + 1 + 0.5k)}{\Gamma(\nu + 1)}(2a)^k \quad (B.2)$$

From ratio 3.8, the shape parameter $\nu$ can be obtained from
Appendix B. Original Formulation of Previous Method

\[ \nu = \frac{4 - \frac{\mu_4}{\mu_2^2}}{\mu_4 - 2}, \quad \text{provided } \frac{\mu_4}{\mu_2^2} \neq 2 \]  
\hfill (B.3)

In the case of fractional lower order moment, for \( p = 1/10 \), the expression for estimating \( \nu \) is given by

\[ \nu = \frac{1.05^2 - \frac{\mu_{21/10}}{\mu_{1/10} \mu_2}}{\mu_{1/10} \mu_2 - 1.05} \]  
\hfill (B.4)

In the Gamma density approximation technique, the parameter \( \beta \) of the equivalent gamma density may be expressed in terms of the parameter \( \nu \) by

\[ \beta = \left[ \frac{4(\nu + 1)\Gamma^2(\nu + 1)}{\pi \Gamma^2(\nu + 1.5)} - 1 \right]^{-1}. \]  
\hfill (B.5)

In GBK approximation technique, the solution of the maximum likelihood function is given as follow

\[ \hat{\alpha} = \frac{1}{2} \exp \left( \frac{\gamma - \psi(\hat{\nu} + 1) + 1}{N} \sum_{i=1}^{N} \ln(x_i) \right) \]  
\hfill (B.6)

and

\[ \sum_{i=1}^{N} \mathcal{K}(x_i) \ln \left( \frac{x_i}{2\hat{\alpha}} \right) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{K}(x_i) \sum_{i=1}^{N} \ln \left( \frac{x_i}{2\hat{\alpha}} \right) + N \]  
\hfill (B.7)

where,

\[ \mathcal{K}(x_i) = \frac{K_{\hat{\nu}-1} \left( \frac{x_i}{\hat{\alpha}} \right) + K_{\hat{\nu}+1} \left( \frac{x_i}{\hat{\alpha}} \right)}{K_{\hat{\nu}} \left( \frac{x_i}{\hat{\alpha}} \right)} \]  
\hfill (B.8)

thus the function \( g_k(\cdot) \) is given by

\[ g_k(\hat{\nu}) = \ln \left[ \frac{\Gamma(\hat{\nu} + 1)}{\Gamma\left(\hat{\nu} + 1 + \frac{k}{2}\right)} \right] + k\psi(\hat{\nu} + 1) \]  
\hfill (B.9)
Appendix C

Matlab Source Code

```matlab
function y=mom(x,p)
% x: K-distributed RV
% p: moment order

function y=raom(x,p)
if nargin<2
    disp('Order of moment needed')
    break
else
    N=length(x);
    p2=p+2;
    mu_2=sum(x.^2)/N;
    mu_p=sum(x.^p)/N;
    mu_p2=sum(x.^p2)/N;
    beta_p=mu_p2/(mu_p*mu_2);
    kb=(0.5*p+1);
    if beta_p~=kb
        y=p*(p+2)/(4*(beta_p-kb));
    else
        beta_p=beta_p+0.0001
        y=p*(p+2)/(4*(beta_p-kb));
    end
end
```

%/r/o method of moment , and
%/ Fractional moment by D R Iskander
%/ function y=mom(x,p)
%/ x: K-distributed RV
%/ p: moment order
%/ function y=mom(x,p)
if nargin<2
    disp('Order of moment needed')
    break
else
    N=length(x);
    p2=p+2;
    mu_2=sum(x.^2)/N;
    mu_p=sum(x.^p)/N;
    mu_p2=sum(x.^p2)/N;
    beta_p=mu_p2/(mu_p*mu_2);
    kb=(0.5*p+1);
    if beta_p~=kb
        y=p*(p+2)/(4*(beta_p-kb));
    else
        beta_p=beta_p+0.0001
        y=p*(p+2)/(4*(beta_p-kb));
    end
```

123
function y = rag(x)
N = length(x);
nu = [0.05 0.1:.05:5];
beta = nu;
rho_n = beta.*exp(-psi(beta));
B = (4.*(nu).*(gamma(nu).^2))./(pi.*(gamma(nu+.5).^2)) - 1).^(-1);

ma = mean(x);  % arithmetic mean
mg = geomean(x);  % geometric mean prod(x)^(1/N)
p = ma/mg;
Be = spline(rho_n,beta,p);  % use twice spline interpolation
y = spline(B,nu,Be);

function y = mlmom(x,k)
if nargin<2
    disp('Order of moment needed')
    break
else
    N = length(x);
    nu = [0.05 0.1:.05:5];
gk = log((gamma(nu))./(gamma(nu+(k/2)))) + (k*psi(nu))./2;

    eu = -psi(1);  % euler's constant
    mlk = sum(log(x.^k))/N;  % k-th order moment of log RV
    lmk = log((sum(x.^k))/N);  % log of k-th order moment

    gke = mlk - lmk + (k*eu/2) + log(gamma(1+ (k/2)));
    y = spline(gk,nu,gke);
% % ilow's method
% % function y=ilow(x,t)
% % x: K-distributed RV
% % t: value of t for CF
% %
function y=ilow(x,t)
N=length(x);
nu=[.05 .1:.1:2];
cf=(1./(1+((t.^2)./(2.*nu))).^nu); % CF of normalised K
U=rand(1,N).*2.*pi;
% transformation envelope
x=x.*cos(U);
% to quadrature component
x=x./std(x);
% normalised K
ecf=abs(sum(exp(x.*t.*i))/(N)); % ECF of normalised K
y=spline(cf,nu,ecf);

% % Lombardo's U-statistics
% % function y = lombardoU(x)
% % x: K-distributed RV

function y = lombardoU(x)

nu=[0.05 .1:.1:5];
U=psi(nu)-log(nu)+psi(1);

mlx=mean(log(x.^2)); % second moment of Log x
lmx=log(mean(x.^2)); % log of second moment x
u=mlx-lmx; % U-statistics
y=spline(U,nu,u);

% % Lombardo's V-statistics
% % function y=lombardoV(x)
% % x: K-distributed RV

function y=lombardoV(x)
m4=mean(x.^4); % fourth moment of x
m2=mean(x.^2)^2; % square of second moment x
V=m4/m2; % V-statistics
y=1/(V/2-1);
%%%% lombardo's W-statistics
%%%% function y=lombardoW(x)
%%%% x: K-distributed RV

function y=lombardoW(x)

nu=[0.01 0.1:.1:5];
W=trigamma(nu)+pi^2/6;

mlxs=mean(log(x.^2).^2);  \% mean of log(x^2)^2
smlx=mean(log(x.^2))^2;  \% square of mean log(x^2)
w=mlxs-smlx;  \% W-statistics
y=spline(W,nu,w);

%%%% Blacknell's zlogz
%%%% function y=zlogz(x)
%%%% x: K-distributed RV

function y=zlogz(x)

zlx=mean(x.^2.*log(x.^2));  \% mean of x^2*log x^2
z=mean(x.^2);  \% mean of x^2
lz=mean(log(x.^2));  \% mean of log x^2
X=zlx/z - lz;  \% X-statistics
y=1/(X-1);
% RV transformation from X to Z
% function y=rvz(x,method)
% x: K-distributed RV
% method: 1 for independent sets
% method: 2 for Z=Y_i-Y_{i+1}
% method: 3 for hadamard transform

function y=rvz(x,method)
if nargin<2
    disp('please choose method number')
    break
else
end
N=length(x);

if method==1
    y=diff(reshape(log(x),N/2,2)');
elseif method==2
    y=diff(log(x));
elseif method==3
    y=((hadamard(N)*log(x)')./sqrt(N/2))';
y(1)=[];
else
end

% Proposed method by variance
% function y=kestimate(x,method)
% x: K-distributed RV
% method: type help rvz for option

function y=kestimate(x,method)
N=length(x);
u=[0.01 0.1:0.1:5];
mu_2=psi(u,2)./2+pi^2/12;
z=rvz(x,method);
variance=var(z);
y=spline(mu_2,u,variance);

% RV transformation
% Variance of RV Z
% Estimate of nu
%%% Proposed method by characteristic function
%%% function y=kestimate(x,method,t,zmethod)
%%% x: K-distributed RV
%%% method: 1 for ECF or 2 for KCFE
%%% t: value of t for CF
%%% zmethod: type help rvz for option

function y=kestimate2(x,method,t,zmethod)

nu=[0.01 0.1:0.1:5];
tr=psi(nu,2);
A=(pi*t/2)/sinh(pi*t/2);

z=rvz(x,zmethod); % transformation from X to Z
N=length(z);
ecf=sum(cos(z.*t))/(N); % ECF of Z

if method==1
    la=(-log(ecf/A)*4/t^2); % ECF method
    y=spline(tr,nu,la);
elseif method==2
    ss=std(z)*(2/3/N)^(1/5);
    ec=ecf.*exp(-ss^2.*t.^2./2); % KCFE method
    la=(-log(ec/A)*4/t^2);
    y=spline(tr,nu,la);
end
Bibliography


