HUMAN AND ANIMAL CLASSIFICATION USING DOPPLER RADAR

by

Willem Daniël Van Eeden

Submitted in partial fulfillment of the requirements for the degree
Master of Engineering (Computer Engineering)

in the

Department of Electrical, Electronic and Computer Engineering
Faculty of Engineering, Built Environment and Information Technology

UNIVERSITY OF PRETORIA

September 2017

SUMMARY

HUMAN AND ANIMAL CLASSIFICATION USING DOPPLER RADAR

by

Willem Daniël Van Eeden

Supervisor(s):	Dr. J.P. de Villiers
Co-supervisor:	W.A.J. Nel
Department:	Electrical, Electronic and Computer Engineering
University:	University of Pretoria
Degree:	Master of Engineering (Computer Engineering)
Keywords:	radar, classification, Doppler, hidden Markov models (HMM), Gaussian mixture models (GMM)

South Africa is currently struggling to deal with a significant poaching and livestock theft problem. This work is concerned with the detection and classification of ground based targets using radar micro-Doppler signatures to aid in the monitoring of borders, nature reserves and farmlands. The research starts of by investigating the state of the art of ground target classification. Different radar systems are investigated with respect to their ability to classify targets at different operating frequencies. Finally, a Gaussian Mixture Model Hidden Markov Model based (GMM-HMM) classification approach is presented and tested in an operational environment. The GMM-HMM method is compared to methods in the literature and is shown to achieve reasonable (up to 95%) classification accuracy, marginally outperforming existing ground target classification methods.

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
AR	Auto Regressive
AIC	Akaike's Information Criterion
CMU	Carnegie Mellon University
CSIR	Council for Scientific and Industrial Research
CW	Continuous Wave
DFT	Discrete Fourier Transform
DPSS	Defence, Peace, Safety and Security
EM	Expectation Maximization
FFT	Fast Fourier Transform
FOPEN	FOliage PENetration
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
HRR	High Range Resolution
HTT	Hilber-Huang Transform
LDA	Linear Discriminant Analysis
LOS	Line Of Sight
LPC	Linear Prediction Coding
MAP	Maximum A-Posterior
MOCAP	MOtion CAPture
MTI	Moving Target Indication
RCS	Radar Cross Section

SAR	Synthetic Aperture Radar
SCR	Signal to Clutter Ratio
SNR	Signal to Noise Ratio
STAP	Space-Time Adaptive Processing
STFT	Short-Time-Fourier-Transform
SVD	Singular Value Decomposition
SVM	Support Vector Machine
ToT	Time On Target
UAV	Unmanned Aerial Vehicles
UHF	Ultra High Frequency
VHF	Very High Frequency

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

1.1 PROBLEM STATEMENT

1.1.1 Context of the problem

The detection and classification of ground based targets has a number of useful applications, ranging from security to military intelligence. Accurate radar classification systems could be used to monitor borders for illegal immigration and to detect poachers within nature reserves. They could also be used to monitor farmlands and nature reserves for livestock thieves. This thesis will focus on the monitoring of farmlands and nature reserves. Livestock theft has become a significant problem in South-Africa. For example, in 2012 approximately R430 million was lost to livestock theft [1]. Furthermore, the poaching problem has escalated drastically in the last decade [2].

The primary concern is to detect, track and classify human targets that are obscured by foliage clutter and who are walking in tall grass. It is infeasible to conduct aerial surveillance of the entire area of nature reserves, large farmlands and borders. In these circumstances, a stationary emplacement that is capable of automatic target classification is required.

1.1.2 Research gap

Recent research performed in the field has shown that it is possible to accurately classify ground targets based on their Doppler return and radar cross section when these targets are in the open [3–8]. Here, in the open is defined as being in clear sight of the radar with no objects in front of the target. However, very little research has been done on classifying targets that are obscured by clutter, such as dense

wooded areas, tall grass and crops. This calls into question the ability of existing classification systems to be applied in areas such as the South African bushveld. The problem now becomes tracking and classifying targets that are obscured by foliage.

It has been shown that lower frequency radar can penetrate foliage to some degree [9, 10]. However, at lower frequencies the observable Doppler frequency shift becomes much less prominent. Low frequency radars also typically require much larger antennas to achieve narrow beam widths [11]. Furthermore, at low frequencies the radar cross section of a target is small in size compared with the wavelength of the radar. This causes it to be difficult, if not impossible, to detect small targets such as humans using low frequency radar. Taking these facts into consideration, a second sub-problem is identified: If low frequencies are required to penetrate foliage but high frequencies are required for accurate classification, then at what frequency would the system be able to do both simultaneously and how accurate would such a system be?

1.2 RESEARCH OBJECTIVE AND QUESTIONS

This research is focused on designing a robust classification system for ground targets. The objective will be to analyze the effect that foliage, operating frequency and radar processing techniques have on our ability to classify ground targets accurately. The research questions follow:

- What is the accuracy by which lower frequency radar can classify targets?
- Is it possible to design a system that is able to classify targets obscured by foliage clutter?
- How accurate would such a system be?
- What limitations does such a system place on the radar itself?

1.3 HYPOTHESIS AND APPROACH

First and foremost, this research relies on the ability to gather sufficient data. The first approach will be to determine what data (if any) is available. Failing to obtain freely available data, the data will have to be recorded or created in some way. Once the data has been recorded, attempts will be made to create a classifier that is capable of distinguishing between humans and animals based on their micro-Doppler features.

1.4 RESEARCH GOALS

- Determine if it is possibly to classify targets at lower radar frequencies.
- Develop a classification system that is able to distinguish between different micro-Doppler motions.

1.5 RESEARCH CONTRIBUTION

This work develops a method for classifying micro-Doppler motions that have been gathered by a radar system and which is able to accurately distinguish between humans and animals in a foliage clutter environment. As far as the author is aware, no such system has previously been designed and tested in a cluttered environment or at ranges exceeding 1km. Furthermore, as far as the author is aware, no research has previously investigated whether individual animal species can be distinguished based on their Doppler profile.

1.6 OVERVIEW OF STUDY

This thesis has 6 chapters, each dealing with different aspects of the work conducted. Chapter 1 is an introduction, and it discusses the approach and goals of the work. It will also give a brief overview of the studies that were conducted.

Chapter 2 is an extensive literature study, which reviews the existing radar systems and the basic approaches that have been developed for Doppler-based classification, including those areas where Doppler signal classification has been used. This chapter also includes an evaluation of the many different classifiers that have been used, as well as their advantages and disadvantages.

Chapter 3 details the basic background theory that is relevant to this work, which includes radar processing, different types of radar that could be used and the effects that are present within the data. It also gives a basic discussion of the Doppler effect.

Chapter 4 evaluates a method for simulating Doppler motion from humans and it determines whether it could be feasibly implemented. Mel-Cepstrum coefficients are extracted from the simulated and measured data, and the simulated human data is compared to measured data at both X- and S-band, based on the mel-Cepstrum coefficients that have been extracted.

Chapter 5 describes the development of a classification scheme for the measured radar data that was obtained by the CSIR in the process of developing an X-band system for the Kruger National Park. The data is described and separability is analysed, a classification scheme is developed to classify targets and, finally, several different tests are performed on the data to determine the effect that different parameters have on the classification accuracy and false alarm rate. A method for using confidence metrics to reduce false alarm rates is also studied.

Chapter 6 draws some conclusions about the work and it will also make several recommendations for future work.

CHAPTER 2 LITERATURE STUDY

2.1 CHAPTER OVERVIEW

This chapter discusses the current state of the art of ground target classification and radar systems. The chapter starts by discussing existing automatic radar classification systems. The commonly used classification models are then discussed and finally the different features used for radar target recognition are discussed.

It is found that existing automatic radar classification systems use a wide variety of classification methods. The most promising among them are the Gaussian Mixture Models (GMM) suggested by [6]. It is further found that most existing systems only classify targets at a very high SNR and at short ranges, this leaves questions as to whether the results can be reproduced in realistic operational environments. Finally, in cases where high bandwidth is not available in the deployed radar system, it is found that an audio based classification scheme using Doppler features, specifically mel-cepstrum coefficients, should result in a system with satisfactory performance.

2.2 EXISTING AUTOMATIC RADAR CLASSIFICATION SYSTEMS

Existing radar classification systems use a variety of classification methods. The most prominent among them are the Support Vector Machines (SVMs) [3, 12, 13], Bayesian classifiers [4], Hidden Markov Models (HMM) [5, 14, 15] and GMM [6, 7]. Artificial Neural Networks (ANN) have also been used successfully [16, 17]. Some of these neural networks have even been trained on raw spectrogram data [18]. In most applications, the classifier requires a long Time on Target (ToT); for example, up to 1 second for accurate classification. In [8, 9] it is shown that the accuracy of Doppler based classification

systems is directly influenced by the coherent integration time of the radar system. Although it was shown that increased ToT would improve classification accuracy, a ToT of greater than 4 seconds would not yield a significant improvement. Very few (if any) of these studies have been performed on animal and human targets observed in their natural habitat, obscured by trees and grasslands.

Almost all existing systems classify targets at a very high SNR at short range and in clear Line Of Sight (LOS) of the sensor. In this low noise clear LOS environment, the existing systems perform extremely well and accuracies in the high 90% are common, regardless of the classifier or the features. However, it is uncertain whether these results would be consistent at higher noise and when targets are moving through cover. This raises doubts about whether existing systems would be usable in areas such as the South African bushveld.

2.3 COMMONLY USED CLASSIFICATION MODELS

Many different models have been proposed for classifying different forms of motions. Here, a short overview of each is given and the advantages of each are described. It should be noted that, given no information, there is no reason to favor any of these classifiers above another. This theorem is called the "no free lunch" theorem [19]. Classical linear combinations and fixed basis functions are capable of classifying almost any dataset but they suffer from the curse of dimensionality [20]. There are many ways to get past the curse of dimensionality. The models discussed in this section each solve this problem and they have different advantages above normal linear basis functions. Nonetheless, even rudimentary linear discriminant functions have been used for classification of radar Doppler targets [6].

2.3.1 Support vector machine

Support vector machines are decision machines and they do not provide probabilistic outputs [20]. They function well on small sets of data, just like other sparse kernel techniques. They have been frequently used to classify human motion exactly for this reason and they perform relatively well on these small datasets [3, 8, 12, 13].

2.3.2 Mixture models

Mixture models are typically used to cluster data into classes. The most prominent amongst these is the GMM, which is widely used in pattern recognition, machine learning and statistical analysis. The parameters are often derived by the Expectation Maximisation (EM) algorithm. [6–8].

2.3.3 Graphical models

Probabilistic graphical models, such as Bayesian networks and Hidden Markov Models (HMM), offer several useful properties for the classification of data [20]:

- They provide visualisation of the structure of the model, and they can be used to design and motivate new models.
- Complex computations can be expressed in terms of graphical manipulations in which underlying mathematical expressions are carried out implicitly.

Bayesian networks allow the user to create links between observations and class, allowing features and their link to classifier output to be explicitly shown in the network. HMMs can be seen as an extension of the Bayesian network, allowing for change over time or state transition to be incorporated into the model. HMM's and Bayesian methods have been used in [4, 21–23].

2.3.4 Neural networks

Neural networks solve the curse of dimensionality by fixing the number of basis functions in advance but allowing them to be adaptive. The parameter values are adapted during training. A neural network comprises multiple layers of logistic regression models [20]. Neural networks have their origins in trying to represent the way that the human brain processes data [24, 25]. Over the years, many different training methods have been developed for neural networks [19, 20] and they have been used for many different classification problems. Neural networks are often considered to be a "black box" solution. A dataset is gathered and a feed forward network with some number of nodes is developed and trained. Although highly successful in their ability to classify almost any data, they give very little information

regarding what features were useful and which ones should be ignored. In addition, unless the network is trained as in [26] with Gaussian activation functions, the system typically gives no indication as to its certainty about a specific classification.

2.3.5 Kernel methods

Typically, when training a classification system, the training data is used to train the model parameters and then the training data is discarded. There is, however, a set of methods that retain a subset of the training data, which are referred to as kernel methods. The K-nearest neighbour methods and particle filters are examples of these methods. Particle filters have been used to classify tracks and they include secondary information that could not be incorporated into the Kalman tracking filters, such as RCS [23, 27].

2.4 FEATURES USED IN RADAR CLASSIFICATION

Next, the features used in classification are analysed. The most common features used in the classification of ground based targets are Doppler signatures [3, 5, 6, 11], Radar Cross Section (RCS) [27] of the target and the range extent [4, 28–30]. The movement of the target could also be used to classify the target and this has been done successfully for many types of targets [31, 32].

By far the best performance for classification in the literature is found when using features extracted from Doppler. This performance can be improved by including the information from the tracker [4, 33] or including features from the high resolution range profile [34].

2.4.1 Doppler features

Radar uses the Doppler frequency to extract the target's radial velocity, and to distinguish between moving and stationary targets. The Doppler phenomenon describes the shift in the centre frequency of an incident waveform due to the direction of the target motion with respect to the source of radiation [35].

Different targets tend to have unique movement models. This uniqueness can be used as a very good feature when classifying targets. Human motion is by far the most unique feature that can be extracted from human targets and the human gait has been extensively studied. The same principles can be used to extract features from different vehicle types. For example, wheeled vehicles differ significantly in their movement profile from tracked vehicles [3].

A higher carrier frequency is required to use Doppler motion, which results in a shorter wavelength. Targets smaller than half the wavelength of the radar tend not to reflect as strongly. Consequently, a higher frequency allows smaller objects and, hence, smaller body parts to reflect more strongly. A higher carrier frequency, therefore, allows for better characterisation of the Doppler spectrogram. The Doppler motion used for classification is typically referred to as micro-Doppler and it is caused by the secondary scatterers on the target, such as: rotations of propellers, swinging arms and legs on animals, and vibrations on trucks and cars [36]. Therefore, the observable micro Doppler motion significantly deteriorates in low frequency radar, such as VHF/UHF radar [37].

There are two main approaches when using the Doppler signature of a target for classification. The first is treating the classifier as an image after plotting the Doppler spectrogram, the second is to treat the Doppler signal as an audio signal.

2.4.1.1 Audio classification approach

Many audio classification methods can be implemented to classify the Doppler signal when the signal is treated as an audio signal. Initial designs for audio based classification simply used the spectrum of the target directly as the feature vector [5]. Later, more advanced feature sets were extracted from the signal. In particular, Linear Discriminant Analysis (LDA) [14], Mel-Cepstrum Coefficient extraction [6], bispectrum estimation [7], and Linear Prediction Coding (LPC) [6] have all been successfully used for human gait recognition.

2.4.1.2 Image processing approach

The classification problem can also be solved by considering the Doppler spectrogram directly as an image, the signal is first cast as an image by performing Short-Time-Fourier-Transform (STFT) to

create a spectrogram of the signal. Spectrograms can be useful analysis tools, even when not using image based classification approaches.

This approach is typically started with time-frequency analysis. The STFT is defined as:

$$X[n,k] = \sum_{r=-\infty}^{\infty} x[r]w[r-n]e^{\frac{-j2\pi rk}{N}}, k = 0,1,\ldots,N-1, \tag{2.1}$$

where n is the time index, k is the frequency index, N is the number of frequency points and $w[n]$ is the window function used, this is typically a Hanning or Hamming window. The window length is chosen such that the signal is stationary over the time period. The spectrogram is then computed by taking the squared magnitude of the STFT

$$S(n,k) = |X[n,k]|^2. \tag{2.2}$$

In [3] two-directional, two-dimensional Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) is performed on a window around the torso frequency. Singular Value Decomposition (SVD) has also been performed on these spectrograms [23] and pseudo-Zernike polynomials have been extracted [38].

2.4.2 Radar cross section

RCS is mainly used as an indication of the size of the target being tracked. This could be an extremely useful feature because it will allow us to accurately separate humans and small animals from other much larger targets. RCS can, however, be an unreliable feature because it varies with range, target aspect angle and many other factors [27]. The significant variation of RCS causes it to be much less useful and, unless significant care is taken to normalise with respect to range and aspect angle, the RCS will vary too significantly to use.

2.4.3 Movement model

The model for tracking in an Interactive Multiple Model (IMM) tracker is also sometimes used [32]. In this method of classification, the target has a certain movement model associated with it in the tracking module. This model gives us information about the movement characteristics, such as velocity, acceleration and the turning radius of the target. Obviously land light vehicles have different movement models from tanks or trucks, and these also differ significantly from humans or animals moving on

the ground. Using Dempster-Shafer evidence reasoning, it is possible to consider the uncertainty of a given hypothesis; thereby, allowing for classification of the target [32].

2.4.4 High range resolution profile

High Range Resolution (HRR) radar is a processing technique that is used to extract a target signature [39]. The signature is formed as a measure of the wideband energy reflected to the radar based on the range of the target. HRR radar techniques enable users to classify targets spatially by creating a range profile for the target. This range profile is then compared to a library of signatures and a nearest neighbour approach is used to classify the target [28]. It is possible to classify targets by building a library of known targets and comparing new targets with the library. However, it is important to note that the target orientation would cause the profile to be different [29, 30]. For example, the range profile for a vehicle facing the radar would be significantly different from a vehicle that has its side turned towards the radar. A very large number of templates would have to be stored or models would have to be created to accommodate almost any orientation with regards to the radar. Generally, there are several sets of orientations for which the scattering physics vary slowly, and the HRR waveforms can be approximated as stationary [40, 41]. The most common method is to develop an Auto Regressive (AR) model to determine the orientation and then, based on an initial orientation prediction, the subsequent measurements are compared to the expected orientation [42]. However, HRR requires an extremely high bandwidth [35].

2.5 COMMON USES FOR DOPPLER CLASSIFICATION

Doppler classification of moving targets has been used in many different situations. This section covers a few of the proposed uses that have been developed in the literature. Most research does not give explicit uses for the work that they propose. This is a fatal shortcoming because the environment that it will be used in directly impacts the methods that are viable. For instance, if you wish to classify a target in a scanning environment, then the ToT plays a significant role in the design of the classifier and radar. However, if the proposed design is for a staring application, then ToT plays almost no role in the system consideration.

2.5.1 Perimeter security

Short range motion detection sensors have been developed in recent years and they are capable of detecting targets approaching sensitive areas, such as airport grounds. They are preferred to normal camera systems due to their all weather any time functionality. They are not dependent upon light conditions or even weather, and they can typically function in dense fog and rain. However, these systems have a very high rate of false alarms; for example, stray dogs and animals can frequently trigger the sensor. To help combat this high rate of false alarms, motion classification has been implemented to distinguish between animals and humans [4].

2.5.2 Border security and military intelligence

In the surveillance of large areas, such as borders and nature reserves, it is impractical to patrol the entire area or survey with Unmanned Aerial Vehicles (UAVS). Consequently, recent development has focused on detecting ground moving targets with radar systems and to distinguish ground moving personnel from vehicles. In the surveillance of borders, it is crucial that this system is able to distinguish humans from other slow moving animals [5].

2.5.3 Fall detection for the elderly and medical diagnosis

Very fine motor vibrations can be detected with high frequency radar and used to distinguish between trucks, tanks and different types of vehicle. This has led to research focusing on the use of motion detection for medical diagnosis, which attempts to identify disease based on the micro-vibrations that are present in the movement of patients. At the same time, short range motion detectors have been developed for implementation in old age homes and hospitals for the elderly to detect if a fall has occurred so that help can be immediately dispatched to the patient's room [43, 44].

2.6 CONFIDENCE SCORES IN CLASSIFICATION

Confidence scores have been investigated to increase accuracy and to reduce false alarm rates. Confidence scores are frequently used in classification systems to reduce the false alarm rates in systems

where the cost of a false alarm drastically outweighs the cost of simple human intervention or waiting for more data. Specifically, confidence scores are often used in bio-medicine, manufacturing and in product quality control [45, 46]. The principle behind using confidence score is that an incorrect classification will tend to have a lower confidence score than correct classifications. Thus, if a good confidence score can be derived, then it could be used to reject incorrect classification, marking them as unsure and requesting further evaluation by secondary classifiers or human operators.

CHAPTER 3 BACKGROUND THEORY

3.1 CHAPTER OVERVIEW

This chapter discusses the background theory related to this study. These basic principles are required to understand the work that has previously been documented. This chapter briefly discusses the Doppler effect, the effects of clutter on a radar signal, as well as methods to deal with clutter. It is found that the effects of clutter will be highly dependent upon the operating frequency of the radar system. Finally basic tools and approaches for simulating radar targets are discussed and it is found that the human kinematic walking model can be used along with very basic RCS models to simulate the radar return of a walking human.

3.2 THE DOPPLER EFFECT

During this work, the Doppler motion induced onto the transmitted wave of the radar by the target is used for classification. Consequently, it is necessary to understand the Doppler effect in detail. The Doppler effect is defined as the change in frequency or wavelength of a wave for an observer who is moving relative to its source. In radar systems, Doppler frequency is used to extract target radial velocity. It is also used to distinguish between moving and stationary targets [47]. The Doppler shift f_d is related to the wavelength λ of the target by the following equation:

$$f_d = \frac{2v}{\lambda}, \tag{3.1}$$

where v is the speed of the target. In more general terms, if the angle θ between the target and the radar is known, then the Doppler shift of the target can be expressed as:

$$f_d = \frac{2v}{\lambda}\cos\theta. \tag{3.2}$$

This effect is illustrated in Fig. 3.1.

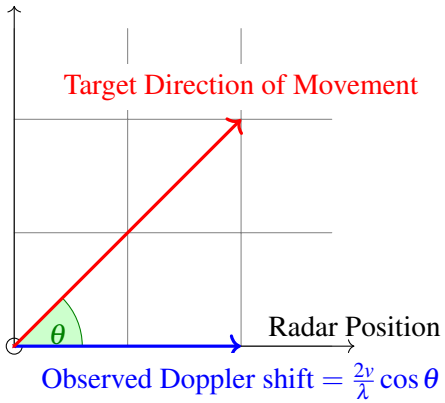


Figure 3.1. Diagram illustrating the Doppler effect. The target is at the origin, walking the direction as indicated by the red line.The blue line can be seen as the reflection from the target in the direction of the Radar. The target is moving at an angle to the radar and, therefore, a receiver on the right will see a positive frequency shift because the signal is compressed due to the movement of the target and a receiver on the left will see a negative shift. Any other incidence angle will cause a shift proportional to the angle of incidence θ. A target moving perpendicular to the radar will not induce any Doppler shift and will, therefore, remain undetected.

An individual that is moving will induce a different shift upon the transmitted wave from the radar for each of their body parts [48]. These effects are then added together to form a wave that has multiple different shifts on it. The intensity of each shift and the relative size of that shift indicates the size and speed of the individual body parts, which varies over time as the target moves. This effect can be analysed in more detail by using a spectrogram. A spectrogram is derived from the STFT. The signal is divided into small segments and then an Fast Fourier Transform (FFT) is performed on the segments to determine the frequency shifts visible in that segment. These segments are placed next to each other in order, with the Z-axis (colour) creates a very memorable image of target motion that can be analysed.

3.3 SURFACE CLUTTER

Given that the targets will be recorded in cluttered environments, it is essential to understand the effects that clutter will have on the received waveform.

Foliage, such as trees, have a tendency to sway in the wind. This causes trees to obscure slow moving

targets even further because simply removing all zero velocity targets will not remove swaying trees and removing higher velocity targets will also remove valid slow moving ground targets.

The polarisation of the radar becomes important when considering foliage penetration. It has been shown that different polarisations will scatter from different parts of the foliage. Horizontal polarisation usually scatters from the branches and soil. Vertical polarisation will usually scatter more from the tree trunks [49–51].

Clutter is a term used to describe any object that may generate unwanted radar returns. In this case, clutter consists of surface clutter—such as trees, rocks and vegetation—or volume clutter—such as rain, birds or insect swarms. Clutter echoes are typically random and are expressed in terms of their statistics. Average clutter Radar Cross Section (RCS) is given by:

$$\sigma_c = \sigma^0 A_c, \tag{3.3}$$

where $\sigma^0(m^2/m^2)$ is the clutter scattering coefficient and A_c is the area of the clutter. The Signal to Clutter Ratio (SCR) is given by

$$SCR = \frac{\sigma_t F_t^2 F_r^2}{\sigma_c F_c^2}, \tag{3.4}$$

where F_c is the clutter propagation factor, F_t and F_r are the transmit and receive propagation factors. In most cases, $F_t = F_r$ and σ_t is the RCS of the target.

Radar clutter power is typically received from all clutter within the processing bin. The SCR from all of the clutter can be expressed using the radar equation as:

$$SCR_{Ac} = \frac{2\sigma_t \cos\Psi_g}{\sigma^0 \theta_{3db} R c \tau} \tag{3.5}$$

where θ_{3db} is the radar beam width, σ_t is the RCS of the target, Ψ_g is the grazing angle, σ^0 is the clutter reflection coefficient, R is the range, c is the speed of light and τ is the pulse width.

It can be seen from (3.5) that there are a number of choices of how to reduce the effect of clutter. For example, the grazing angle of the radar can be raised by placing it on a high hill or a large emplacement. Alternatively, narrower pulses can be used at the expensive of increased bandwidth. Finally, a narrower beam width would also reduce the effects of clutter; however, a narrower beam width would come at the cost of increased antenna size [47]. These effects are illustrated in Fig. 3.2.

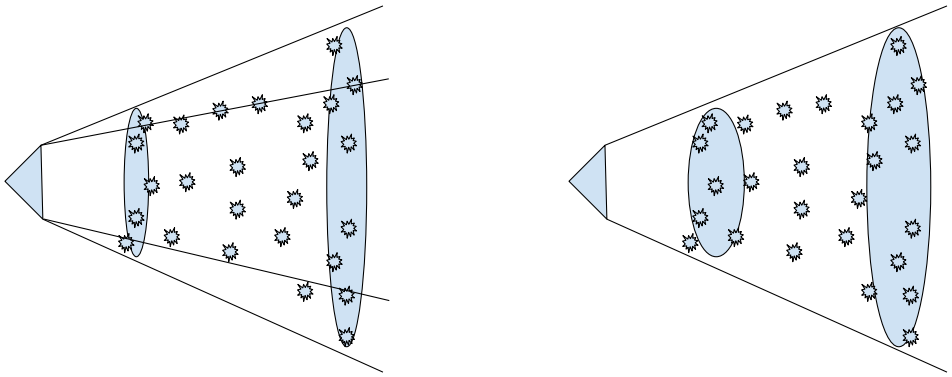


Figure 3.2. Illustration of SCR effects. The triangle represents the radar transmitter while the marks indicate trees or shrubs within the beam of the radar. From the image it can be seen that as the range increases more clutter falls within the processing bin because the bin increases in size based on the range of the target, the pulse width of the radar and beam width of the radar.

Next, the common methods for suppressing clutter will be evaluated. Since clutter in the scene slowly moving from side to side will also induce a Doppler return, clutter suppression techniques will have to be studied.

3.4 CLUTTER SUPPRESSION TECHNIQUES

There are a number of ways to mitigate clutter. The simplest of these is simple delay line cancellation and coherent signal processing. Stationary targets can be removed by analysing the Doppler shift caused by moving targets when a pulse Doppler radar is used. However, in the presence of foliage clutter, the target returns will only be suppressed and not entirely removed because tree branches and so on might be moved by the wind. To remedy this situation, extremely slow moving targets are also usually ignored. This allows slow moving targets to also avoid detection. Another method for clutter suppression is to develop the classification system to be able to recognise clutter and then suppress this output to the user. This technique is widely used in the existing classification systems, with clutter almost always being considered as one of the classes that need to be classified [3–6]. The main type of radar that is used will be investigated next.

Clutter suppression will not remove all of the clutter from the scene and it could possibly allow targets to escape detection if they move slow enough. Consequently, it is important to understand what happens to a radar signal as it passes through foliage.

3.5 FOLIAGE PENETRATION PHENOMENA

This section will discuss the foliage penetration phenomena. Extensive research was done on the effects of foliage on radar in the early-1990s. There are a number of phenomena that occur when radar signals penetrate dense foliage. First, signal attenuation occurs, resulting in a decrease in Signal to Noise Ratio (SNR) [51]. It was found that at VHF frequencies (200 MHz) foliage attenuation was much less than at any other frequency. This could be explained by the fact that the wavelength was very large when compared to the trunks and branches of the trees. It was also found that rain forests caused much more attenuation than northern latitude forests, meaning that the density of the forests played a role in the amount of attenuation that the signal experienced. The most important conclusion of the foliage penetration research that was conducted in the literature was that it was infeasible to try and detect targets within dense foliage at frequencies higher than L-Band [49]. The expected loss from penetrating foliage, such as dense woodlands, is approximately 20dB loss per 100m of foliage at UHF frequencies [49].

Phase shift and polarisation are other effects that occur due to foliage penetration of radar signals. Radar waves reflecting off the trees scatter and this causes an observable effect in the polarisation of the signal. These effects are diverse and will have to be kept in mind when developing FOPEN systems. Phase error has to be considered when trying to develop a SAR image from the data and it should have less of an effect on stationary emplacements because the phase shift effects it less. It was found that UHF had the least phase error and that the standard deviation of the phase error was less than a quarter of a wavelength [49–51].

Finally, normal Doppler extraction techniques are much less effective because the amount of backscatter and attenuation of the signal through foliage. One proposed solution is the Hilbert–Huang transform [52]. Although this transform allows for a much better signal to clutter ratio, this typically comes at the cost of a much higher required bandwidth, which is something the available radar systems do not have access to.

These studies where conducted in dense rain forest and northern woodlands. Although no formal studies have been performed on propagation studies of radar in savannah environments, the savannah regions of the borders are not as dense as the forests investigated in these studies. Tall grass should not significantly obscure targets at S-band. Interaction with game rangers has also indicated that poachers

tend to avoid heavily overgrown areas because traversing these terrains may prove difficult and they usually use animal paths. Consequently, a system operating at X-band is also feasible while accepting that the legs of the target may be obscured, leaving only the main body and arm sway visible.

3.6 FREQUENCY-MODULATED CONTINUOUS WAVE RADAR

Frequency-Modulated Continuous Wave (FMCW) radar is a special type of radar that is frequently used. The data described in Chapter 5 was recorded using an FMCW radar.

Similar to a CW-radar, a FMCW radar continuously transmits power; however, it is capable of changing its operating frequency during measurement. The distance measurement is accomplished by comparing the frequency of the received signal to a reference (this is usually simply the transmission signal) and the duration of the transmission signal is substantially greater than the required receiving time for the distance measuring range.

FMCW radar has the ability to measure very small ranges to the target. It is also able to both measure the target range and velocity (Doppler content of signal) simultaneously and it has a very high accuracy of range. There are many design considerations when developing a FMCW radar that fall outside the scope of this work. In general, the Doppler return for a FMCW radar can be treated much like that of an MTI-CW radar, which just has the advantage of also having an accurate range measurement and no downtime between pulses. This allows coherent integration over separate bursts without having to worry about discontinuities between bursts. Most modern radar systems are FMCW systems [47].

3.7 SIMULATING RADAR MEASUREMENTS

Most radar systems are expensive to operate and deploy, and initially obtaining data was difficult. Consequently, simulated radar data has been used to assist many studies. Simulation also gives insight into data and it is usually a useful tool to when conducting studies. In Chapter 4, radar data is simulated and compared to a small amount of available measurements.

One solution for the problem of limited data is training and testing classification systems on artificial datasets. In the case of radar, this can be achieved by simulating the radar backscatter of a target using simulation software, such as FEKO and SigmaHat. FEKO is a commercially available software suite that is developed by EMSS S.A. (Pty) Ltd. and it is widely used for antenna analysis. SigmaHat is a software tool that is developed by the CSIR DPSS for the prediction and analysis of EM scattering of electrically large complex objects [53]. These types of packages have been used to model ships and aircraft, and they have been shown to be good approximations for real observations. However, when used on ground targets, it is difficult to model the ground plane of the earth and this results in a simulation that does not necessarily correlate well with reality. In the application of simulated models, care will have to be taken to ensure that any observations from the simulation that could be used for classification are not artefacts of the simulation process.

The most common way to simulate radar Doppler measurements is by summing the approximate back scatter of each body part individually. Each body part is then approximated as a cylinder and the backscatter of each of these cylinders is added over time to approximate human motion. Doppler motion could be approximated by combining this data with a human kinematic walking model [9,10,48].

One of the problems with these simulations is their inability to simulate clutter, such as trees around the target. To model the target inside foliage clutter, a statistical model of the clutter will have to be applied to any measurement.

CHAPTER 4 SIMULATING HUMAN DOPPLER MOTION FROM MOTION CAPTURE

4.1 CHAPTER OVERVIEW

Owing to an initial lack of data, alternative data sources had to be found. This chapter documents efforts by the author to simulate radar Doppler data from existing motion capture databases. To establish the viability of this type of data source, cepstral coefficients extracted from this data are compared to cepstral coefficients extracted from a small amount of measured data that was available from the Council for Scientific and Industrial Research (CSIR) Radar and Electronic warfare research group. A simulation for human Doppler response is developed based on the Carnegie Mellon University Motion Capture database (CMU MOCAP) [54]. This data is used to simulate human Doppler response as it would be seen by a radar system and this data is then compared to measured radar data in S- and X-band. Cepstrum analysis is used to evaluate the features of each time frame and the synthetic data is compared to measured data. It is shown that motion capture data can be used to simulate the Doppler response of human targets. It is also shown that, whereas the motion of most body parts of a human target can be observed in the X-band data, only the main torso sway can be observed in the S-band. This implies that X-band data is well suited to cepstrum based human motion classification, whereas S-band is not.

However, there are some discriminative features that could be extracted from the S-band data of the main body sway of running and walking individuals. Finally, the statistical differences between cepstrum coefficients of the data of walking and running individuals are highlighted, indicating their discriminative significance.

The work in this chapter was adapted and submitted to the 2015 radar conference and has since been published.

4.2 INTRODUCTION

Baulic walking models [9] have been used to model and classify human motion [10]. Baulic describes the human walking model in terms of three global parameters: the cycle length, cycle frequency and human speed. Although this model has a certain level of human motion personification, not all personifications of the human motion are presented in the models. Different left/right step lengths are one of the features that is not included in this model. This chapter will attempt to derive a more accurate model of human motion that can be used to simulate a wider variety of targets. The suitability of motion capture data, such as the Carnegie Mellon University Graphics Lab Motion Capture (MOCAP) database, will be investigated and this data is converted into accurate human Doppler measurements that can be used in the development of classifiers for the radar domain.

4.3 CMU GRAPHICS MOTION CAPTURE DATABASE

The CMU MOCAP database [54] is a freely available database of multiple motion captures that includes various formats. Targets have been recorded performing a wide variety of motions. Walking, running, dancing and a wide variety of arm motions have been recorded. The MOCAP data was collected using 12 Vicon infrared MX-40 cameras, each of which is recording at 120Hz with a 4 megapixel resolution. The cameras are placed in a rectangular area of approximately 3m x 8m in the center of the room. Only motion taking place within the rectangle is captured. To capture the position of the joints, the test subject wears a black jumpsuits and each joint is marked with a small grey marker the position of which can be seen in Fig. 4.1. In total, 41 markers are taped to each individual and the position of the joints are inferred from the position of these markers and used to create the skeleton files. These skeleton files are then animated and body positions can be extracted at each time step.

The actual data that is recorded is the positions of the joints over time. This means that the marker position is not recorded and instead multiple markers are used to extrapolate the position of the joint and those positions are calculated. These joint positions can then be used to determine the position of

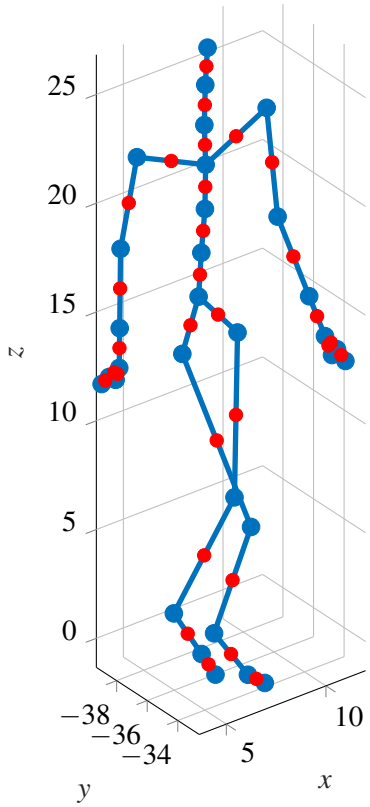


Figure 4.1. A wire-frame model of a human subject. The blue dots represent the joints as obtained from the data. The red dots illustrate the centre of the Body parts (The point from which the speed of the body part is calculated). Not all body parts are used but they could be included in the model if desired. For instance, the neck and shoulders are excluded in this model for simplicity.

the most crucial body parts at each time instance. The speed and acceleration of each body part as the person performs movement in front of the camera system can be inferred using these data points.

4.3.1 Target model

To model the targets using the CMU MOCAP database, the target is modeled as a set of ellipsoids. An ellipsoid is placed on the position of each body part, the dimensions of which are extracted from the skeleton file. Using the joint markers, each ellipsoid is placed and its center calculated for every time frame, as can be seen in Fig. 4.1. In total, 12 body parts are used as part of the target model.

For each body part, the RCS is calculated as:

$$\sqrt{\sigma_b} = \frac{R_e^2 \frac{1}{2} H_e \sqrt{pi}}{(R_e sin(\theta_e))^2 + (\frac{1}{2} H_e cos(\theta_e))^2} e^{(j2\pi/\lambda)2r}, \tag{4.1}$$

where R_e is radius of the ellipsoid, H_e is the length of the body part and θ_e is the aspect angle. Once the return for each body part has been computed, the total target return is calculated as:

$$\sigma_T = \sum_{0}^{N} \sqrt{\sigma_b} e^{j\pi(fv/c)}, \tag{4.2}$$

where N is the number of body parts being considered, v is the speed of the specific body part and f is the carrier frequency. For the head, the return is calculated as a sphere that is $H_e = R_e$. The values for H_e and R_e for other body parts are average values for a human using the same values as in [48], these values are given in Table. 4.1.

Table 4.1. Values for H_e and R_e used in simulation from [48].

Body part	H_e (m)	R_e (m)
Head	0.2	0.2
Torso	0.8	0.25
Upper leg	0.5	0.1
Lower leg	0.5	0.07
Feet	0.2	0.07
Upper arm	0.45	0.05
Lower arm	0.45	0.04

4.3.2 Data processing

MOCAP data for radar return modelling presents two mian problems. The first is the low sampling rate and the second is the high frequency noise that is present in the estimation of the joint positions. A typical radar system usually samples the data in the kilohertz range while the MOCAP cameras only sample target positions at 120 Hz. To deal with these problems, the following data processing is performed:

First a radar matching the parameters, shown in Table 4.2, of the available measured data is simulated and placed at an arbitrary point (1 km straight ahead of the target). Then, the position of the 12 body parts of interest are calculated using the position of the joints from the skeleton file. Using the

position of the body parts in the subsequent frames, the speed and angle of incidence relative to the radar are calculated. Then, for clarity and because human targets with relatively low speeds are under consideration, the data is passed through a simple low pass filter to remove the high frequency noise present in the estimation of the joint position and error from the measurement. This low pass filter is a simple 20'th order, 5 Khz low pass FIR filter using a Hamming window. The data is processed to accommodate the desired radar. The speed and angle of incidence data are interpolated by buffering the signal with zero's and the filtering with a low pass FIR filter of length 4 and with an alpha value of 0.5 in order to increase the sampling frequency to the desired amount. Finally, the radar Doppler return of the target at each time step is calculated as the sum of the individual body parts.

Table 4.2. Operating frequencies for radar used in the simulation validation.

Radar Parameters	
Operating Frequency	X-Band
Carrier Frequency	10.5 GHz
Operating Frequency	S-Band
Carrier Frequency	4 GHz

4.4 MEL-CEPSTRUM COEFFICIENT EXTRACTION

The Doppler of the radar return is treated similarly to an audio signal and audio classification techniques are used to extract features from the radar data. Historically, it has been shown that trained human operators can distinguish between different classes of targets when listening to the Doppler return of targets [55]. Using this knowledge it should be possible to leverage an audio based classification scheme for classifying targets based on their radar Doppler return. Mel-frequency cepstrum coefficients are frequently used in audio classification. They appear to produce a more efficient representation of the frequency spectra than other analysis methods such as Linear Predictive Coding (LPC) [56].

Throughout this work two terms will be used to discuss the method in which the data is processed. These concepts are more clearly illustrated in Fig. 4.2.

The first concept discussed is that of a burst. A burst of data is a segment of data that is considered stationary and is a length chosen based on the required Doppler resolution. This value should ideally

be the longest possible length of data such that the signal is considered stationary for the time period. The length of this burst will determine the maximum Doppler resolution. As such a higher value is almost always considered better, keeping in mind that the target Doppler statistics should be considered stationary for the duration of the burst. In typical radar systems this will usually limit the maximum burst length to approximately 100ms because after this time frame some smearing effects will occur due to the target no longer being stationary for an individual burst. The term "window" could also be used instead of "burst" but this causes some confusion with the idea of a windowing function. Considering that any burst of data will always be passed through a windowing function to avoid high frequency noise arising from the segmentation of the data, this confusion is exacerbated. As already stated, the term burst is used instead here because it relates to the terminology frequently used in radar. The burst length is the term used to describe the length of the physical transmitted pulse/burst. In some systems there is dead time in between bursts. Processing over multiple bursts could cause some errors to arise owing to the dead time of the system, and as such in these systems the maximum burst length of our classifier is dictated by the burst length of the radar system, although in most FMCW systems this tends to be an artificial concept because there is little to no dead time.

The second is a frame of data, a frame is a segment of data simulated to be a length equal to our desired simulated observation period (which will henceforth be referred to as ToT). This is the period for which the target was directly in the beam of the radar system. As the scanning rate of the radar is adjusted, this length changes. Typical values for this are usually between 50ms to 100ms, but if the radar scan rate is lowered enough, any value is theoretically possible. In a staring radar environment (where the antenna is not rotating) the frame length can be any desired length and is a purely artificial construct used to evaluate performance under a given hypothetical observation period.

In summary, a burst will refer to a small segment of data that is the result of FFT processing (explained below) and has been passed through a windowing function. A frame of data will refer to a series of bursts that have been observed in sequence the length of which will be referred to as the ToT.

A burst is composed of N samples which have been observed consecutively upon which a length-N FFT is performed resulting in FFT energies $S_1, \ldots, S_n$ at frequencies $f_1, \ldots, f_N$. Bursts can be allowed to overlap in some instances, and are typically chosen in such a manner as to have the signal be considered stationary for the duration. A windowing function (Hamming window in this case) is applied to the window of data to reduce side-lobes. For each burst, a set of mel-cepstrum coefficients are calculated

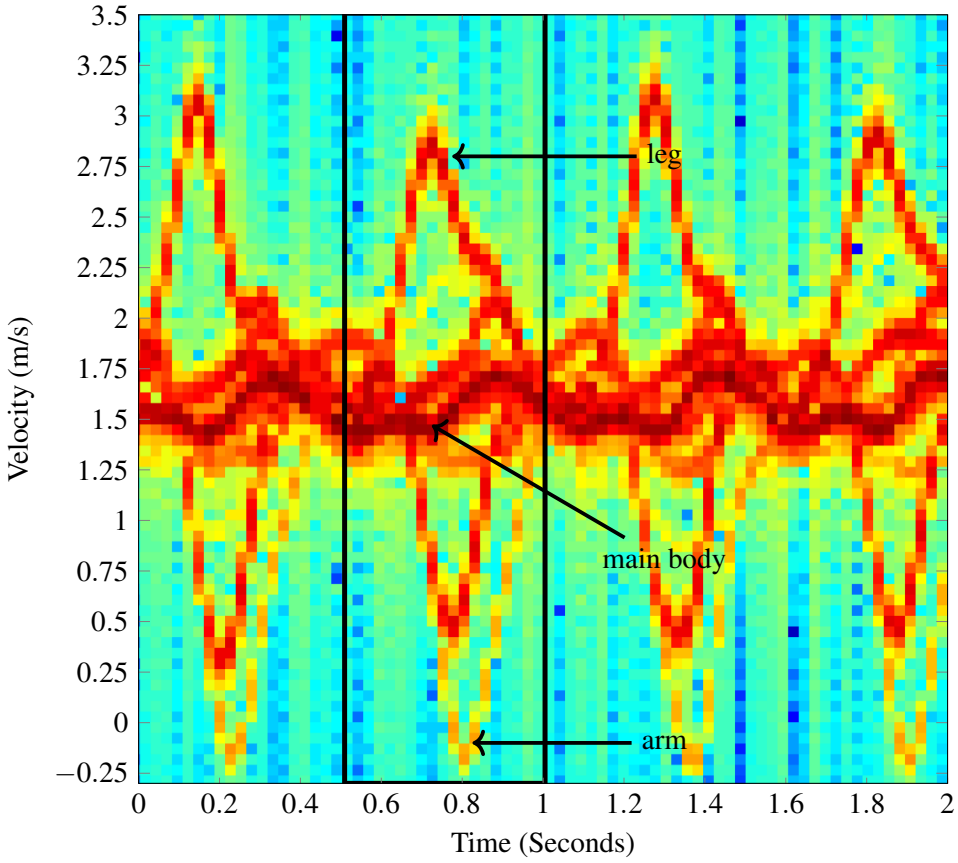


Figure 4.2. Each vertical line of the spectrogram corresponds to 1 burst of 128 samples upon which a N-point FFT (where N=128) has been computed. The rectangle indicates the boundaries of a frame. Here, a frame represents the chosen (simulated) ToT, were the radar to operate in scanning mode. The frame in turn, consists of a specified number of bursts (In this example there are 18bursts in the frame). The main body sway, as well as the movement of the legs and arms are clearly visible in the spectrogram, and are indicated with arrows in the spectrogram.

and stored. A set of consecutive bursts for a specific set of mel-cepstral frequency ranges is defined as a *frame*. A spectrogram of a simulated human illustrating these concepts is shown in Fig. 4.2.

To compute the mel-frequency cepstrum coefficients, the following steps are performed:

1. The time domain signal $s(n)$ is divided into frames of length L corresponding to the ToT one wishes to consider. The length of these frames are up to the user but will usually be dictated by the radar system in question. The ToT has been varied in different studies.
2. Next, the frames are divided into B bursts, the burst length N correspond to the burst length of

the radar system, when this is applicable, and will usually be equal to the length of the FFT N. The framed time domain signal will be referred to as $s_i(n)$, where n is the index of the nth data point in the burst and the subscript i is the index of the ith burst in the frame of length L.

3. For each burst in the frame, take the FFT, thereby obtaining the frequency domain signal $S_i(k)$:

$$S_i(k) = \sum_{n=1}^{N} s_i(n)h(n)e^{\frac{-j2\pi kn}{N}} \quad 1 \leq k \leq N, \tag{4.3}$$

where $h(n)$ is an N sample long Hamming window, n is the index over samples, and k is the Fourier domain index. The periodogram-based power spectral estimate for the burst $s_i(n)$ is given by:

$$P_i(k) = \frac{1}{N}|S_i(k)|^2. \tag{4.4}$$

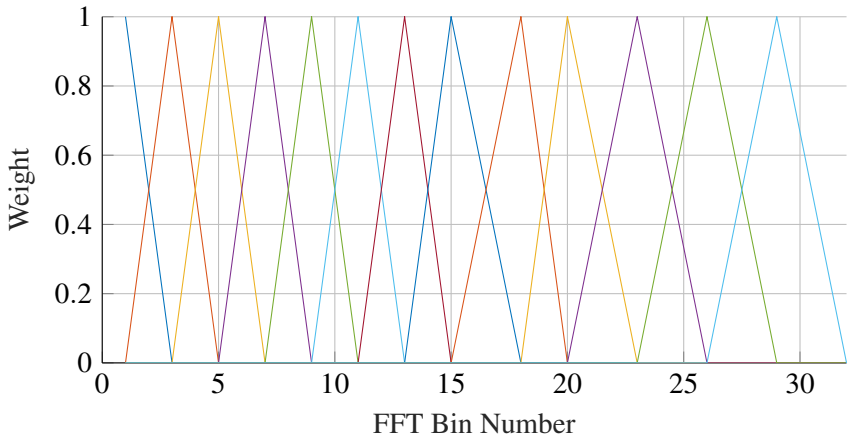


Figure 4.3. Cepstral weighting functions for 13 cepstrum with an FFT length of 64.

4. Compute the mel-spaced filterbank $H_m(k)$. This is a set of triangular filters that are applied to the periodogram power spectrum of Step 3. An illustration of the cepstral wavelet functions is shown in Fig. 4.3. These are essentially wavelets spaced according to the mel-scale. Each individual mel-cepstrum coefficient covers a certain range of the spectrum and sums all of the data in that region to give a single value corresponding to the power in that region. The filterbank will be M vectors of length N, where M is equal to the chosen number of cepstral coefficients. To calculate the filterbank energies, multiply the coefficients of each filter $H_m(k)$ with the power spectrum $P_i(k)$ of the burst and sum all the values to obtain a single value A_m, $m \in \{1,\ldots,M\}$ for each of the M filters; that is,

$$A_m = \sum_{k=0}^{N} P_i(k)H_m(k). \tag{4.5}$$

5. Take the log of each of the filterbank energies from step 4; that is,

$$Y_m = \log A_m. \tag{4.6}$$

6. A Discrete Cosine Transform (DCT) is used to determine the spectral content (e.g., signal amplitude). The initial K coefficients y_n, $n \in \{1,\ldots,K\}$ of an inverse DCT of the first K components of $Y_1, Y_2, \ldots Y_K$, where $K < N$ is given by
$$y_n = \sum_{k=1}^{M} Y_k \cos\left[n\left(k-\frac{1}{2}\right) * \frac{\pi}{20}\right]. \tag{4.7}$$
The initial value y_0 represents the average power of the signal, which can vary widely, and is usually discarded. Only the first 5–10 values are usually used in recognition depending on the spectrum of the target being observed. If a higher degree of resolution is desired the higher values can be retained.

To calculate the coefficients $H_m(k)$ of the mel-filterbanks from Step 4, the next four steps can be followed:

1. First, convert the upper and lower frequencies of the relevant frequency interval of the data to the mel scale with the equation:
$$\mathcal{M}(f) = 1125\ln(1+\frac{f}{700}). \tag{4.8}$$
where f is the frequency being converted. In this case, the lower frequency limit is 15Hz (23.85 mels) and the upper frequency limit is dependent on the sampling rate of the system and is usually equal to the Nyquist frequency. If the radar system has a Set Repetition Frequency (SRF) of 1.3kHz the upper limit will be around 650Hz (738.87 mels).
2. Next, the points are linearly spaced between the lower limit and upper limit in the mel-spectrum to obtain $M+1$ mel-frequency points μ_m. We then convert these points back to frequency domain (from the mel-domain) with the equation:
$$h(m) = \mathcal{M}^{-1}(\beta_m) = 700(e^{\frac{\beta_m}{1125}} - 1), \tag{4.9}$$
where β_m indexes a frequency on the mel-scale, to obtain m points $h(m)$ in Hz.
3. Since the frequency resolution does not allow these points to be placed anywhere in space where desired, these points can be converted to frequency bin numbers $f(m)$ based on the length of the FFT (N) as follows:
$$f(m) = \lfloor (Nh(m)/T_s) \rfloor, \tag{4.10}$$
where T_s is equal to the sample period and $\lfloor\rfloor$ is the floor of the function. This results in a sequence of bin numbers $f(m)$.
4. Now, the filters $H_m(k)$ are created. Each filter will start increasing at a point $f(m-1)$ peak at $f(m)$ and decrease back to zero at $f(m+1)$ as can be seen in Fig. 4.3 while being equal to 0 at

all other points according to the following equation:

$$H_m(k) = \begin{cases} 0 & k < f(m-1), \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m), \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) \leq k \leq f(m+1), \\ 0 & k > f(m+1). \end{cases} \tag{4.11}$$

Only $D = M - 1$ coefficients x_1 to x_M are used.

This description should make it apparent that the Mel-cepstrum coefficients can be considered as a method for simplifying the representation of the signal by summing spectral data in bands and then representing the signal based on the coefficients of a DFT that would recreate that spectral distribution. If the FFT does not contain more data points than there are cepstrum coefficients, then the cepstrums serve no purpose and will be equivalent to simply using the values within the FFT bins.

4.5 COMPARISON OF DATA

To determine if the CMU motion capture data is a good representation of the radar data, the simulation results are compared to three different radar datasets. The three corresponding radars are:

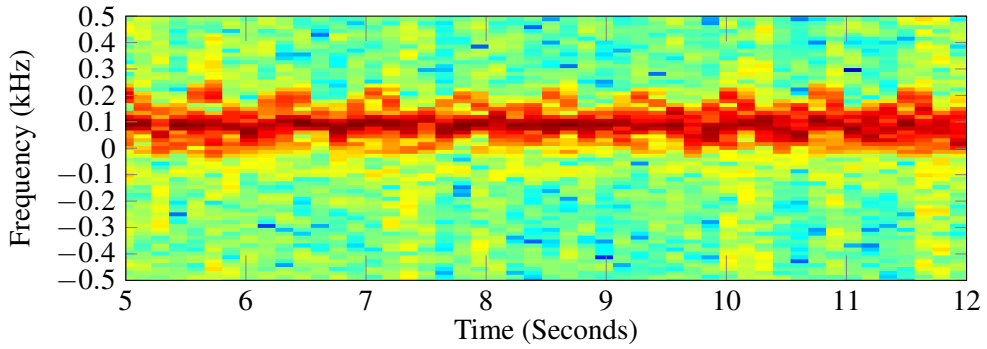
- An S-band measuring facility at the CSIR,
- A short range continuous wave (CW) S-band radar,
- An X-band radar gathered by [55].

A number of pre-processing steps have been preformed on the data. First, the X-band measured data was passed through a simple 50Hz low pass notch filter centred around 0Hz to mitigate the effect of clutter in the data. Next, the power of all the data in each time frame of the frequency axis was normalised for every time step t and bin n of the FFT

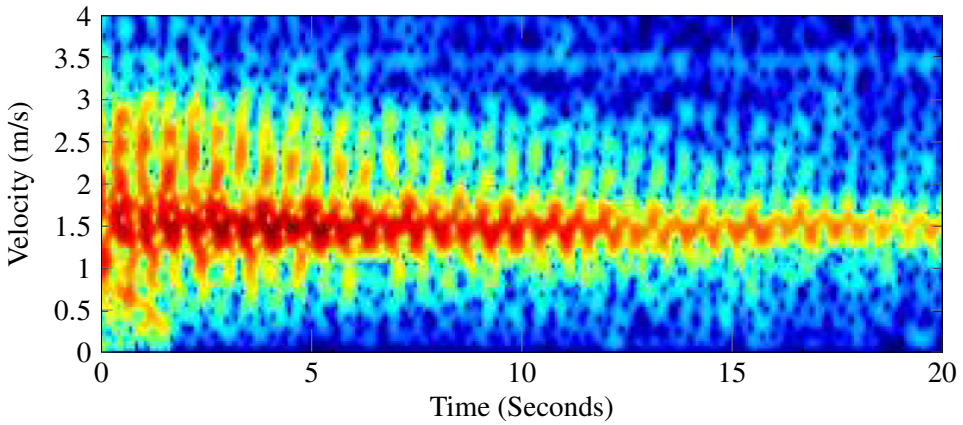
$$f_{norm}(t,n) = \frac{|f(t,n)|}{\sum_{n=0}^{N_{fft}}(f(t,n))}, \tag{4.12}$$

where f is the spectrogram of the data at time index t as generated by an FFT of length N_{fft}.

4.5.1 S-band data



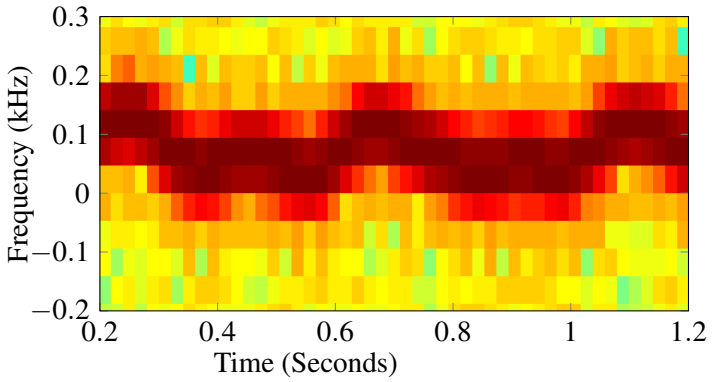
(a) Measurement of human walking approximately 200m from CSIR S-band measuring facility.



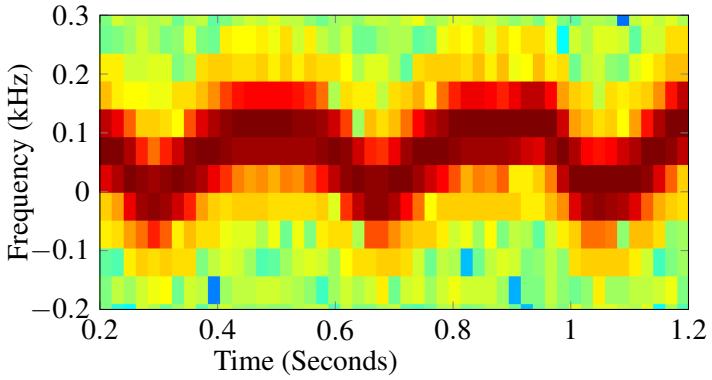
(b) Measurement of human walking away from S-band continuous wave radar from [55].

Figure 4.4. Spectrogram of measured S-band data.

The spectrogram generated from the CMU MOCAP data is shown in Fig. 4.5(a) and Fig. 4.5(b) alongside the spectrogram generated from the CSIR's S-band measuring facility Fig. 4.4(a) and the CW S-band data Fig. 4.4(b) from [55]. The time resolution of the signal from the CSIR's S-band measurement facility is relatively low; however, the simulated model could be put under the same constraints. As can be seen, at this frequency and range the micro Doppler features of the human motion are not visible, and only the torso can be observed. In Fig. 4.4(b), where the measurements have been performed at extremely close ranges, the arms and legs are only observable at extremely short ranges and they decay rapidly as the target moves away. This limits the utility of cepstrum based classification because cepstrum analysis works on the principle of placing weighted triangular shaped functions around the frequency spectrum for the micro Doppler signatures caused by the swinging of the arms and legs. In this case, such detailed coefficient extraction could not be performed owing



(a) Spectrogram of walking person extracted from MOCAP library. In this simulation all of the body parts except the torso have been removed.



(b) Spectrogram of running human simulated from MOCAP library. In this simulation all of the body parts except the torso have been removed. The spectrogram for a walking and running individual even with only the main body visible differ significantly from one another.

Figure 4.5. Spectrogram of simulated S-band data.

to limited Doppler resolution (i.e. the entire frequency extent of the main body falls only within five Doppler cells).

For the spectrograms generated from the CMU MOCAP data, the targets are simulated moving towards the radar and the simulation is put under the same constraints. Since the arms and legs of the target can be observed at all stages in the simulation, they were removed from the S-Band simulation to match the spectrogram to that observed in the measurements.

There are still some observable differences between the walking and running spectrograms that could

be used as features as seen in Fig. 4.4, The walking human has a slightly smaller frequency span (around 50Hz) than the running human. The running human's main body spends more time on average at a speed higher than the average speed of the target, while the walking human exhibits the opposite with the main body spending a higher percentage of its time slower than the average speed of the target. Furthermore, the stride length can still be extracted from these features by finding the period of the sinusoidal waveform. Using this waveform, it should be possible to extract the main body velocity and frequency span of the main body sway. The extraction and evaluation of these features is not investigated in this work but has been researched [57].

Although the S-band results show that spectrogram based analysis is not really a viable option, it can still be seen that simulating data using the CMU-MOCAP library will give good results to use in the generation of additional features and investigation of human motion. The most significant difference between the measured data and simulated data is the result of the filtering process to remove the zero-Doppler line and the surrounding clutter. In the simulated data, the main body sway is much more sinusoidal while the measured data from the CSIR S-band measuring facility has cut off the bottom part of the sinusoid owing to the higher clutter and the notch filter designed to remove the clutter. In Fig. 4.4(b) (the S-band continuous wave data) it is evident the main body sway is much more sinusoidal (as expected) owing to the lower level of clutter in the measurement.

4.5.2 X-band data

A number of preprocessing steps have been performed for the measured X-band data. The most significant step is the application of a 10Hz notch filter centred around 0Hz to remove the clutter and normalisation of the Doppler line for each time frame.

4.5.2.1 Visual inspection

The simulation is adjusted by changing the frequency of the simulation and including the additional body parts Fig. 4.6. It is then compared to high range resolution X-band radar data (Fig. 4.7). It is evident that the arms and legs are prominently visible and that the target motion looks as expected. It is important to note that most of the negative Doppler and low frequency oscillation of the arms are lost as a result of notch filters designed to remove the low frequency clutter. The presence of the

negative Doppler within a human spectrogram could be a useful feature because it is believed to be only bipedal walking that exhibits negative Doppler when the arms counteract the forward motion of the legs to preserve balance. Although this is prominently visible within the radar simulation of the MOCAP data (which does not include any clutter or filtering), the back-swing of the arms is lost. In the X-band data, the arms and legs are still visible and it is reasonable to expect them to remain visible at much greater ranges.

This data allow for the application of cepstrum based features. When the bursts of data are clustered based on the cepstrum coefficients, it is evident that each burst is easily assigned to at least one of three possible classes (states). Differentiating between more states within the walking model is possible but would require more cepstra to be placed within the space along with higher frequency and time resolution to avoid having the cepstrum be only a single point within the FFT.

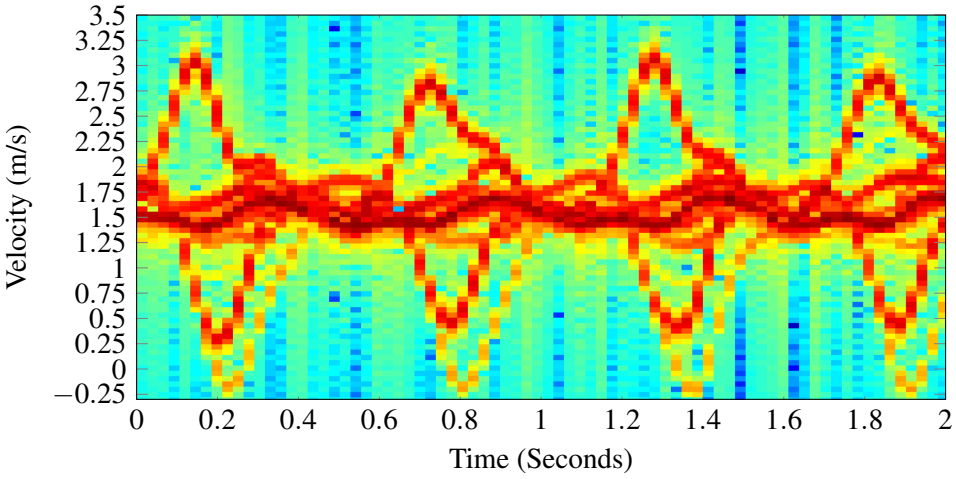


Figure 4.6. Spectrogram of human walking derived from Mocap Data at X-band.

4.5.2.2 Cepstral feature analysis

X-band data cepstral features were extracted. These features were then statistically compared to give an indication of how well the MOCAP spectrogram matches the real data, as well as how these features can assist in the discrimination of types of human motion. To extract cepstral coefficients, ten triangular functions were linearly spaced across the positive frequency extent of the spectrogram. It was observed here and in Chapter 5 that very little data is observed in the opposite spectrum of the direction of movement. Although from the simulation it is evident that negative Doppler could be observed, the filtering to remove the zero-Doppler return and the worst of the clutter filters out most if not all of the

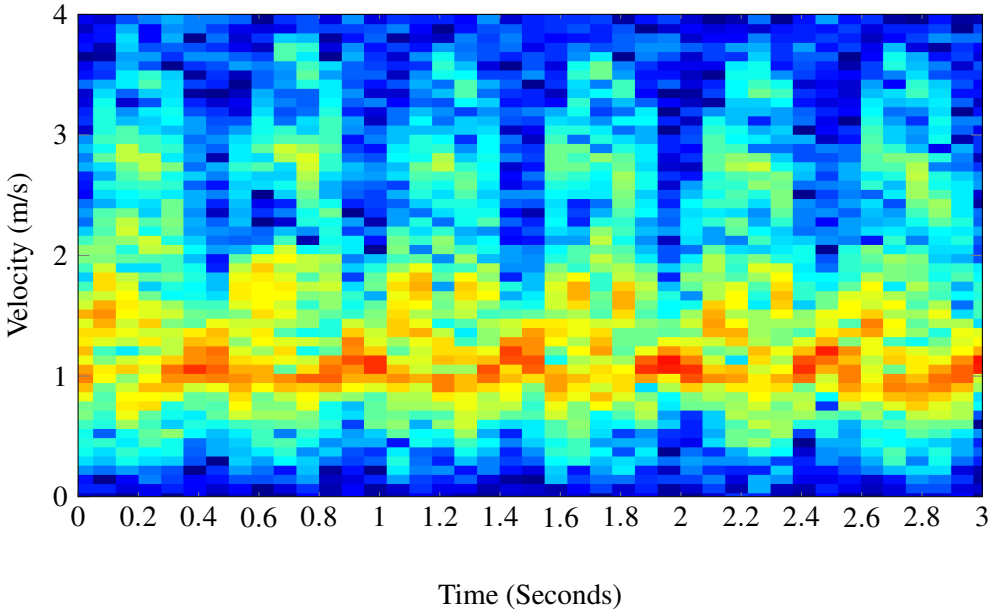


Figure 4.7. Measured spectrogram of human walking away from X-band radar from [55].

observable micro-Doppler. Consequently, it was chosen to only extract cepstrum for one side of the spectrogram. This is also the case for the data collected in Chapter 5. Few targets were observed to exhibit both positive and negative Doppler.

Once the cepstral coefficients are extracted from each time frame, the data is clustered into one of three states using K-means clustering based on the Euclidean distance between the cepstrum coefficients. The data is clustered into only three states to investigate if the data clusters in a neat periodic fashion as expected. Once the final classifier is designed, the amount of states is allowed to vary based on class. Methods such as Akaike's information criterion (AIC) [58] could be used instead (as demonstrated in Chapter 5) to decide on the optimal amount of "states". Fig. 4.8 indicates the classes to which each bursts has been assigned over time. It is evident that the cepstral features allow a distinction to be made between the different phases of the walking process. Classification of motion type can be performed by comparing the transitions of extracted cepstrum coefficient classes of walking human data to those of the training data.

The mean and variance of each cepstrum is show in Table 4.3 and 4.4.

Although the process works for both the simulated data and the measured data, it can be seen that

Table 4.3. Mean and variances of clustered mel-cepstrum coefficients for simulated data.

Mean and variance of each cluster of simulated data						
	class 1		class 2		class 3	
Num	Mean	Var	Mean	Var	Mean	Var
1	-5.28	0.96	-6.76	1.19	-8.42	1.92
2	4.01	0.063	3.44	0.38	1.98	2.39
3	1.78	1.5	3.25	0.45	2.29	0.27
4	-1.78	0.51	0.73	0.37	1.04	0.76
5	-0.46	0.63	0.61	0.18	1.22	0.08
6	-0.32	1.16	-0.72	0.22	0.59	0.2
7	0.88	0.15	-0.38	0.06	0.53	0.03
8	0.35	0.08	-0.50	0.03	0.19	0.02

Table 4.4. Mean and variances of of clustered mel-cepstrum coefficients for measured data.

Mean and variance of each cluster of measured data						
	class 1		class 2		class 3	
Num	Mean	Var	Mean	Var	Mean	Var
1	-16.3	0.07	-15.9	0.09	-15.05	0.08
2	-0.24	0.02	0.48	0.14	0.24	0.22
3	-3.27	0.06	-3.35	0.08	-2.64	0.1
4	0.22	0.05	-0.20	0.12	-0.08	0.05
5	1.96	0.12	0.81	0.08	0.44	0.12
6	-0.39	0.03	-0.77	0.04	-0.63	0.09
7	-0.47	0.15	-0.01	0.08	-0.21	0.07
8	0.03	0.04	-0.06	0.05	0.17	0.05

the actual values (mean and variance) of the cepstrum coefficients differ significantly between the simulated radar and real radar data. This is due to the ideal ellipsoids used within the simulation model, which caused a difference between the ratios of RCSs of body parts over time. A second observation is that the spectral power within the first used cepstrum is significantly lower and provides almost no distinction in the measured data because of the notch filter designed to remove the clutter, thereby

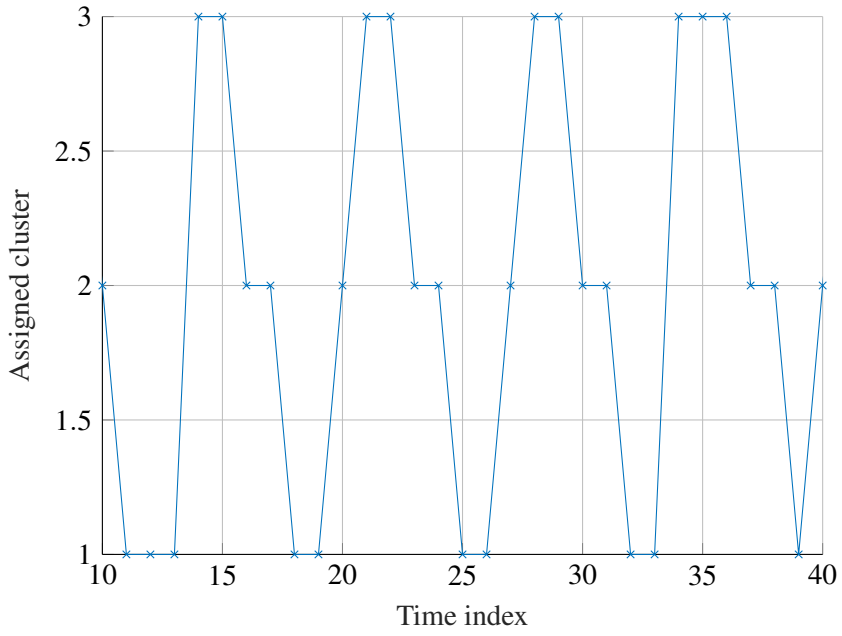


Figure 4.8. State transition over time for measured X-band data. From the state transition of the clustered data, we can see that the clusters represent different phases within the human walk cycle and that the cepstrum features should allow for classification of human motion.

suppressing all low frequency human motion information. The results of the analysis of the cepstral features makes it evident that training a radar on simulated data when no training data is available could have severely negative effects because the cepstrum features differ so significantly.

4.6 CONCLUSION

The results of the study indicate that MOCAP data can be effectively used to simulate human motion at different bandwidths. However, it is very important that the simulation be validated on measured data to ensure that the simulated data is similar in terms of the features extracted. In these tests, it was evident that a simple system such as this could not be used to supplement training data. This study mainly considered S-band and X-band data and it can be seen that for any long range application a cepstrum based analysis of human walking is not viable at S-band. It can also be seen that it is important to follow the same processing procedure on the simulation data as on the real data. However, clutter needs to be handled in the real data through the use of a 0 Hz notch filter and this effect would need to be included within any simulation. To this effect, if MOCAP data is to be used effectively to supplement classification data, then a more complex simulation of the model is required which

includes the clutter and calibration that needs to be preformed to accurately reflect the RCS of the body parts at different ranges. If, however, the goal is to only test different classification techniques, then using only MOCAP data will give reliable, reproducible results. However, if the intention is to perform classification with a practical radar system, then a cepstrum based classifier would need to be trained on real data, unless significant care is taken to ensure that the simulated data matches the measured data closely in all regards.

A key point to remember during the processing of the data is the normalisation of the power of the Doppler data in every time frame. It can be seen that cepstral analysis of the data focuses on the ratio's between different cepstrum coefficients. If care is not taken to ensure that the total power within each frame is normalised, then the magnitude of the cepstral coefficients could vary widely between frames, which would result in faulty classification.

Given the periodic nature of the simulated data, it is reasonable to assume that a HMM based classifier could add a further degree of accuracy to existing classification models. Chapter 5 will test this hypothesis on recorded data.

CHAPTER 5 CLASSIFICATION OF RADAR DOPPLER TARGETS

5.1 CHAPTER OVERVIEW

In this chapter, a classification is performed on measured radar data obtained by the CSIR. The data was collected in the Kruger national parks using an X-band radar that was staring over the grasslands and observing animals and humans walking towards the radar. The human targets were the rangers and researchers walking in a semi-controlled manner while the animals were targets of opportunity that happened to be walking towards the radar. A combined GMM-HMM model is developed to distinguish between slow moving animal and human targets using mel-cepstrum coefficients. For comparison, Bilik's method [6] is also implemented to establish a baseline for comparison with a GMM model. A database of slow moving Doppler targets in a cluttered environment was recorded and used to evaluate the performance of the models. It is shown that this approach can accurately distinguish between different classes of animals and humans walking in cluttered environments. Although the dataset has limitations, especially in the limited quantity and the variability within the data, the results are promising. The accuracy of the model is found to be dependent upon continuous observation time on target and this accuracy can range from 65% to approximately 90% accurate classification for times on target between 250ms to 1.25s, respectively. Furthermore, the effects of a confidence score and its ability to reduce false alarms will also be evaluated.

5.2 INTRODUCTION

Due to the severe problem of rhino poaching, the CSIR's DPSS research group was approached to help develop an anti-poaching radar. In the initial testing of this system, staring data was collected by

staring over an undisclosed region of the Kruger National Park from a hilltop. This X-band radar was actually designed for another application but it was re-purposed for detecting and tracking humans and animals. As such, the design is not ideal for this application. However, it was able to function sufficiently to detect slow moving ground targets, such as poachers walking across the border. The data described and used in this section was recorded as part of an initial test of the system. The radar was operated in its staring mode and targets of opportunity were recorded, along with a researcher that was walking with a guide towards the radar to determine the detectable ranges for the system. Most of the recordings that were used as part of the classification dataset were recorded at ranges from 1–3 km.

5.3 CLASSIFICATION METHOD OVERVIEW

This section will describe how the two classifiers function on a general level. The GMM-HMM model was created by the author of this dissertation, similar methods have been used in audio classification but have yet to be tested on radar classification problems, the GMM method (Bilik's method) is used as a baseline for comparison. Subsequent sections will describe the different methods in more specific detail. The images used to describe the processes in this section are generalisations that serve as an example. The actual data clusters into many more than three components if a model order criterion such the Akaike Information Criterion (AIC) is used. However, the exact length of the bursts may not be consistent with that shown in the figures here.

Regardless of the classification method used, the recorded data is divided into three equal sized datasets and the data processing described in Section 5.4 is applied. Mel-cepstrum coefficients are extracted from the FFT spectrum of each burst and of each frame, and they are then stored in a file while keeping the order of the observations intact. This process is visually described in Fig. 5.1. The mel-cepstra are then clustered separately for each class using a GMM. Each spectral line of the spectrogram can be assigned to a state or cluster when observing the spectrograms of different classes. A "state" is a region of mel-cepstrum feature space, where each spectral line which belongs to that state is a point in that region. This clustering concept is shown in Fig. 5.2. Once the mel-cepstra have been clustered, it is treated in a different manner based on the classification method being used.

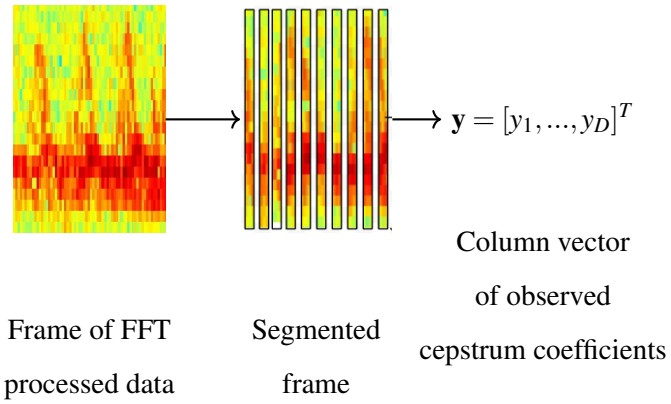


Figure 5.1. A visual illustration of the Segmentation (dividing the frame into bursts). Each frame of of FFT processed data is segmented into smaller indivisible spectral lines. For each of these segments the mel-cepstrum coefficients are extracted and stored. Mel-cepstrum coefficients can be thought of as a compact representations of the frequency domain.

5.3.1 Bilik's method

Bilik's method [6] is a method from the literature that performs exceptionally well and is frequently used for comparison with new methods [8]. Bilik's method is a GMM based all versus all classification method. For each class, a GMM model is trained on the mel-cepstrum coefficients.

Bilik's method is a two stage voting procedure. First, the cepstrum coefficients are calculated for each burst and then for each burst a likelihood ratio test (LRT) is performed between the different classes. The LRT for each class is then compared in a one versus one "competition" to a trained separation boundary γ. A class is classified as either belonging to Class 1 or Class 2, and then Class 2 is compared to Class 3 and Class 1 is compared to Class 3 individually. The class that receives the most votes is classified as the class for that specific burst. This concept is illustrated in Fig. 5.3. The separation boundary is trained by preceptron learning to deal with the difference in the amount of data available for each class. If the classes have exactly the same amount of data, then this boundary should theoretically be equal to 1.

Once the individual bursts have been classified, the frame is classified based on the votes that each class has received in a burst-by-burst basis. This concept is shown in Fig. 5.4.

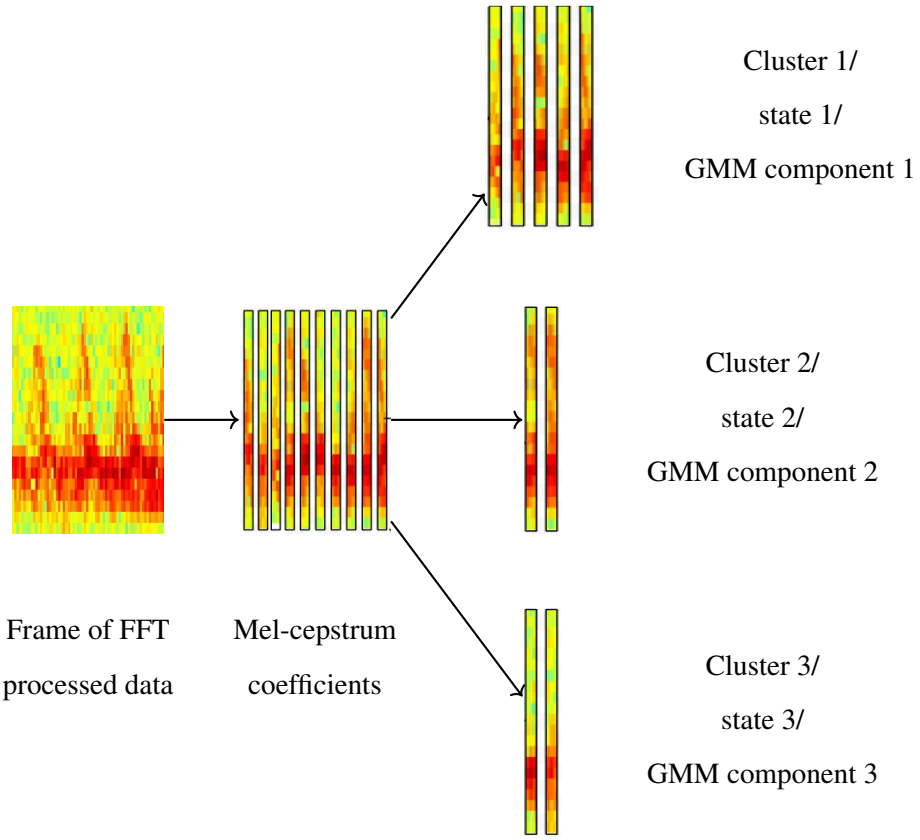


Figure 5.2. A visual illustration of the clustering. Each frame of of FFT processed data (or observation) is segmented into smaller segments of data. Visually, it can be seen that as time progresses, the signal will repeat in a periodic manner and certain bursts will have very similar frequency representations. It is expected that these representations will cluster together logically.

5.3.2 GMM-HMM design

A previous study [55] has shown that trained human operators can distinguish between different classes of targets when listening to the Doppler frequency characteristics of the return from dismounts. Using this knowledge, it should be possible to develop an automatic scheme for classifying targets based on the audio-like characteristics of their radar micro-Doppler return. In the GMM part of the GMM-HMM model, mel-frequency cepstrum coefficients are used to distinguish between the different phases or "states" of the Doppler spectrogram. Mel-frequency cepstrum coefficients are frequently used in audio classification. They appear to produce an efficient scale that is more representative of human audio perception representation of the frequency spectra than other methods, such as Linear Predictive Coding (LPC) [56]. The HMM part of the GMM-HMM model is used in speech recognition

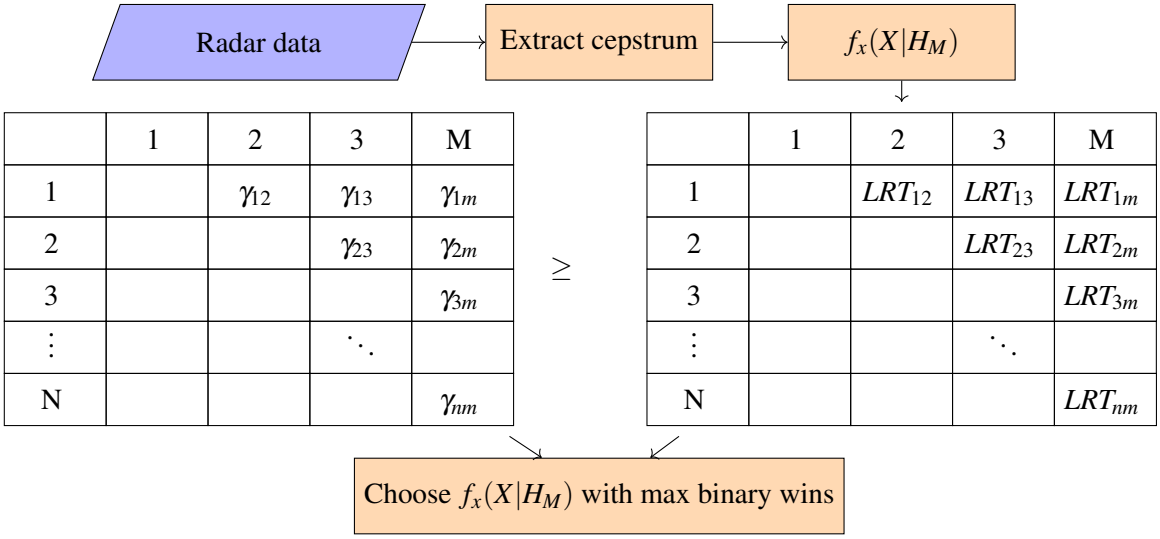


Figure 5.3. A visual illustration of Bilik's method voting scheme adapted from [6]. For each burst, a likelihood ratio test is performed between the different classes. This likelihood ratio is then compared to a trained γ value for the binary classifier and the class that wins the most of these binary competitions is chosen as the class for that window. For example, if LRT_{12} is smaller than γ_{12}, then Class 2 is chosen as the "winner"; and if LRT_{12} is larger than γ_{12}, then Class 1 is chosen as the "winner".

to model and recognise transitions between formants in speech and it has been shown to produce satisfactory performance [59].

The GMM-HMM method is designed around the observed periodicity of the animal movement cycles. Similar to the observation in audio classifiers. HMMs have been shown to be good classifiers for repeated sequences, such as audio signals. In Chapter 4, it was observed that the mel-cepstrum coefficients repeated in a periodic manner for the simulated data and this phenomenon was expected to repeat itself in measured data. The cross correlation for two random windows of giraffe data is shown in Fig. 5.5. The cross correlation shows that there is approximately 60–70% of correlation between random observations and indicates that the assumption of periodicity is acceptable. As such, a combined GMM-HMM method was designed for classification. This method was envisioned to combine the good performance of existing GMM based system while also considering the time varying cyclic information that is present in the movement of the target.

Once the GMMs have been trained for each class, they are re-combined into a single GMM, merging

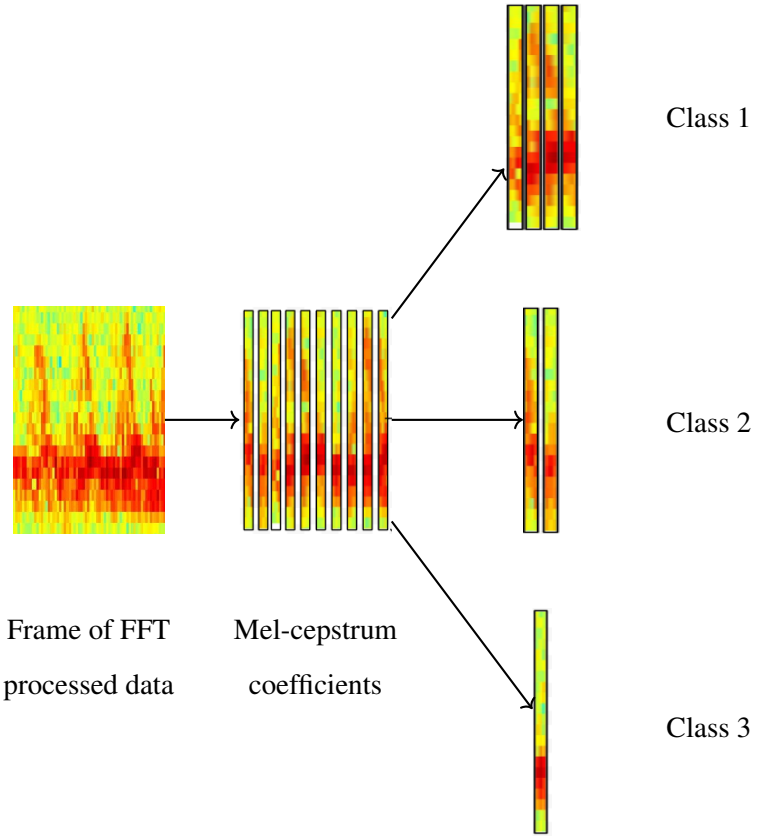


Figure 5.4. A visual illustration of Bilik's method. In this example, four bursts are classified as Class 1, two as Class 2 and one as Class 3, which results in the frame being classified as Class 1.

the components of each GMM based on their Kullback-Leibler divergence (KL-divergence) [60]. If two components had a KL-divergence of less than 1, then those components are merged into a single Gaussian, while any other components are left intact and rescaled according to their relative sizes in the GMM. Although this was never required in the measured dataset, it would be required if any of the individual components overlapped in space.

Once the GMM models are merged, each component of the new GMM is treated as a "state" and a HMM is trained on the state transitions that occur. To determine the class, the relative probability of each HMM is calculated based on the sequence of observations and the class with the highest probability is chosen as the class for the frame of data. The state transitions are illustrated in Fig. 5.6. The cross correlation of the state transitions for measured Giraffe data is shown in Fig. 5.5. As expected, it can be seen that the transitions occur in a cyclic manner, however the cross correlation is much less than expected. For the actual training of the model, the HMM model is fully connected and any transition that is not observed is set to a small non-zero number (1×10^{-4}).

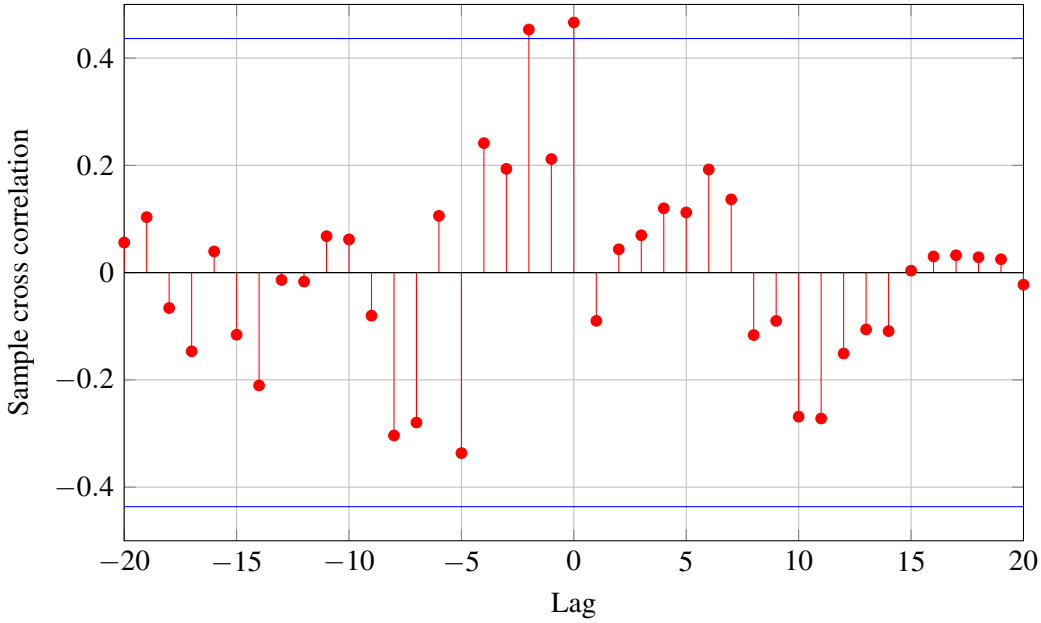
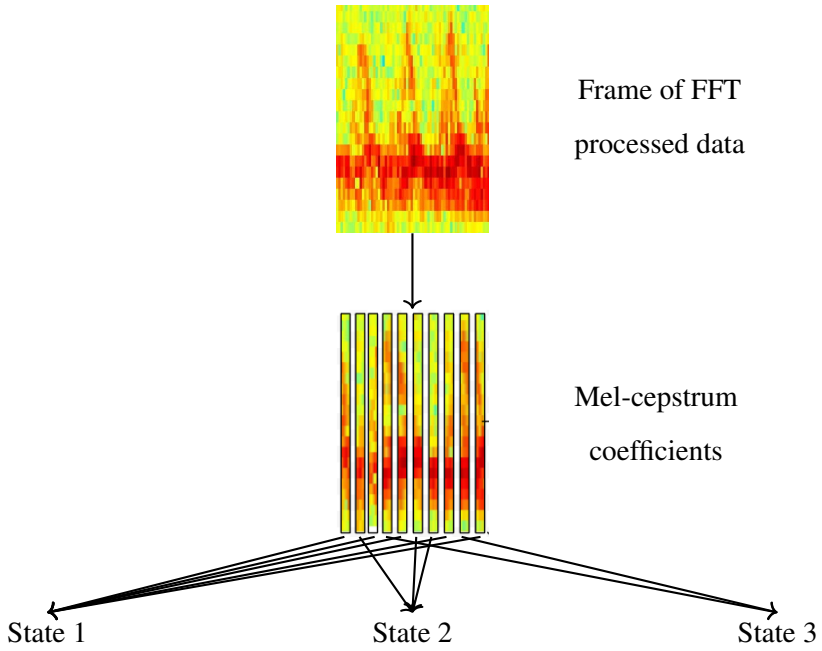


Figure 5.5. Cross correlation for two random 2 second frames of Giraffe data, negative correlation of the frames can indicate periods for which the data is out of phase.

5.4 DATA DESCRIPTION

As far as the author is aware, no large database of animals and humans walking in a woody savannah environment has ever been recorded with an FMCW radar. In an effort to develop robust classifiers for these radar systems, a data collection effort has been conducted over the course of the last year. Human and animal targets have been recorded in the wild using a staring FMCW Doppler radar. The human targets were willing participants, whereas the animals were targets of opportunity that happened to be moving towards or away from the radar site at the time. For this reason, most animal recordings are not controlled in any manner, their angle of incidence is not necessarily walking exactly towards the radar nor are their exact speeds consistent, which causes some variance within the datasets. Nonetheless, different animals could be accurately distinguished from one another.

The data was collected from a hilltop using an X-band radar system staring over savannah regions with grazing and moving animals. Most recordings were performed on animals between 1 and 3km from the radar site, the radar parameters that can be shared are given in Table 5.1.



S1 ⟶ S2 ⟶ S1 ⟶ S3 ⟶ S1 ⟶ S2 ⟶ S2 ⟶ S1 ⟶ S3 ⟶ S1

Figure 5.6. A visual illustration of the state transition after clustering. It can be seen that the transitions occur in a cyclic manner.

Table 5.1. Radar Measurement Parameters

Parameter	**Value**
Sweep Repetition Frequency	1.33 kHz
Range Resolution	$\approx$ 5m
Transmit Frequency	10.2 GHz
Beam Width	0.6°

5.4.1 Class partitioning

In this data, three different classes are considered. Walking humans, walking animals and grazing animals. Grazing animals are defined in this context as animals that are standing still, eating grass, raising their head and looking around. These types of movements are frequently detected by the radar and need to be distinguished from clutter because they are close to the zero-Doppler line and will slowly migrate from cell to cell. Each class consists of roughly 6–10 minutes worth of recorded data,

while using the radar in staring mode.

This study has two parts. The first part involves broad class classification in which the data was partitioned into walking humans, walking animals (zebra, giraffe) and grazing animals (zebra, giraffe and rhino). The second part involves a binary classification between two species, namely zebra and giraffe.

In the dataset, the targets present with either negative or positive Doppler, not both. Since the cepstral coefficients are only defined for positive spectra, the spectra of targets moving away from the radar (with negative Doppler spectra) were flipped such that every negative Doppler frequency presented as a positive Doppler frequency and *vice versa*. The implicit intuitive assumption for the sake of simplicity, although not probably accurate, is that the Doppler spectrum of a target moving away from the radar with some aspect angle would present as the mirror image of the Doppler spectrum of a target moving towards the radar with the opposite aspect angle. As such, the radial direction of the target is not assumed to contain discriminative features for the purpose of classifying between species. Future studies with more representative data may be employed to test the assumption that closing and receding targets would exhibit mirror image spectra.

A fourth "nondescript" class is added to the GMM-HMM classifier, which represents other classes that are not trained for in the model. This class is analogous to a uniform distribution over motion types and in our model is represented by an equi-probable random walk through the hidden Markov model states (i.e. all state transitions are equally probable). Primarily, it serves as a method to normalise the classification results and lower the model's confidence in its own answer, owing to the fact that not all types of target motion are accounted for in our model. In practice, this class has almost no effect on the classification outcomes, except for reducing confidence in the final answer if the exhibited motion is close to random walk behavior. Practically, it serves to distinguish unknown and yet unobserved motions that could be exhibited by animals, humans and other known motion. This model can be combined with a required confidence score before classification, which allows for false alarm rates to be significantly reduced. A similar approach was suggested in [20].

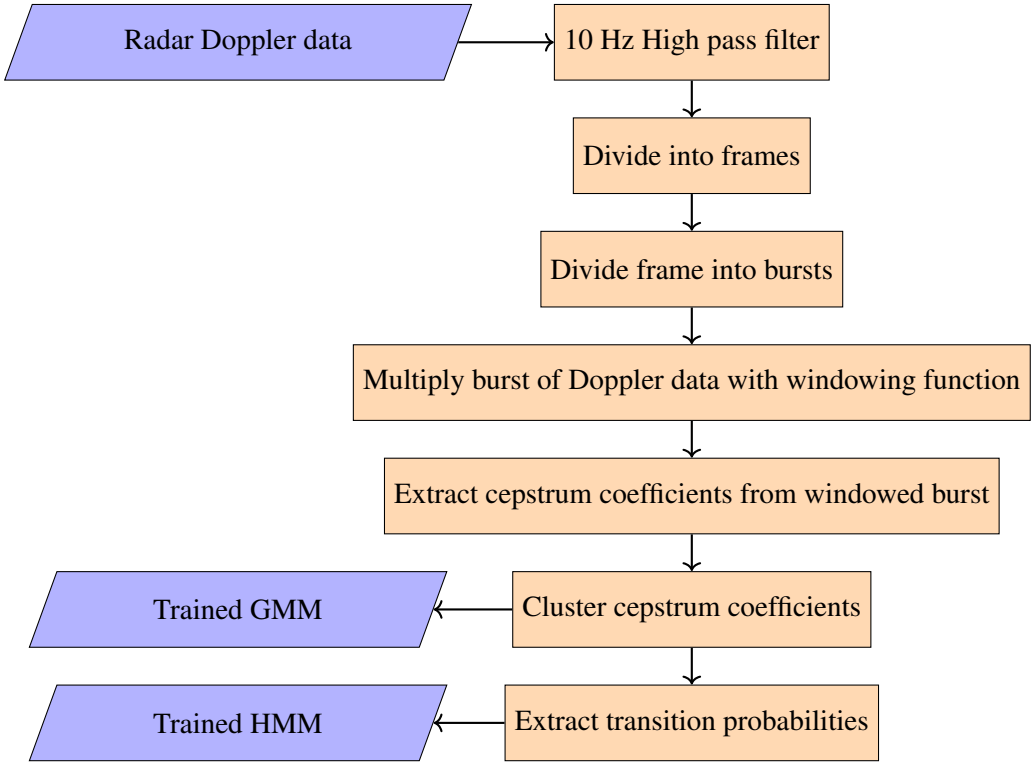


Figure 5.7. Flowchart of data processing

5.5 CLASSIFIER ALGORITHM DEVELOPMENT

The processing steps that have been performed on the data are shown in Fig. 5.7. These steps are followed when training the classifiers on any given dataset. First, targets are extracted from the radar Doppler return. The targets are then filtered to remove the ground clutter. After this, every target is divided into "Observation" frames and the frames are subsequently divided into bursts (equal to the FFT length). The cepstrum coefficients for these bursts are then extracted and clustered for each class individually using a GMM. Once the GMM models have been extracted, the GMM model components are compared using Kullback-Leibler divergence (KL-divergence) [60] and all of the separate class GMMs are merged into a single GMM for the GMM-HMM approach. In none of our tests was it found necessary to merge individual components because their KL-divergence showed that this was unnecessary and the individual components did not overlap significantly. Finally, transition probabilities are extracted for use in the HMM part of the classification. For the GMM approach of Bilik [6], the data processing stops at having obtained a GMM for each individual class.

5.5.1 Mel-cepstrum coefficient extraction

The radar return is treated similarly to an audio signal and audio classification techniques are used to extract features from the data. In a previous study [55], it was shown that trained human operators can distinguish between different classes of targets when listening to the Doppler frequency characteristics of the return from dismounts. Using this knowledge, it should be possible to develop an automatic scheme for classifying targets based on the audio-like characteristics of their radar Doppler return. Mel-frequency cepstrum coefficients are frequently used in audio classification.

Before any processing is performed to extract features, a simple 10Hz high pass filter was applied to the recorded data to remove the quasi-stationary ground clutter. The rest of the mel-cepstrum extraction process is identical to that of Chapter 4.

5.5.2 Gaussian mixture model

Once the mel-cepstrum coefficients have been determined, they are separately clustered for each class by a Gaussian Mixture Model (GMM) using the iterative expectation-maximisation (EM) algorithm [26]. The vector of observed cepstrum coefficients is given by the column vector $\mathbf{y} = [y_1, ..., y_D]^T$. A set of these vectors are then modelled as a mixture of Gaussian distributions. Consider a $D \times B$ dimensional matrix $\mathbf{Y}$, where the nth column is populated with the elements of column vector $\mathbf{y_n}$, where D is equal to the number of cepstrums per burst and B is equal to the number of bursts. Similarly, the corresponding latent variables will be denoted by an $K \times B$ matrix $\mathbf{Z}$ with rows $\mathbf{z}_n$. Assuming that the data points are drawn independently from the distribution, then the log likelihood of the Gaussian mixture model for this dataset is given by

$$\ln p(\mathbf{Y}|\kappa, \mu, \Sigma) = \sum_{n=1}^{B} \ln \left\{ \sum_{k=1}^{K} \kappa_k \mathcal{N}(\mathbf{y_n}|\mu_k, \Sigma_k) \right\}, \tag{5.1}$$

where $\mathcal{N}(\cdot|\mu, \Sigma)$ represents a multivariate Gaussian distribution with mean vector μ and covariance matrix Σ. The likelihood function in (5.1) needs to be maximised with respect to the parameters $\{\kappa_k, \mu_k, \Sigma_k\}$ for $k = 1, \ldots, N$. Dempster's EM algorithm is used to maximise this likelihood [61]. The first step of the EM algorithm is to initialise the means μ_k, covariances Σ_k and mixing coefficients κ_k and evaluate the initial value of the log likelihood. Then, perform the **EStep** by evaluating the weights

using the current parameter values

$$\gamma(z_{nk}) = \frac{\kappa_k \mathcal{N}(\mathbf{y}_n|\mu_k,\Sigma_k)}{\Sigma_j \kappa_j \mathcal{N}(\mathbf{y}_n|\mu_j,\Sigma_j)}. \tag{5.2}$$

Next, perform the **MStep** by re-estimating the parameters using the current weights

$$\mu_k^{new} = \frac{1}{B_k}\sum_{n=1}^{B}\gamma(z_{nk})\mathbf{y}_n, \tag{5.3}$$

$$\Sigma_k^{new} = \frac{1}{B_k}\sum_{n=1}^{B}(\mathbf{y}_n - \mu_l^{new})(\mathbf{y}_n - \mu_l^{new})^T, \tag{5.4}$$

$$\kappa_k^{new} = \frac{N_k}{B}, \tag{5.5}$$

where

$$B_k = \sum_{n=1}^{N}\gamma(z_{nk}). \tag{5.6}$$

Finally, the log likelihood and parameters are checked for convergence and if the values have not changed by more than 1e-6, then the training process is terminated. If convergence is not satisfied, then return to the **EStep** (5.2) and continue until the algorithm converges.

The number of components of the GMM for each class of animal or human class are automatically computed using the Akaike's Information Criterion (AIC) [58]. A number of models are trained with different numbers of components for each model. The model with the lowest AIC score determines the GMM model order. For the training dataset that is used in this study, the GMM typically has $Q = 6-10$ components dependent on the class (10 for humans and walking animals, 6 for grazing animals). This can obviously vary between datasets but is unlikely to vary much unless substantial differences are present in the structure of the cepstrum coefficients. Finally, the states (GMM components) of the classes are combined into a single GMM to allow the probabilities of the different components to be related. The separate GMMs for each of the classes are merged into a single augmented GMM, in which all the components of the separate class dependent GMMs are merged in a single "augmented" space. To determine whether or not any components significantly overlap and should be merged, the closed form equation for a bi-directional version of the Kullback-Leibler (KL) divergence between the Gaussian components is used as a similarity score between components [60]. The equation for the bi-directional KL divergence is given by:

$$\begin{aligned} D_{KL}(\mathcal{N}_a||\mathcal{N}_b) = {} & 0.25(tr(\Sigma_b^{-1}\Sigma_a) + (\mu_b-\mu_a)^\top\Sigma_b^{-1}(\mu_b-\mu_a) - k + \ln\frac{\det\Sigma_b}{\det\Sigma_a}) \\ & + 0.25(tr(\Sigma_a^{-1}\Sigma_b) + (\mu_a-\mu_b)^\top\Sigma_a^{-1}(\mu_a-\mu_b) - k + \ln\frac{\det\Sigma_a}{\det\Sigma_b}), \end{aligned} \tag{5.7}$$

where $\mathcal{N}_a$ and $\mathcal{N}_b$ are the two components of the Gaussian mixture models being compared. Σ_a and Σ_b are the covariance matrices of the Gaussian mixture components. μ_a and μ_b are the means of the Gaussian mixture components. In this bi-directional version, the value is not dependent on the order of $\mathcal{N}_a$ and $\mathcal{N}_b$, as is the case with the standard KL-divergence. In this application, it was found that no components overlap when using the KL divergence criterion. This implies that the states of the different animal and human classes are well separated in the cepstrum feature space. If any were to overlap, a merging strategy would have to be followed. A reasonable merging strategy would be to simply add all of the points of the two components together and fit a single Gaussian over these data points. Next, a HMM is constructed to classify the targets based on their state transitions.

5.5.3 Hidden Markov model

The HMM determines the probability of a class C given a sequence of observed cepstrum coefficients $\mathbf{y}_{0:t}$, where the subscript $0:t$ represents all of the observations from discrete time 0 to t. The corresponding sequence of hidden states is denoted by $x_{0:t}$. At any time instant t, x_t is a scalar that can take on an integer value between 1 and Q, where Q is the number of mixture components of the GMM. The probabilities of these states are inferred by the expectation maximisation algorithm, which has already been described. This is the standard state space description of the HMM. Using Bayes theorem

$$P(C,x_{0:t}|\mathbf{y}_{0:t}) = \frac{P(\mathbf{y}_{0:t}|C,x_{0:t})P(C,x_{0:t})}{P(\mathbf{y}_{0:t})}. \tag{5.8}$$

The prior probability in the numerator can be expanded using the product rule as follows

$$P(C,x_{0:t}) = P(C)P(x_{0:t}|C). \tag{5.9}$$

The first order Markov assumption present in the model results in

$$P(x_{0:t}|C) = P(x_0|C)P(x_1|x_0,C)P(x_2|x_1,C)\ldots P(x_t|x_{t-1},C), \tag{5.10}$$

or

$$P(\mathbf{x}_{0:t}|C) = P(x_0|C)\prod_{l=1}^{t} P(x_l|x_{l-1},C). \tag{5.11}$$

Using the conditional independence properties of the model—that is, that C is independent of y given x—, the first term of (5.8) can be expanded as

$$P(\mathbf{y}_{0:t}|C,x_{0:t}) = P(\mathbf{y}_{0:t}|x_{0:t}) \tag{5.12}$$

$$= \prod_{j=0}^{t} P(\mathbf{y}_j|x_j). \tag{5.13}$$

Replacing (5.11) and (5.13) into (5.8) results in

$$P(C,x_{0:t}|\mathbf{y}_{0:t}) = \frac{P(C)\prod_{j=0}^{t}P(\mathbf{y}_j|x_j)P(x_0|C)\prod_{l=1}^{t}P(x_l|x_{l-1},C)}{P(\mathbf{y}_{0:t})}. \tag{5.14}$$

Only the probability of the class given the observations is of interest, as such a marginalisation over the states $x_{0:t}$ can be performed and the denominator can be ignored due to it only being a normalisation constant, hence:

$$P(C|\mathbf{y}_{0:t}) \propto \sum_{x_{0:t}} P(C)P(x_0|C)\prod_{l=1}^{t}P(x_l|x_{l-1},C)\prod_{j=0}^{t}P(\mathbf{y}_j|x_j). \tag{5.15}$$

The conditional probability in (5.15) can then either be computed using full enumeration or the Viterbi sequence can be computed to approximate the probability of the class

$$\max_{x_{0:t}} P(C|\mathbf{y}_{0:t}) = \alpha P(C)P(x_0|C)\max_{x_t}(P(x_t|x_{t-1},C)\max_{x_{0:t-1}} P(y_{0:t}|x_{0:t}). \tag{5.16}$$

5.6 METHOD

Once the mel-cepstrum coefficients have been extracted and clustered appropriately, the transition probabilities for each class between the different "cepstral-states" is computed and stored through supervised learning. A frame is observed, divided into appropriate sized bursts and each burst is assigned a "state" based on its cepstrum coefficients and the GMM cluster center it is closest to. Once this has been done, the state transitions are calculated. To avoid errors based on low emission probabilities, transition probabilities that are zero are set to small non-zero values. ($1e-4$) The class membership for each cepstral state is also computed. It should be apparent that certain cepstral-states are more likely to occur in certain classes than others. This information is also useful and it is collected at the same time. It also forms the basis for GMM based classification.

In the classification model, the most likely sequence through the Markov model has been determined using the Viterbi method. Tests are performed to investigate the effects of different FFT lengths and required confidences, and the GMM-HMM method developed here is compared to Billik's method [6]. For Billik's method, the implemented confidence metric is the percentage of total votes received. Meanwhile, the GMM-HMM method uses the likelihood output of the HMM.

Two different studies on two different datasets have been conducted for this research. The first focuses on distinguishing between broad classes (human/Bipedal motion, walking animals and grazing animals) for practical purposes. The second study focuses on distinguishing between two specific species. The specific species are zebra and giraffe. They were chosen because these two animals usually walk at

roughly equivalent speeds in our data, while their relative body motions and spectrograms differ in terms of periodicity and the top Doppler shift achieved by the legs of the animal. This difference is believed to directly relate to the physiology of their movement.

5.6.1 Stationarity tests

In speech recognition, 20–30ms is considered stationary [62]; however, Doppler walking motion differs from this. To determine the length that the data is stationary, the data was subjected to stationarity tests. Specifically, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was performed [63]. The KPSS test determines if the time series is one where statistical properties such as the mean and variance are constant over time. To determine the average length that the data is stationary, the burst length was varied and the percentage of bursts that passed the stationarity test were recorded. These tests were performed to determine the maximum logical burst length because the assumption is made that the data is stationary for the duration of of the burst length. This varied over different tests and classes but the goal was to obtain a general sense of the maximum burst length to limit the scope of the study.

The results showed that for a burst length of 128 points (100ms), 71% of samples were considered stationary. For 64 points (50ms), 85% of bursts would pass a stationarity test. At longer burst lengths, more bursts would be considered non-stationary than stationary. Consequently, it was decided to limit the scope of the study to burst lengths of 128 points or less because many of the bursts that would not pass a stationarity test could still be considered psuedo-stationary at these lengths. Finally, the length of overlap needed to be determined. No overlap was included between the bursts because Bilik's method specifically does not use an overlap, and also because votes and "states" would not be statistically independent with the addition of an overlap.

5.6.2 Broad class tests

The classes have been distinguished as follows in the broad class classification tests. Moving animals (zebra and giraffe) that were observed walking straight (or relatively straight) towards the radar were grouped together as a group called walking animals. Animals that were observed while grazing (zebra, giraffe and rhino) will also frequently present themselves as slow moving targets as they lift their heads looking for predators and move across the fields slowly grazing. This Doppler profile is similar to

clutter in some regards but it is much more consistent and independent of wind. These observations have been grouped into a class called grazing animals. Finally, willing human participants were observed at a range of around 3km walking towards the radar installation and were grouped as human targets. No distinction is made between whether both the ranger and the researcher or only one of the two is present in the processing cell. Distinguishing between multiple and single humans in a cell is possible and has been done in [6]. This will typically be moved to a secondary classifier for a more accurate distinction.

Another class was added, which has no members for classification and compares a sequence of Cepstral states with a random walk. If a class is classified as ”Other”, then the classifier believes that the target is more likely to have been generated by a random walk than an actual target in its profile and as such defers the decision until more observations can be obtained. This class should reduce the classifier's confidence in itself, giving a more accurate representation of how sure the system is of its classification. All accuracies are recorded as resultant remaining accuracy; that is, the percentage correct classification after rejections have been removed from the totals.

The FFT length was varied between half burst (32 Point, 25ms), full burst (64 Point,50ms) and double burst (128 Point,100ms), and a study was performed to examine the effects of the burst length. The results were also compared to a GMM based classification system that was presented by Bilik [6] (with different observation times) without the addition of the HMM. The GMM is the only classification method employed for the first observation of each test of the GMM-HMM method because no transitions can be observed for a single observation and no time varying or information is considered. These tests include a test on the effect of classifying with the individual species clustered separately, as well as clustering them together.

5.6.3 Giraffe–zebra tests

The FFT size for the window was set equal to a full burst length of 128 (100ms) and the overlap was set to 0 for the zebra and giraffe classification. The observed frame size was varied for illustrative purposes between 1 and 40 consecutive bursts. This ToT is an important factor on scanning radar systems. An increased ToT will also increase the revisit time, which will make accurate tracking more

difficult to achieve. This study is conducted on zebra/giraffe data because of their significant similarity when looking at them in terms of spectral distribution observable in the spectrogram.

5.6.4 Varying required confidence study

The effects of using a percentage based confidence scores for rejection are also investigated. For the percentage rejection based classification, instead of simply using a MAP approach for classification, the class needs to obtain above a certain percentage confidence to be classified. This percentage is varied to observe the effect.

Finally, tests were performed to determine the ability of a required confidence metric to reduce false alarm rates and improve overall accuracy. Research has shown that there is a high correlation between low confidence classifications and incorrect classification. Consequently, the use of confidence metrics to avoid false alarms in classification is often used. For the GMM-HMM method, the required confidence score is the probability of the highest probability model having spawned the observed sequence. For Bilik's method, the required confidence score is the percentage of votes obtained. To determine the effects of different thresholds, the required threshold is plotted versus rejection rate and classification accuracy. An example of this is shown for a 1s ToT in Fig. 5.8 and Fig. 5.9.

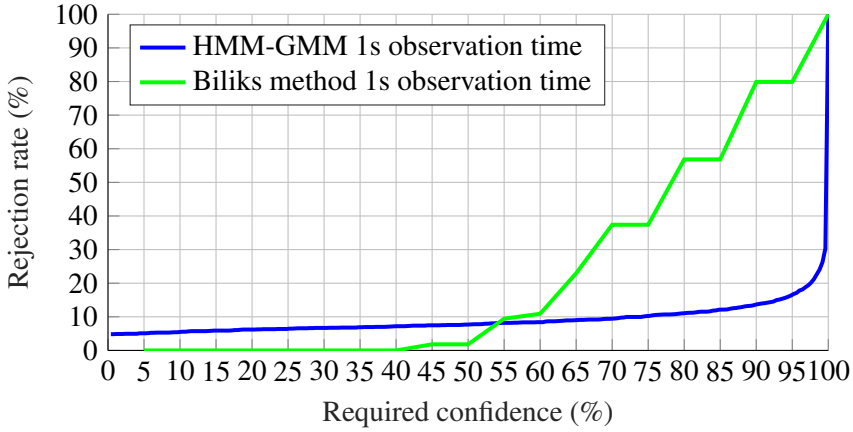


Figure 5.8. Comparison of required confidence versus percentage of rejected samples. It can be seen that the thresholds are not comparable and this results in different rejections rates at a specific required confidence.

Fig. 5.8 and Fig. 5.9 show the required confidence versus the rejection rate and rejection rate versus accuracy for ToT of 1s. Using these graphs, a satisfactory operational point can be chosen for testing.

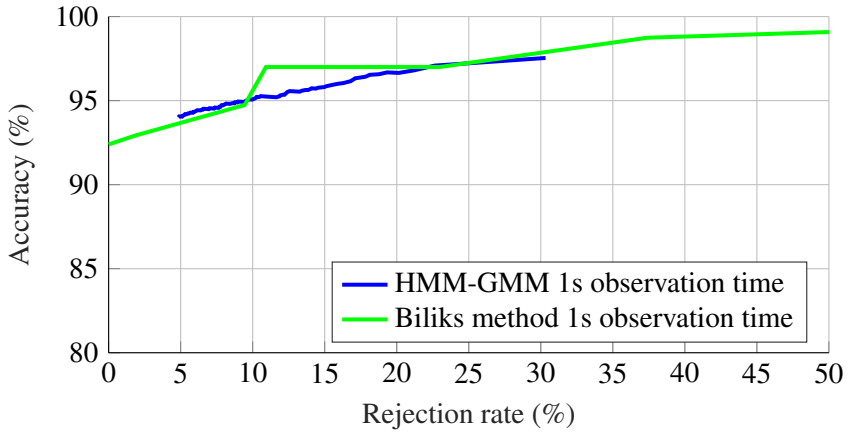


Figure 5.9. Rejection rate versus remaining accuracy for different classifiers. The figure shows that for any given rejection rate, the improvement to the accuracy is approximately equivalent regardless of the scoring system or method used. The spike in the rejection rate versus accuracy improvement for Bilik's method can be attributed to the discrete nature of the scoring system.

In the tests in this work, the confidence was varied to determine the exact effect over longer ToT's. It is difficult to compare the two methods based on their rejection rate and required confidence because the two score's function differently and the scores for the methods are not relateable. The GMM-HMM scoring system is heavily dependent upon the ToT because the confidence of the model drastically increases when the ToT increases, while Bilik's method scoring remains relatively constant over all ToT but only fractional values can be considered for shorter ToT. The important point from Fig. 5.9 is that for any given rejection rate, the improvement to the accuracy is approximately equivalent. However, controlling the rejection rate is difficult because the confidence score requirement can be easily varied but not the rejection rate.

5.7 RESULTS

5.7.1 Broad class separation study

The histogram plots of the cepstrum coefficients are shown in Fig. 5.10. This histogram plot shows that only Class 2 and Class 3 are easily separable. However, once a GMM is trained on the data individually and the components of the GMM compared using KL-Divergence, it quickly becomes apparent that the trained GMMs are highly separable and that the individual components of the GMM

are not similar in any way with KL-divergence scores never being less than 2. This explains the very good results obtained by Bilik's method in the following sections and why the GMM-HMM method does not give the expected improvement.

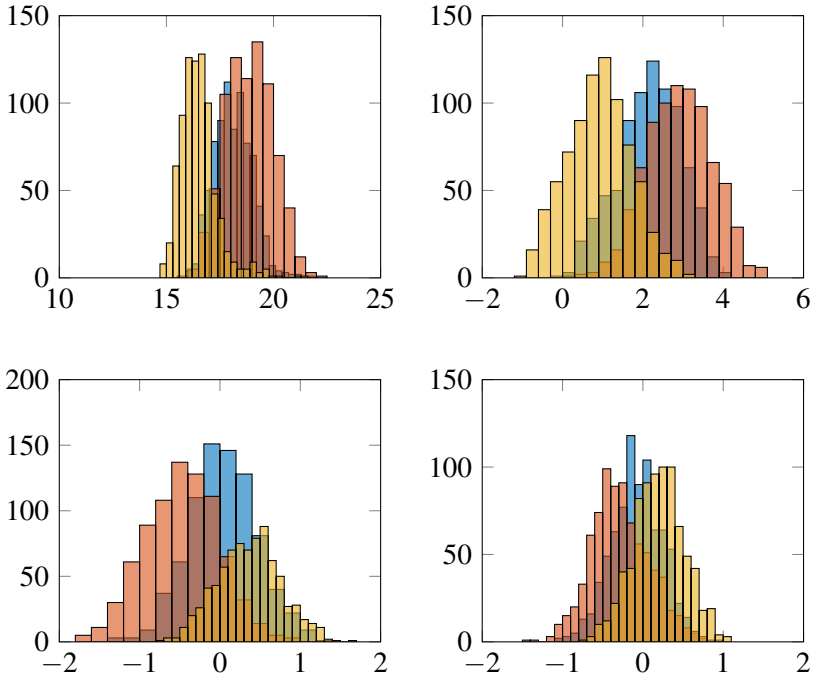


Figure 5.10. Histogram of cepstrum for broad classes showing the separation between the cepstrum coefficients for the 4 most discriminative cepstrum, from the histogram plot of the cepstrum coefficients it would seem that the cepstrum coefficients fall within the same range, and only Class 2 and 3 are significantly separable based on only their cepstrum values.

5.7.2 GMM clustering complexity and model order selection study

There are two possible approaches to solving the clustering for the GMM models for the GMM-HMM approach. The first approach combines the cepstral data from all the classes together and then fits a GMM model to the combined feature space. This method fits the idealised idea of how the GMM-HMM method should work best but leads to a more complex GMM space to cluster. The second approach splits the cepstral data of the classes up class by class and then fits a GMM to each class separately. However, this approach has one technical difficulty. When GMM models are clustered separately, the likelihoods produced by the model are not relateable to other models. For instance, if the model for Class 1 reports that the likelihood for component 1 is 0.1 and the model for Class 2 reports that the likelihood for component 3 is 0.5, then those numbers have nothing to do with one another. Unless

the models are combined in some fashion, the likelihoods produced by the models mean nothing and they cannot be related to one another. To solve this problem, the individual GMMs are merged after clustering by combining all of the components into a new GMM using the number of samples per component to re-estimate the PDFs of the different components. After clustering the GMM models the result of each class is a set of Gaussian components, each with a mean, variance and mixing proportion, this mixing proportion is equivalent to the number of samples assigned to the component. Using these variables from each class a new GMM can be formed using all the components of the different classes and re-estimating the PDF's using the number of samples assigned to each component to estimate the PDF of the new GMM.

To determine the effects of the clustering complexity, a study was performed to investigate the effects of training the GMM models with the classes combined before clustering, and for each class clustered individually and then combined. The results are shown in Fig. 5.11. It can be seen that separating the data of the classes and then training the GMM models for each class individually before re-merging them into a single GMM for the GMM-HMM model provided a significant improvement above that of the GMM trained on the cepstral coefficients of all the classes combined. However, the standard deviation in the three-fold cross validation results seemed excessive. Once the results were inspected in more detail, it became apparent that the HMM part of the GMM-HMM model had insufficient data to train upon. The standard deviation was very large and compared to Bilik's method the performance was rather poor. It was expected that this model would outperform the GMM method of Bilik because it includes additional time varying information that is not considered by Bilik's method. The model also showed results that are indicative of an over fit model in that it would make errors with a very high confidence. These results are shown in Table 5.5.

The AIC score used to determine the number of components in the GMM would create a good fit for the GMM model but this AIC score would not take into account the parameters of the HMM part of the classifier, as such the AIC score would indicate that a model order of 27 (10 components for human, 10 for Grazing animals, 7 for grazing animal) was the best. This large amount of components results in an extremely complex transition space. As an illustration, the 27 components would result in 729 possible transitions while the actual training dataset would only contain approximately 3000 transitions. Consequently, there may be insufficient data available to train the HMM part of the classifier and the poorly trained HMM model causes a significant increase in standard deviation and inconsistency in the results.

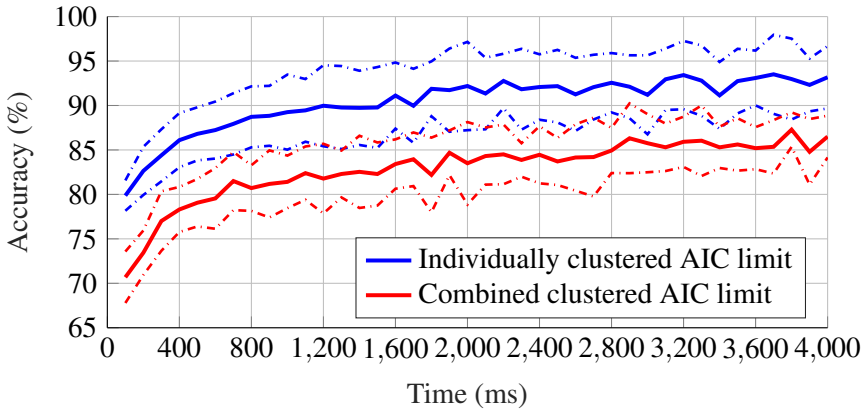


Figure 5.11. Comparison of the results obtained by different clustering approaches, it can be seen that clustering the classes for the GMM individually and then re-merging the separate GMM models gives a drastic improvement over combined clustering. The dash-dot curves indicates confidence bands.

To remedy this situation, it was determined that the GMM model would have to be restricted to a lower number of components. Given that there is no known alternative to AIC for an GMM-HMM combined model and the creation of such a scoring system falls beyond the scope of this work, it was chosen to scale down the GMM (while keeping the ratios of the number components per GMM for each class given by the AIC score intact). To determine the effect of this decision, the GMM model was re-clustered with a fixed amount of components while maintaining the ratio of the components from the GMM clustered by the AIC. The results of this method are shown in Fig. 5.12. It can be seen that halving the number of components in the GMM improved the accuracy above that of using the AIC score to determine the number of components. It was determined that a thorough cross-validation-test study would have to be conducted to compensate for these results.

A custom three-fold cross validation strategy was then implemented, which is depicted in Table 5.3. The data fold partitioning is performed such that in any sub-fold shown in Table 5.3, one set is completely held out until testing. This ensures that the test set results are never used to determine model order.

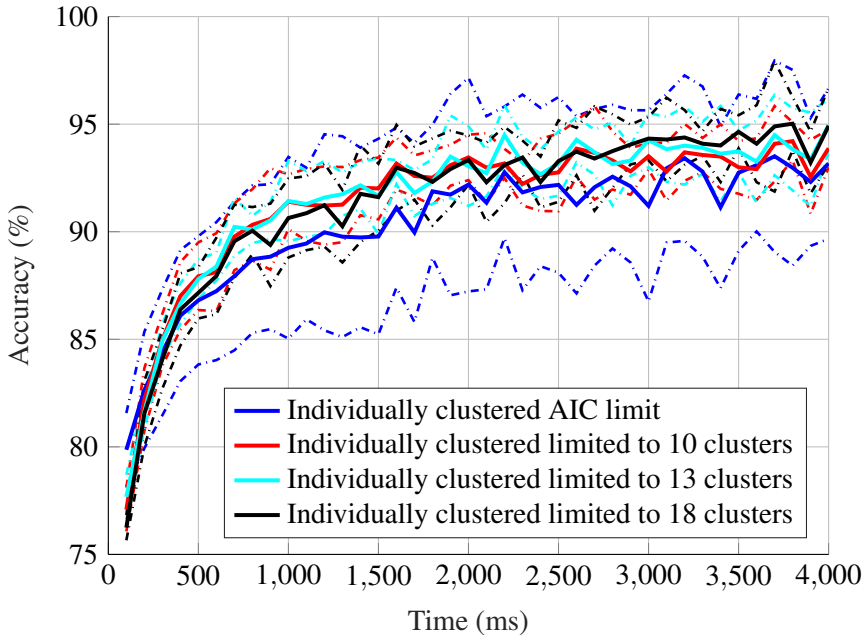


Figure 5.12. Comparison of the results obtained by different numbers of components. From the figure, it would seem that 13 components (5 human, 5 walking animal and 3 grazing animal) give the best results.The dash-dot curves indicates confidence bands.

Table 5.3. This table indicates how the data was partitioned for training, validation and testing over three-folds. Consider for example a "sub-partition" a). Fold number 3 is held out for a test set, and as such no test data is used to in any way determine the models resulting from the training/validation set combinations 1/2 and 2/1. The same can be said for sub-partitions b) and c). The validation results are averaged within each sub-partition (i.e. average of only two results per sub-partition), however the test results are averaged over all six test results.

	Training	**Model Validation**	**Testing**
a)	Train on 1	Validate on 2	Test on 3
	Train on 2	Validate on 1	Test on 3
b)	Train on 2	Validate on 3	Test on 1
	Train on 3	Validate on 2	Test on 1
c)	Train on 1	Validate on 3	Test on 2
	Train on 3	Validate on 1	Test on 2

It was decided to train and validate models by increasing and decreasing the number of components per class one at a time. It was assumed that, much like a gradient decent problem, the accuracy would decrease if the choice was worse and increase if the change was positive. The dataset was divided into three partitions of roughly equal size. The validation strategy of Table. 5.3 was used then used to obtain the results in Table. 5.4. The model order that had the highest accuracy averaged between its two-fold validation procedure on its validation sets was chosen to be run on the test set. In the case of Table. 5.4 this corresponds to the model with 14 components (5/5/4). This model would then be tested on the third partition of the data to obtain the test results for this fold of the data. This process is repeated for every fold of the data, rotating the sets and recording results for each fold separately. All of the tests performed for the rest of this work are reported as the mean and standard deviation for three different folds of data with the model order for the three tests being chosen based on their specific validation results. This should avoid contamination of the test set results because a model order decision is never made based on any result in which the test set is used.

Table 5.4 shows that the model order is an interesting space to explore. Increasing the model order increases training set accuracy but degrades validation accuracy, while reducing the model order reduces both training and validation accuracy. This can be explained if you consider that reducing the model order reduces the problems associated with the poor training of the HMM transition space but reduces the accuracy of the models fit in the cepstral (GMM) space. The HMM model also increases the complexity of the model and this complexity is not accounted for in the AIC score. Without this scoring system, it would be very difficult to justify specific model order choices. To deal with this problem, cross validation is employed to obtain a general idea of classifier performance.

5.7.3 Broad class classification

In Fig. 5.13 and Table 5.6 it can be seen that the accuracy of the system is a function of the ToT or, to put it differently, the number of consecutive bursts observed from the target. The longer that the target is observed, the more accurate the classification becomes. However, in a non-staring application, the revisit time will increase because the radar system has to rotate slower to observe more bursts from each cell. This results in an interesting trade off: as the ToT is increased, a scanning radar's revisit time will increase, making accurate tracking of the target more difficult but yielding a higher accuracy for classification. This problem could be reduced by increasing the number of beams on the radar system;

however, this increases the power and financial cost of such a radar system. Bilik *et. al.* [6] showed that approximately 4 seconds of observation was optimal and that after this time frame longer staring would not significantly increase the accuracy of the classification system. Here, almost the same result is obtained with the actual staring time being reliant on the species. A human specifically requires an observation time of around 1 seconds because this is the period of their movement cycle, after 1 second the classification performance does increase but significant extra ToT is required for very little gain.

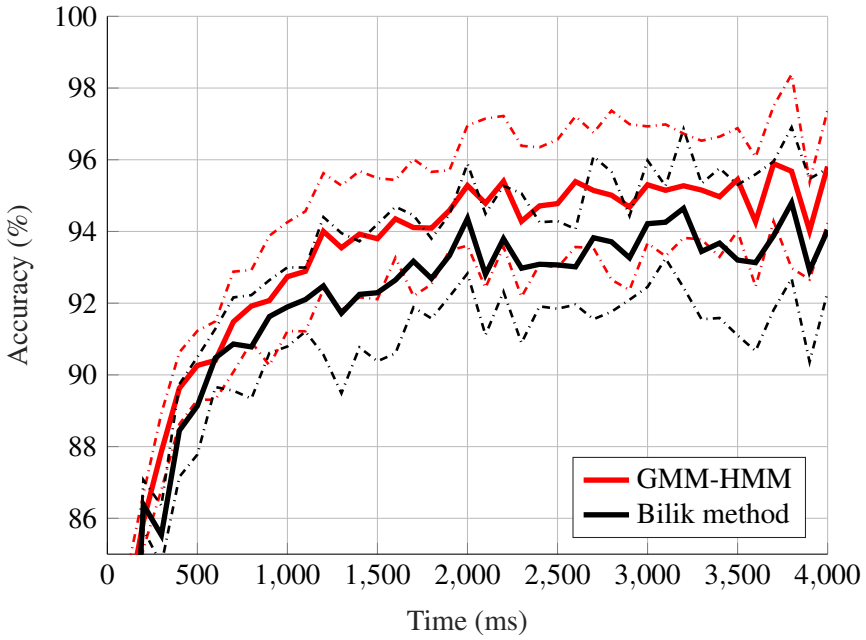


Figure 5.13. A comparison of human classifier accuracy versus observed time for Bilik's method and HMM-GMM. The dashed-dotted lines indicate $\pm$ single standard deviations for the different folds.

The results in Table 5.5 caused some initial fear that the model was over fitting the data due to classes being confused with a relatively high confidence. However, it was determined that the error was most likely caused by incorrect labelling of the spectrogram data. Specifically, when walking in the field, the ranger would periodically look back to see if the researcher was still following closely or would stop and scan for animals. The human targets would often have to walk around scrubs and obstacles. As such, their angle of incidence relative to the radar would vary and their profile would match grazing animals from time to time. This effect can be seen in Fig. 5.14.

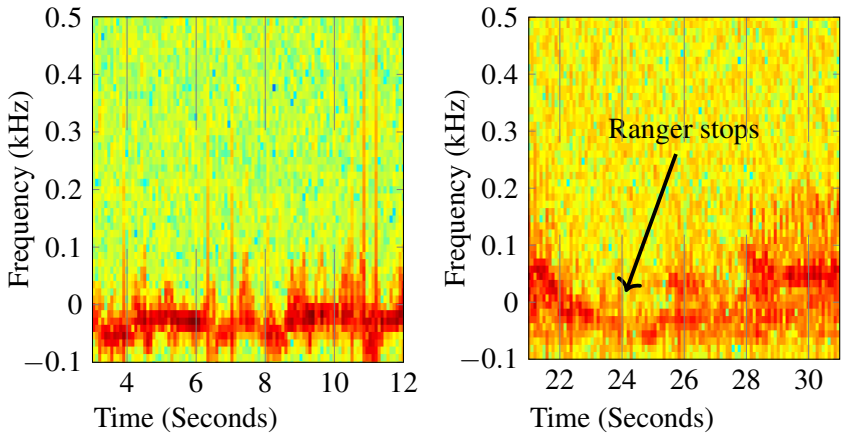


Figure 5.14. Comparison of grazing animal on the left and human on the right, from the image it can be seen why humans and grazing animals may be confused. During the measuring process, the ranger would often stop to look back if the researcher was following closely. Hence, for these short periods of time, the human and grazing animal classes could match due to labelling errors.

5.7.4 Varying FFT lengths and GMM comparison

In Fig. 5.15 a comparison of the accuracy of the classifier with varying burst lengths and ToT is shown. It is known that the length of the FFT effects the Doppler resolution of the system and, hence, the information available for exploitation. However, for longer ToT, the radar system would have to observe the target for a substantial amount of time, slowing down the scanning and, therefore, creating difficulties for the radar tracking system. The results show that the FFT resolution provides much more information and separability than transition space and as such the higher bursts lengths always perform better. The test was not run for 200ms burst lengths because at this length the burst can no longer consistently be considered stationary.

5.7.5 Varying required confidence study

The required confidence of each of method is varied to examine the level of improvement that is achievable with the implementation of the confidence scores described in Section 5.6.4.

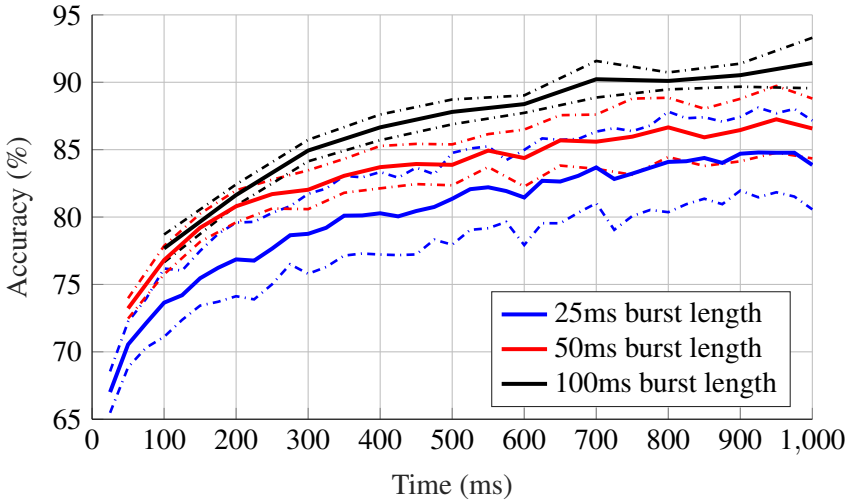


Figure 5.15. Comparison of Classifier Accuracy versus ToT for varying burst lengths. For each burst length, the first observation does not use a HMM and instead the first classification is done purely on a GMM of that length.The dash-dot curve indicates confidence bands.

5.7.5.1 Bilik's confidence score

The results for the required confidence study of Billik's method are shown in Fig. 5.16. The results show that the inclusion of a required confidence greatly increases the accuracy and reduces the false alarm rate at the cost of not classifying an amount of uncertain observations. Deferring classification and waiting for more observations is preferable to making a false classification and deploying park rangers without proper confirmation. Bilik's method is capable of obtaining an almost perfect classification accuracy with a high enough confidence score. However, this comes at the cost of not classifying almost half of all of the samples. If a lower threshold of 65% is instead chosen, then the accuracy is only improved by 5%.

Overall, the confidence metrics for Bilik's method performs much better than expected. This is not surprising because M/N strategies are often used in radar track validation and they are usually a powerful yet simple solution to false track confirmations [64]. This works just as well in a classification environment to defer classification decisions in the presence of uncertain classifications.

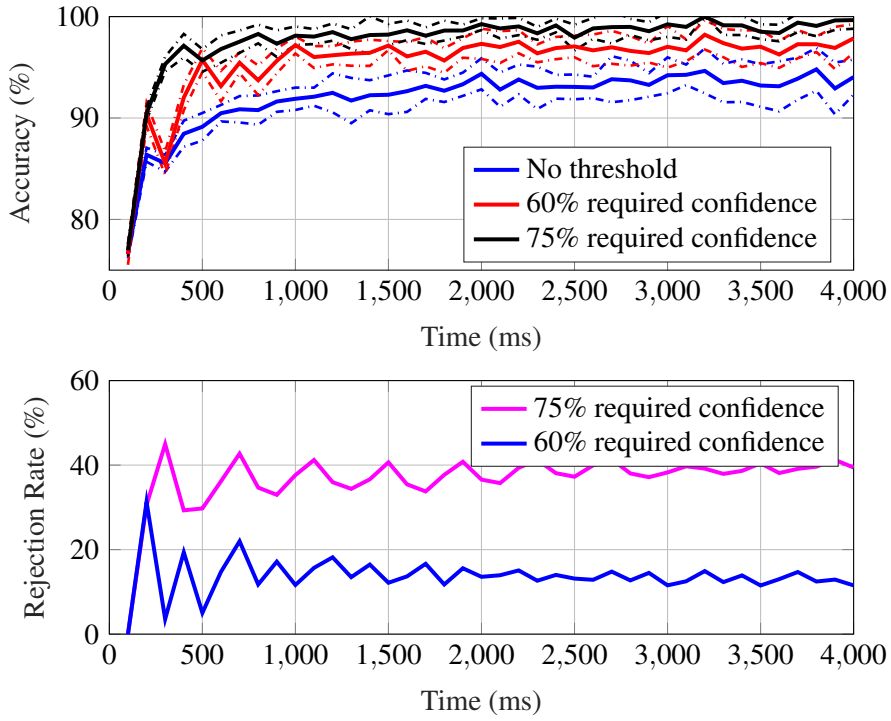


Figure 5.16. A comparison of classifier remaining accuracy and rejection rate versus ToT for Bilik's method using confidence scores.The dash-dot curve indicates confidence bands.

5.7.5.2 GMM-HMM confidence score

The results for the required confidence study of the GMM-HMM method are shown in Fig. 5.17. The results show that the confidence metric is again capable of increasing the accuracy and reducing false alarms. However, unlike Bilik's method, where a specific confidence requirement would cause a uniform improvement to the system over all ToT's, the improvement and rejection rate are highly based on the ToT. The HMM-GMM model tends to become more confident in its accuracy with more observation. Even wrong classifications may have relatively high confidence scores for long ToT because there are reasonable transitions to other class states (for instance, a series of less likely states for the correct class are observed in between two short sequences of very likely states for the incorrect class). However, investigation of the confidence of samples has shown that the confidence for wrong classifications will still be lower (even though that value may seem high) than correct observations. As such, for long ToT, higher confidence scores should be required for significant improvement.

Fig. 5.14 compares a grazing animal and a human. The human targets tended to have to walk around

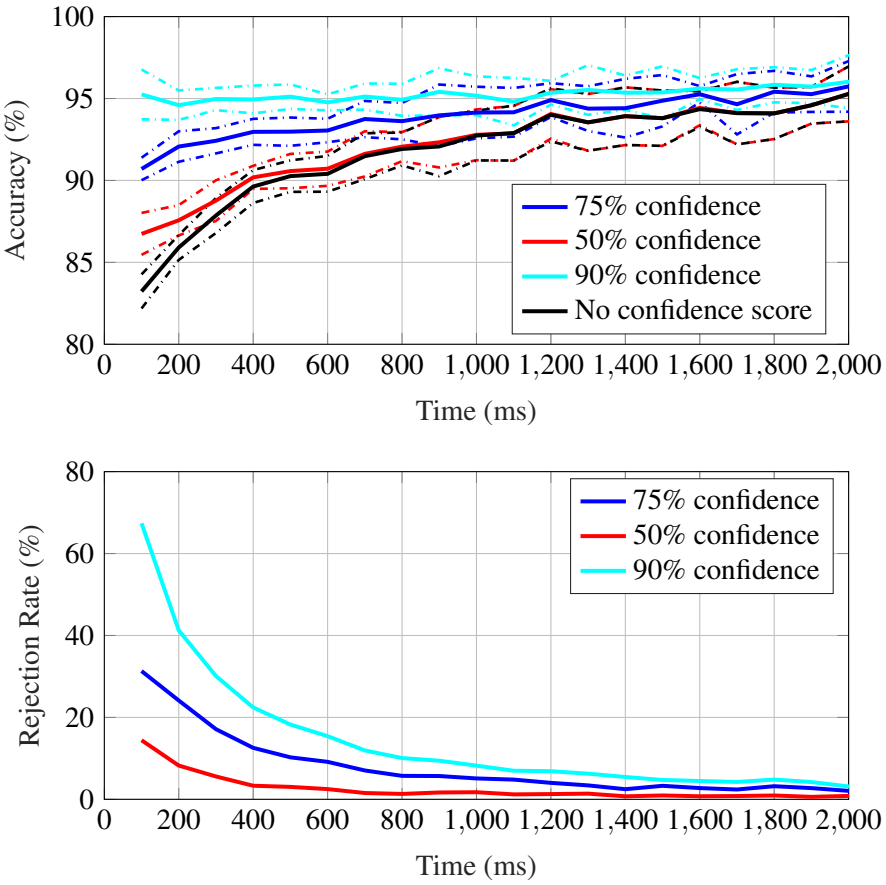


Figure 5.17. A comparison of classifier remaining accuracy and rejection rate versus ToT for GMM-HMM method using confidence scores. The dash-dot curve indicates confidence bands.

obstacles and the ranger would often look back to see if the researchers were still following close or they may stop to scan the environment for wild animals that may be close by. As such, for short periods of time the human and grazing data could be confused with a relatively high confidence. This phenomena requires further study and methods to deal with it should be investigated.

5.7.6 Giraffe–zebra separation study

Separation studies were performed on all data in this work. Here, the separation study for the giraffe and zebra is explicitly shown to illustrate some key observations. The histograms of the cepstrum extracted from the data for giraffe and zebra classes are shown in Fig. 5.18. The separation boundaries between the different classes do not appear to be clear at all in the histograms. However, once the data had been clustered using the GMM, the resulting components of the GMM were very distinct. This

can be seen when evaluating the KL-Divergence between the individual components of each class. A snapshot of this is shown in Table. 5.7. Even though the cepstrum do not appear to give significant distinguishing attributes, once they are clustered, the resulting clusters are significantly separated in space.

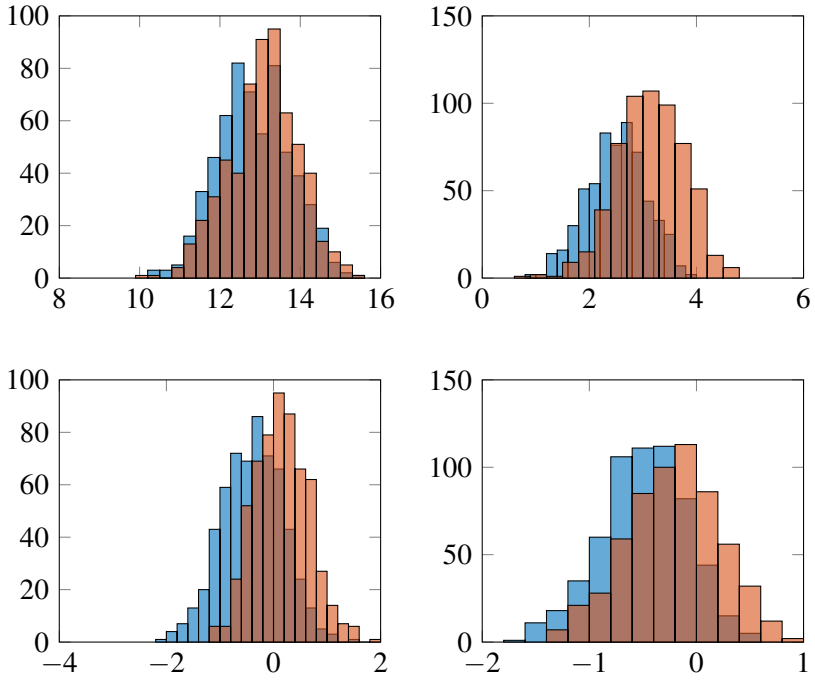


Figure 5.18. Histogram of most discriminative cepstrum for giraffe and zebra showing the separation between the cepstrum coefficients, from the histogram plot of the cepstrum coefficients it would seem that the cepstrum coefficients very similar.

5.7.7 Giraffe–zebra classification

A comparison of a typical giraffe spectrogram and a typical zebra spectrogram is shown in Fig. 5.19. The pattern of the spectrogram repeats itself in a different fashion based on the species being observed. Comparing arbitrary windows of giraffe and, for instance, zebra, their cesptrum coefficients could be the same at that time instance. However, it was believed that the way that they transition through different "cepstral-states" allows them to be accurately classified by the HMM. Notably, the difference between a zebra and giraffe occurs in the frequency of leg movements as well as the frequency shift obtained by their legs. This ratio is expected to correspond to the length of the animals legs and should be a distinguishing feature for every species.

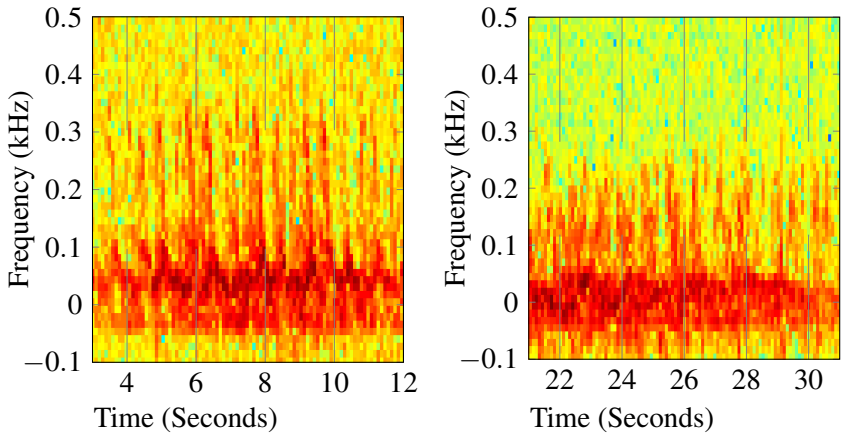


Figure 5.19. Comparison of zebra (left) and giraffe (right) spectrograms.

Data from these two species was collected and then compared. These targets were moving at roughly the same main body speed and accurate classification of these species would be reliant on accurately distinguishing based on micro-Doppler features. An example of the spectrogram extracted for each of these species is shown in Fig. 5.19. Three-fold cross validation on the results was performed and the final combined results are tabulated in Table 5.6.

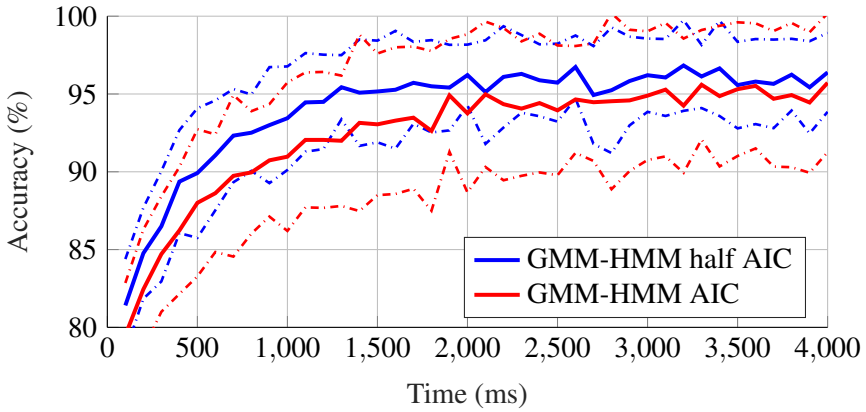


Figure 5.20. Comparison of zebra and giraffe results when reducing the component order derived by AIC score by half.The dash-dot curve indicates confidence bands.

For the giraffe zebra data the AIC score was again used to determine the order of the models; however, it was again determined that the resulting model was too complex for the GMM-HMM approach and instead the cross-validation strategy described in Section 5.7.2 was used. This resulted in a model with half the number of components of the AIC method. The comparison between the accuracies in these models is shown in Fig. 5.20, resulting again in improved classification performance.

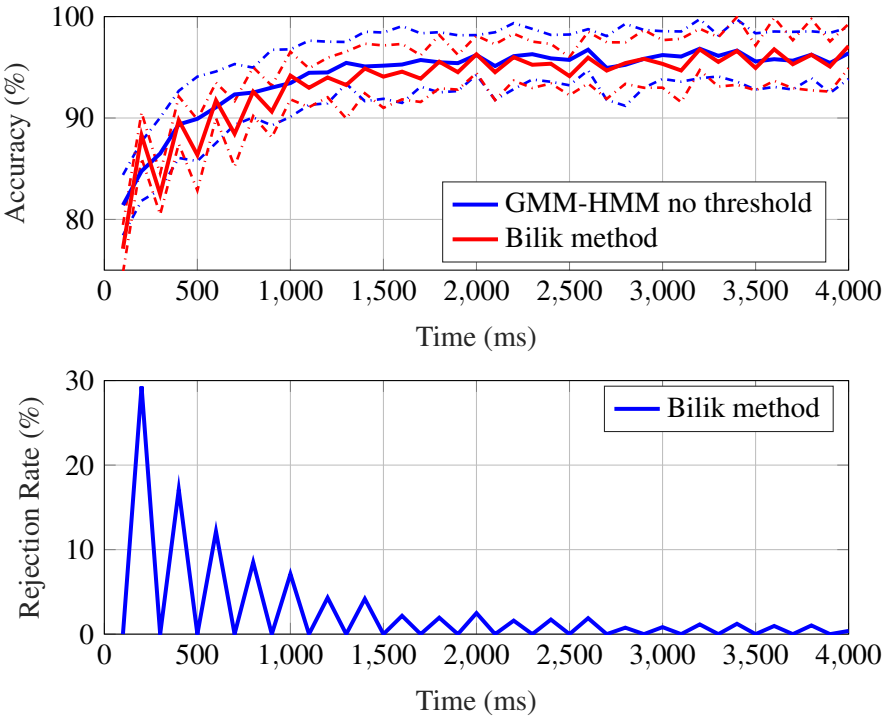


Figure 5.21. Comparison of zebra And giraffe results.The dash-dot curve indicates confidence bands.

The results show that if the classifier is trained to distinguish between two different species and as long as the target was observed for a long enough time-frame, accurate classification is possible using either method. The longer that a target is observed, the more accurate the classification is. If the classifier is trained to distinguish between only two narrow species, then it can be seen that higher degrees of accuracy is possible due to the model having less variance in its classes and the specific differences can be latched onto more easily. Thus, for accurate species classification, one could use a hierarchy tree form classifier. First, species would be differentiated based on their body type and behaviour (walking/running/grazing) and then subsequently classified by a secondary classifier to distinguish them from other species based on their micro Doppler.

5.8 CONCLUSION

This chapter has shown that accurate human-animal classification of slow moving land based targets is possible when they are observed in their natural habitat. The classification performance is dependent on the ToT. It should be noted that since the diversity of the dataset on which this study is based is somewhat limited, further data is required to definitively characterise the performance of the classifier

and identify possible shortcomings. Using the currently available dataset, the approach used produces promising results. It is shown that human-animal classification is possible with reasonable accuracies while the targets are in their natural environments at ranges exceeding 1km.

Although the results obtained for the GMM-HMM classifier are not statistically significantly better than Billik's method for any specific ToT, the GMM-HMM consistently performs better over all ToTs. As mentioned in Section 5.7.2, if only the AIC score were to be used to determine model order, then this dataset would not contain enough transitions per fold to accurately estimate the HMM transition probabilities between GMM components. This is a form of over fitting because, in addition to the GMM parameters determined by the AIC, HMM transition parameters are added, which results in an overly complex model. Hence, by using a validation set, the best model order was determined in Section 5.7.2; however, the performance seems to be insensitive to model selection choices over a fairly broad band close to the best model order.

The design of the GMM-HMM model was based on the results in Chapter 4, where it was shown that the cepstrum for human targets would repeat in a periodic manner and it was believed that this periodicity would be present in many other species. Even if individual mel-cepstrum observations were similar, the state transitions over time would serve as a discriminating feature. Although the HMM-GMM method seems to outperform Billik's method, the performance increase is not particularly marked. This can be better understood when the bi-directional KL-distances between the GMM components are investigated. A bi-directional KL-distance of less than one would mean that the specific components were very similar. Upon studying the KL-distances for the giraffe and zebra GMM components, it was found that the smallest KL-distance was approximately 2 and the average was approximately 50. This implies that the cepstral features for different animals are not similar. Consequently, the cepstral information contained in the GMM provides most of the discriminative power of the features. On the other hand, because there are very few transitions that overlap between different species (classes), the transitions do not significantly contribute to the classification results. It is believed that with a more representative data, clusters would overlap more and the contribution of state transitions in the GMM-HMM model could contribute to improved performance of the GMM-HMM classifier.

The classification method suggested in this work improves the accuracy of the classification for short ToTs (between 150ms and 300ms). However, if the target is observed for a very short time (shorter than 50ms), then any temporal or spectral based methods would struggle because of the time scales at

which human or animal motion occur.

It is very important to note that the AIC model order obtained for the clustering of the GMM models is not incorrect. Even though the results would suggest that the AIC score over fits the data, these exact model orders are used in Bilik's method and give excellent results. The reduction in accuracy for the GMM-HMM model at this high order is believed to be caused by the insufficient data available for training the transition space of the HMM. It is believed that if given sufficient training, the data the accuracy will be improved by the higher model order.

The GMM-HMM model derived here seems to out perform that of Bilik's method. However, the dataset is too limited to make broad sweeping conclusions about the superiority of either method.

Table 5.2. Three fold cross validation confusion for broad before the model complexity was dropped. The results are reported as percentages for GMM-HMM method followed by Biliks method.

500ms ToT			
HMM-GMM/Bilik	**True Class**		
Classifier Out	**Human**	**Walking**	**Grazing**
Human	80.8/84.58	3.7/3.48	11.55/9.11
Walking Animal	10.13/5	93.14/94.06	2.3/0.96
Grazing Animal	9/10.4	3.14/2.46	86.09/89.93
Other	2.1/-	2.13/-	1.63/-
1s ToT			
HMM-GMM/Bilik	**True Class**		
Classifier Out	**Human**	**Walking**	**Grazing**
Human	84.4/87.06	2.81/0.237	7.69/5.62
Walking Animal	8/3.41	95.38/96.1	0.7/0.37
Grazing Animal	7.6/9.54	1.8/1.53	91.58/94.01
Other	0/-	0/-	0/-
1.5s ToT			
HMM-GMM/Bilik	**True Class**		
Classifier Out	**Human**	**Walking**	**Grazing**
Human	86.8/88.39	0.27/1.7	7.42/5.5
Walking Animal	6.7/3.8	94.5/96	0.5/0.8
Grazing Animal	6.37/7.7	2.7/2.2	92.03/93.5
Other	0/-	0/-	0/-
2s ToT			
HMM-GMM/Bilik	**True Class**		
Classifier Out	**Human**	**Walking**	**Grazing**
Human	89.63/91.15	2/2.3	5.88/3
Walking Animal	5.85/2.68	96.69/96.36	0/0.37
Grazing Animal	4.52/6.17	1.32/1.32	94.12/96.64
Other	0/-	0/-	0/-

Table 5.4. Table of accuracy for model order of a single fold of data (average of two validation results per sub-partition). Model order is denoted as human/walking animal/grazing animal model order. There is a broad range close to the best point where model order does not affect performance significantly; that is, performance is insensitive to choice of model order within this range. However, performance does degrade for very high and low model orders.

Model order	GMM Component Partition	Training Accuracy	Validation Accuracy
6	2/2/2	96.62% ±1.74	92.26% ±1.15
8	3/3/2	96.97% ±1.66	94.13% ±1.64
10	4/4/2	97.68% ±1.48	93.99% ±0.27
12	5/5/2	97.29% ±0.45	93.28% ±0.054
12	4/5/3	97.83% ±1.66	92.96% ±0.1
12	5/4/3	97.90% ±1.66	94.42% ±1.49
13	5/5/3	97.61% ±0.7	93.68% ±2.21
14	5/6/3	97.61% ±0.81	93.41% ±2.36
14	6/5/3	98.38% ±0.73	92.53% ±2.52
14	5/5/4	98.96% ±0.062	94.66% ±0.24
15	6/6/3	97.28% ±1.41	93.22% ±2.72
18	7/7/4	98.74% ±0.5	94.08% ±3.83
21	8/8/5	99.02% ±0.072	93.11% ±4.78
23	9/9/5	99.18% ±0.009	93.93% ±4.96

Table 5.5. Three-fold cross validation confusion matrices for broad classes using MAP approach. The results are reported as percentages for HMM-GMM method followed by Bilik's method.

500ms ToT			
HMM-GMM/Bilik	**True Class**		
Classifier Out	**Human**	**Walking**	**Grazing**
Human	85.87/84.58	3.39/3.48	11.09/9.11
Walking animal	4.33/5	93.06/94.06	1.36/0.96
Grazing animal	6.93/10.4	1.49/2.46	86.55/89.93
Other	2.87/-	2.07/-	1.0/-
1s ToT			
HMM-GMM/Bilik	**True Class**		
Classifier Out	**Human**	**Walking**	**Grazing**
Human	89.2/87.06	2.31/0.237	8.79/5.62
Walking animal	3.47/3.41	94.39/96.1	3.7/0.37
Grazing animal	4.93/9.54	1.65/1.53	90.48/94.01
Other	2.4/-	1.65/-	0.37/-
1.5s ToT			
HMM-GMM/Bilik	**True Class**		
Classifier Out	**Human**	**Walking**	**Grazing**
Human	91.04/88.39	1.49/1.7	7.14/5.5
Walking animal	3.98/3.8	96.29/96	0.55/0.8
Grazing animal	4.18/7.7	0.99/2.2	91.48/93.5
Other	0.8/-	1.24/-	0.82/-
2s ToT			
HMM-GMM/Bilik	**True Class**		
Classifier Out	**Human**	**Walking**	**Grazing**
Human	93.09/91.15	1.66/2.3	6.25/3
Walking animal	3.19/2.68	96.69/96.36	0/0.37
Grazing animal	2.66/6.17	0.33/1.32	93.75/96.64
Other	1.06/-	1.32/-	0/-

Table 5.6. Three-fold cross validation results for giraffe and zebra classification.

500ms ToT		
HMM-GMM/Bilik	**True Class**	
Classifier Out	**Zebra**	**Giraffe**
Zebra	86.13/78.9	7.79/9.04
Giraffe	13.87/21.1	92.21/90.96
1s ToT		
HMM-GMM/Bilik	**True Class**	
Classifier Out	**Zebra**	**Giraffe**
Zebra	90.7/88.08	4.88/2.18
Giraffe	9.3/11.92	95.12/97.82
1.500s ToT		
HMM-GMM/Bilik	**True Class**	
Classifier Out	**Zebra**	**Giraffe**
Zebra	93.4/89.58	3.78/3.11
Giraffe	6.6/10.42	96.22/96.89
2s ToT		
HMM-GMM/Bilik	**True Class**	
Classifier Out	**Zebra**	**Giraffe**
Zebra	94.44/93.1	2.68/1.79
Giraffe	5.56/6.9	97.32/98.21

Table 5.7. KL-distances for giraffe and zebra class components.

Giraffe Zebra KL-Distance between components					
Component Number	1	2	3	4	5
1	16.89	17.12	30.28	27.86	33.72
2	3.45	8.01	15.37	15.60	8.67
3	4.64	6.65	5.91	7.35	5.09
4	7.62	8.42	9.78	2.87	11.64
5	30.41	15.77	37.44	34.37	20.25

CHAPTER 6 CONCLUSION AND FUTURE WORK

6.1 DISSERTATION SUMMARY

Chapter 4 has shown that human motion from MOCAP data could be used to simulate radar data; however, this data tends to be very "clean" and results in cepstrum coefficients that are not representative of recorded data. A significant effort would have to be done to make simulated data that matches recorded data. This effort could in fact be so great that it would be counter productive because the data needed to validate the model would most likely be enough to successfully train a classifier. Chapter 4 also investigated whether or not S-band radar could be used. An S-band radar has the advantage of some inherent foliage penetration capability; however, this foliage penetrations capability comes at the cost of Doppler resolution. More specifically, the discriminating features of the Doppler profile are lost; that is, the appendages (arms and legs) become unobservable. It should be noted that the beam width of the test radar may not have been narrow enough and that a radar more specifically designed for ground target observation would most likely present better Doppler profiles; however, narrowing the beam width of the radar comes at an increase in the antenna size. For a very narrow beam width at S-band, the radar would no longer be mobile and such a large antenna would come with other logistical challenges. Therefore, the X-band is still the best current solution for ground target tracking.

Chapter 5 has shown that it is possible to classify targets using their Doppler profiles when given a long enough observation time. Depending on ToT, classification rates of 75% to 90% are achievable. Although the dataset is limited, it has been shown that a GMM-HMM or GMM model works well to classify targets. While the GMM-based Bilik's method works much better than expected and the

implemented GMM-HMM approach performs worse than expected, this deviation from the expected results can be described by a closer analysis of the data.

Upon investigation of the clustering of the cepstral features, it was noticed that, contrary to initial belief, the cepstral features do not overlap in their space as expected and as such the order in which a frame of data moves through the cepstral states is not as significant as which spectral states the data moved through. In essence, most of the classification power of the model comes from its GMM-model and not from the HMM transition model. This may change with a more comprehensive database but investigating the clustering of individual species is recommended before implementing the more complex GMM-HMM approach. Nonetheless, it has been shown that this method (and Bilik's method) can distinguish between different species of animals, something that has not been tested or done before this work.

6.2 FUTURE WORK

Although this research has shown to be promising, many questions remain unanswered. For example, there is no reason why Bilik's method as implemented would not work in a scanning radar environment, and it is recommended that data should be gathered and tested for this purpose. There is an opportunity for a massive data gathering operation due to the fact that the Meerkat radar systems in the Kruger National Park are now operational. This data gathering will allow for more robust testing of classification systems and should allow for the integration of classification aided tracking. Furthermore, the data in this research only considers targets moving towards the radar and the effect of varying incidence angles has to be investigated and a solution should be derived for any variations that might occur due to change in incidence angle.

Fusion of tracking and classification systems to improve both should be investigated. This fusion of the data could further benefit from tracking filters that are designed to classify radar tracks based on their movement characteristics. Poachers tend to move in a goal orientated fashion, sticking to well defined known paths and their radar tracks tend to differ from the animals that graze and loiter in an area. Research by [31] has shown that such characteristic behavior in maritime pirates can be used to classify motion. A similar approach could be used with moving ground targets and this form of classification could be fused into the surveillance systems.

REFERENCES

[1] P. Mashala. (2013, may) Developing farmers feel the scourge of stock theft. [Online]. Available: www.farmersweekly.co.za

[2] R. Duffy, "War, by conservation," *Geoforum*, vol. 69, pp. 238–248, 2016.

[3] J. Li, S. L. Phung, F. H. C. Tivive, and A. Bouzerdoum, "Automatic classification of human motions using Doppler radar," *The 2012 International Joint Conference on Neural Networks*, pp. 1–6, 2012.

[4] S. Challa and G. W. Pulford, "Joint target tracking and classification using radar and ESM sensors," *IEEE transactions on Aerospace and Electronic Systems*, vol. 37, no. 3, pp. 1039–1055, 2001.

[5] A. Stove and S. Sykes, "A Doppler-based automatic target classifier for a battlefield surveillance radar," 2002.

[6] I. Bilik, J. Tabrikian, and A. Cohen, "GMM-based target classification for ground surveillance doppler radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 267–278, Jan 2006.

[7] P. O. Molchanov, J. T. Astola, K. O. Egiazarian, and A. V. Totsky, "Target classification by using pattern features extracted from bispectrum-based radar Doppler signatures," in *Radar Symposium (IRS), 2011 Proceedings International*. IEEE, 2011, pp. 791–796.

[8] P. Molchanov, J. Astola, K. Egiazarian, and A. Totsky, "Classification of ground moving radar targets by using joint time-frequency analysis," in *Radar Conference (RADAR), 2012 IEEE*. IEEE, 2012, pp. 0366–0371.

[9] R. Boulic, N. M. Thalmann, and D. Thalmann, "A global human walking model with real-time kinematic personification," *The visual computer*, vol. 6, no. 6, pp. 344–358, 1990.

[10] P. Van Dorp and F. Groen, "Feature-based human motion parameter estimation with radar," *IET Radar, Sonar & Navigation*, vol. 2, no. 2, pp. 135–145, 2008.

[11] A. Cilliers and W. Nel, "Helicopter parameter extraction using joint time-frequency and tomographic techniques," in *Radar, 2008 International Conference on*. IEEE, 2008, pp. 598–603.

[12] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328–1337, May 2009.

[13] D. P. Fairchild and R. M. Narayanan, "Classification of human motions using empirical mode decomposition of human micro-Doppler signatures," *IET Radar, Sonar Navigation*, vol. 8, no. 5, pp. 425–434, June 2014.

[14] M. Jahangir, K. Ponting, and J. O'loghlen, "A robust doppler classification technique based on hidden markov models," 2002.

[15] M. O. Padar, A. E. Ertan, and Ç. ĝatay Candan, "Classification of human motion using radar micro-Doppler signatures with hidden Markov models," in *Radar Conference (RadarConf), 2016 IEEE*. IEEE, 2016, pp. 1–6.

[16] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 8–12, Jan 2016.

[17] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using an artificial neural network," in *2008 IEEE Antennas and Propagation Society International Symposium*, July 2008, pp. 1–4.

[18] T. S. Jordan, "Using convolutional neural networks for human activity classification on micro-Doppler radar spectrograms," in *SPIE Defense+ Security*. International Society for Optics and Photonics, 2016, pp. 982 509–982 509.

[19] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[20] C. Bishop, "Novelty detection and neural network validation," *IEE Proceeding on Image Signal Processing*, vol. 141, pp. 217–222, 1994.

[21] J. A. Nanzer and R. L. Rogers, "Bayesian classification of humans and vehicles using micro-Doppler signals from a scanning-beam radar," *IEEE Microwave and Wireless Components Letters*, vol. 19, no. 5, pp. 338–340, May 2009.

[22] G. E. Smith, K. Woodbridge, and C. J. Baker, "Naïve Bayesian radar micro-Doppler recognition," in *Radar, 2008 International Conference on*. IEEE, 2008, pp. 111–116.

[23] F. Fioranelli, M. Ritchie, S. Z. Gurbuz, and H. Griffiths, "Feature diversity for optimized human micro-Doppler classification using multistatic radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. PP, no. 99, pp. 1–1, 2017.

[24] B. Widrow, M. E. Hoff *et al.*, "Adaptive switching circuits," in *IRE WESCON convention record*, vol. 4, no. 1. New York, 1960, pp. 96–104.

[25] M. I. Jordan and D. E. Rumelhart, "Forward models: Supervised learning with a distal teacher," *Cognitive science*, vol. 16, no. 3, pp. 307–354, 1992.

[26] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[27] S. Herman and P. Moulin, "A particle filtering approach to FM-band passive radar tracking and automatic target recognition," in *Proceedings, IEEE Aerospace Conference*, vol. 4, 2002, pp. 4–1789–4–1808 vol.4.

[28] A. A. Ksienski, Y. Lin, and L. J. White, "Low-frequency approach to target identification," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1651–1660, Dec 1975.

[29] S. Hudson and D. Psaltis, "Correlation filters for aircraft identification from radar range profiles," *IEEE Transactions on Aerospace and Electronic systems*, vol. 29, no. 3, pp. 741–748, 1993.

[30] H.-J. Li and S.-H. Yang, "Using range profiles as feature vectors to identify aerospace objects," *IEEE Transactions on Antennas and Propagation*, vol. 41, no. 3, pp. 261–268, 1993.

[31] J. J. Dabrowski and J. P. de Villiers, "Maritime piracy situation modelling with dynamic Bayesian networks," *Information fusion*, vol. 23, pp. 116–130, 2015.

[32] G. Kouemou, C. Neumann, and F. Opitz, "Exploitation of track accuracy information in fusion technologies for radar target classification using Dempster-Shafer rules," in *Information Fusion, 2009. FUSION'09. 12th International Conference on*. IEEE, 2009, pp. 217–223.

[33] S. M. Herman, "A particle filtering approach to joint passive radar tracking and target classification," DTIC Document, Tech. Rep., 2002.

[34] S.-K. Han, H.-T. Kim, S.-H. Park, and K.-T. Kim, "Efficient radar target recognition using a combination of range profile and time-frequency analysis," *Progress In Electromagnetics Research*, vol. 108, pp. 131–140, 2010.

[35] M. A. Richards, J. A. Scheer, W. A. Holm *et al.*, *Principles of modern radar*. Citeseer, 2010.

[36] V. C. Chen, "Analysis of radar micro-Doppler with time-frequency transform," in *Proceedings of the Tenth IEEE Workshop on Statistical Signal and Array Processing (Cat. No.00TH8496)*, 2000, pp. 463–466.

[37] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-Doppler effect in radar: phenomenon, model, and simulation study," *IEEE Transactions on Aerospace and electronic systems*, vol. 42, no. 1, pp. 2–21, 2006.

[38] C. Clemente, L. Pallotta, A. D. Maio, J. J. Soraghan, and A. Farina, "A novel algorithm for radar classification based on doppler characteristics exploiting orthogonal pseudo-zernike polynomials," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 417–430, January 2015.

[39] M. N. Cohen, "An overview of high range resolution radar techniques," in *Telesystems Conference, 1991. Proceedings. Vol.1., NTC '91., National*, Mar 1991, pp. 107–115.

[40] P. R. Runkle, P. K. Bharadwaj, L. Couchman, and L. Carin, "Hidden Markov models for multiaspect target classification," *IEEE Transactions on Signal Processing*, vol. 47, no. 7, pp. 2035–2040, 1999.

[41] P. Runkle, L. H. Nguyen, J. H. McClellan, and L. Carin, "Multi-aspect target detection for SAR imagery using hidden Markov models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 1, pp. 46–55, 2001.

[42] S. P. Jacobs and J. A. O'Sullivan, "Automatic target recognition using sequences of high resolution radar range-profiles," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 364–381, Apr 2000.

[43] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy, "Automatic fall detection based on Doppler radar motion signature," in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. IEEE, 2011, pp. 222–225.

[44] Y. Geng, J. Chen, and K. Pahlavan, "Motion detection using RF signals for the first responder in emergency operations: A PHASER project," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*. IEEE, 2013, pp. 358–364.

[45] E. Scheme and K. Englehart, "A comparison of classification based confidence metrics for use in the design of myoelectric control systems," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 7278–7283.

[46] C. Demant, C. Demant, and B. Streicher-Abel, *Industrial image processing*. Springer, 1999.

[47] B. R. Mahafza, *Radar systems analysis and design using MATLAB*. CRC press, 2002.

[48] P. van Dorp and F. Groen, "Human walking estimation with radar," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 150, no. 5, pp. 356–365, 2003.

[49] S. Gallone, "FOPEN radar for UGS applications," in *Radar (Radar), 2011 IEEE CIE International Conference on*, vol. 1. IEEE, 2011, pp. 173–175.

[50] M. Davis, "Foliage penetration radar: Detection and characterization of objects under trees, scitech. pub," *Inc. Â, Â New YorkÂ*, 2011.

[51] J. B. Billingsley, *Low-angle radar land clutter: measurements and empirical models*. IET, 2002.

[52] C. Cai, W. Liu, J. Fu, and Y. Lu, "Doppler frequency extraction of foliage penetration radar based on the Hilbert-Huang transform technology," in *Radar Conference, 2004. Proceedings of the IEEE*. IEEE, 2004, pp. 170–174.

[53] J. Smit and J. Cilliers, "Insights into factors contributing to the observability of a submarine at periscope depth by modern radar: Part 2-EM simulation of mast RCS in a realistic sea surface environment," in *Antennas and Propagation in Wireless Communications (APWC), 2012 IEEE-APS Topical Conference on*. IEEE, 2012, pp. 989–992.

[54] Carnegie Mellon University, "CMU graphics lab motion capture database," 2014, data obtained from the CMU Graphics Lab Motion Capture Database , http://mocap.cs.cmu.edu/.

[55] J. De Witt, M. Alahmadi, and A. Alzamil, "Design and use of a mobile, X-band, high range resolution, radar research facility," 2012.

[56] L. Deng and D. O'Shaughnessy, *Speech processing: a dynamic and optimization-oriented approach*. CRC Press, 2003.

[57] D. Tahmoush, J. Silvious, and J. Clark, "An UGS radar with micro-Doppler capabilities for wide area persistent surveillance," *SPIE Defense, Security, and Sensing*, pp. 766 904–766 904, 2010.

[58] L. Ljung, *System identification*. Wiley Online Library, 1999.

[59] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[60] J. R. Hershey and P. A. Olsen, "Approximating the Kullback Leibler divergence between Gaussian mixture models," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 4. IEEE, 2007, pp. IV–317.

[61] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

[62] I. V. McLoughlin, *Speech and Audio Processing: a MATLAB-based approach*. Cambridge University Press, 2016.

[63] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of econometrics*, vol. 54, no. 1-3, pp. 159–178, 1992.

[64] S. Blackman and R. Popoli, "Design and analysis of modern tracking systems(book)," *Norwood, MA: Artech House, 1999.*, 1999.