CRANFIELD UNIVERSITY

GIOVANNI MARINO

ANALYSIS OF PERFORMANCE OF AUTOMATIC TARGET RECOGNITION SYSTEMS

CRANFIELD DEFENCE AND SECURITY

THESIS

PhD

CRANFIELD UNIVERSITY

CRANFIELD DEFENCE AND SECURITY

DEPARTMENT OF INFORMATICS AND SYSTEMS ENGINEERING

THESIS

PhD

Academic year 2011-2012

Giovanni Marino

Analysis of performance of automatic target recognition systems

Supervisor: Dr. E. J. Hughes

September 2011

Graecia capta ferum victorem cepit

To my family

To my Grandmother

To Prof. Carmelo Chemi

Abstract

An Automatic Target Recognition (**ATR**) system is a sensor which is usually able to recognize targets or objects based on gathered data. The application of automatic target recognition technology is a critical element of robotic warfare. ATR systems are used in unmanned aerial vehicles and cruise missiles. There are many systems which are able to collect data (e.g. radar sensor, electro-optic sensor, infra-red devices) which are commonly used to collect information and detect, recognise and classify potential targets. Despite significant effort during the last decades, some problems in ATR systems have not been solved yet.

This Ph.D. tried to understand the variation of the information content into an ATR system and how to measure as well as how to preserve information when it passes through the processing chain because they have not been investigated properly yet. Moreover the investigation focused also on the definition of class-separability in ATR system and on the definition of the degree of separability. As a consequence, experiments have been performed for understanding how to assess the degree of class-separability and how the choice of the parameters of an ATR system can affect the final classifier performance (i.e. selecting the most reliable as well as the most information

preserving ones).

As results of the investigations of this thesis, some important results have been obtained: Definition of the class-separability and of the degree of class-separability (i.e. the requirements that a metric for class-separability has to satisfy); definition of a new metric for assessing the degree of class-separability; definition of the most important parameters which affect the classifier performance or reduce/increase the degree of class-separability (i.e. Signal to Clutter Ratio, Clutter models, effects of despeckling processing). Particularly the definition of metrics for assessing the presence of artefacts introduced by denoising algorithms, the ability of denoising algorithms in preserving geometrical features of potential targets, the suitability of current mathematical models at each stage of processing chain (especially for clutter models in radar systems) and the measurement of variation of information content through the processing chain are some of them most important issues which have been investigated.

Acknowledgments & Preface

A Ph.D. is always a very difficult intellectual challenge and summarizing these my last three years of research is still more difficult because of the huge number of experiences, nevertheless I would like to share some considerations I acquired during my presence in the United Kingdom.

Firstly, a very interesting issue is concerned with the empirical approach of British scientists with whom I cooperated (a special thank to Dr. Evan J. Hughes, my supervisor, and to my thesis committee: Dr. Clive Alabaster and Dr. Keith Morrison). They indeed gave some suggestions which helped me to understand better the problems I am about to expose in this thesis.

During my studies I had also the opportunity of reading three very interesting authors who helped me in choosing which scientific approach was more suitable for my investigations: Teilhard de Chardin, Prof. Roger Scruton and Prof. Paul Karl Feyerabend. As a consequence the development of my Ph.D. research had been based on the idea of trying to answer the question 'Why an event happens' instead of explaining 'How an event happens'. Rephrasing previous statement, my thesis is based on the principle that answering to the question 'Why an event happens?' is more important than understanding 'How' the process occurs physically. To be more precise, the

answer of the second question maximizes its information content when it helps the researchers in understanding the problem (i.e. answering 'Why an event happens?').

The reason of viewing the scientific research in such a way can be summa-

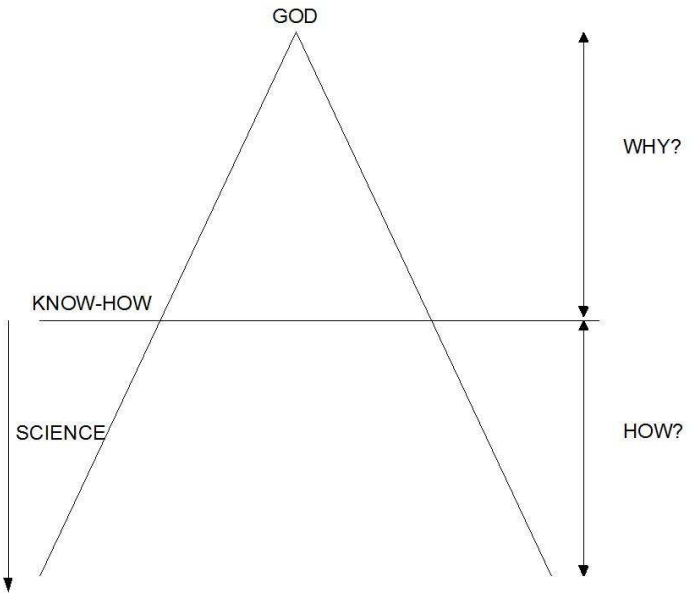


Figure 1: *How Teilhard de Chardin described the dual problem science and faith.*

rized as follows. In Figure 1 the model reported by Teilhard de Chardin in '*Science and Christ*' is shown. It represents how scientific research has increased the knowledge of natural phenomena historically. As a consequence, the scientific research introduced a materialist view of the world. Many authors during history have already described this view of the world (e.g. Thales of Miletus, Anaximenes, Anaximander, Democritus etc.) however this topic is not concerned with this preface. The interesting issue shown in Figure 1 however is how the increasing of knowledge of the world by answering 'How a phenomenon happens' makes more and more important answering to the

question 'Why this phenomenon happens?', which contains more information.

In scientific terms this dilemma can be translated by considering the dual problem between Galilean and non-Galilean scientific methods. As known, the classical scientific approach is based on hypothetic-deductive model and it can be summarized as:

1. Use your experience: consider the problem and try to make sense of it.
2. Form a conjecture: try to state an explanation.
3. Deduct a prediction: Generalize the problem solution.
4. Test: check if your generalization is correct.

The question 'How?' hence is important in order to find a solution for a problem, by basing it on the observations of the phenomena. As reported by Prof. Fayerabend in '*Against method*', the Galilean scientific approach hence tends to simplify too much the problem under investigation in terms of information content, to be more precise, it tends to make problem investigation more dependent on the observations. In practice Galilean scientific method fails because it tends to generalize a solution of a problem by considering just a finite number of observations.

As a consequence, what is the best method of research to generalize a problem solution? In my opinion, the best way of investigation is to answer the 'Why' question because it tends to generalise automatically the problem solution. It indeed investigates the correlation between elements which create problems instead of understanding only how those elements work.

Finally, another important suggestion about the importance of answering 'Why a phenomenon happens?' is given by Prof. Roger Scruton. He has reported in his book '*An Intelligent Person's Guide To Philosophy*' that human beings are rational people, therefore they tend to answer automatically, when something happens to them, to the question 'Why?', therefore it is closer to human nature to answer this question in terms of 'Why?' instead of answering to 'How?'.

In conclusion the reasons of considering the 'Why' question instead of 'How' question can be summarised as follows: It is more natural to answer to the 'Why' question; 'Why' tends to generalise the problem solution automatically and it has a bigger information content. That is why I considered the mentioned guideline during my investigation.

Giovanni Marino

Contents

List of Tables

List of Figures

Glossary

P_{fa} Probability of False Alarm, 30

ATR Automatic Target Recognition, ii, 1, 6, 11, 12, 18, 21, 22, 24, 25, 27, 28, 30, 34, 35, 38, 40–44, 47, 48, 53, 57, 58, 62, 63, 65, 67, 112, 128, 130, 136, 153, 163, 189, 190, 194–197

AUC Area Under the Curve, 41, 51, 79, 86, 87, 94

BC Before Classification, 50

CAD Computer-Aided Design, 40, 43

CDF Cumulative Distribution Function, 83, 84, 94, 102, 103, 170, 172, 173, 175, 178, 196

CFAR Constant False Alarm Rate, 6, 8, 11, 67, 163, 166, 167, 187, 195, 196

CR Contrast Ratio, 113

ECG Electrocardiogram, 13

EEG Electroencephalogram, 13

ENL Equivalent Number of Look, 113, 114, 116, 123, 128, 130, 137, 139, 153, 194, 195

FIFO First In First Out, 17

FPPS Fixed Priority Pre-emptive Scheduling, 18

HMM Hidden Markov Model, 11, 30, 189

HRR High Range Resolution, 6, 44

ICA Independent Component Analysis, 25, 68

IFF Identify Friend or Foe, 6

IQM Image Quality Metric, 113

ISAR Inverse Synthetic Aperture Radar, 6

ISLR Integrate SideLobes Ratio, 113

JEM Jet Engine Modulation, 6

KAPS Kolmogorov Appropriate Prediction of Separability, 91, 94, 99, 102, 107, 108, 192, 193, 197

LDA Linear Discriminant Analysis, 79

LIDAR Light Detection And Ranging, 12

MI Mutual Information, 36, 38, 68, 71

ML Maximum Likelihood, 139

MNR Multiplicative Noise Ratio, 113

MSE Mean Square Error, 11, 30, 123, 128, 139, 194

MSE Mean Squared Error, 113

MSTAR Moving and Stationary Target Acquisition and Recognition, 30, 43

MTI Moving Target Indication, 3

MWW Mann-Whitney-Wilcoxon, 128

NRA Non-Return-Area, 114

NVM Normalized Variance Metric, 116, 117, 123, 128, 130, 139, 153, 194, 195

OS Operating System, 15

PC Principal Component, 65

PCA Principal Component Analysis, 25, 65–67, 71, 190

PCI Probability of Correct Identification, 30

PDF Probability Density Function, 39, 57, 63, 69, 71, 83, 90, 92, 103, 107, 115–117, 119, 128, 130, 139, 143, 144, 149, 163, 169, 170, 172, 190, 195

PSF Point Spread Function, 118

PSNR Peak Signal-to-Noise Ratio, 113

RADAR RAdio Detection And Ranging, 3

RCS Radar Cross Section, 6, 23, 34, 115–119, 123, 128, 130, 136–139, 143, 144, 149, 153, 167, 189, 194, 195

ROC Receiver Operating Curve, 41

ROC Receiver Operative Curve, 51, 79, 86, 87, 94, 192

ROI Region Of Interest, 6, 8, 166

SAR Synthetic Aperture Radar, 3, 6, 21–23, 25, 27, 28, 30, 34, 36, 40, 42–44, 53, 57, 62, 63, 65, 67, 112–115, 117–119, 128, 138, 153, 163, 167, 172, 174, 182, 187, 189, 190, 194, 195, 197

SCR Signal to Clutter Ratio, 63, 113, 138, 180

SIAP Single Integrated Air Picture, 30

SJF Shortest Job First, 18

SNR Signal-to-Noise Ratio, 113

SRT Shortest Remaining Time, 18

SSR Secondary Surveillance Radar, 6

SVD Singular Value Decomposition, 66

SVM Support Vector Machine, 30, 84, 189

UT Unscented Theory, 24

VaR Value at Risk, 19

WWII Second World War, 3

Chapter 1

Introduction

1.1 Work overview

Classification is a very common task in human activities. People, objects and animals are usually classified to be managed in a more efficient manner. Historically, the introduction of automatic systems for managing and storing information has made easier and faster the classification of huge quantities of data. As a consequence, the interest of researchers in this topic has increased the during last decades especially in safety critical systems (i.e. military applications, medical systems, etc.), producing systems that perform fast classification (termed Automatic Target Recognition (**ATR**)). For instance, in military systems awareness of the presence of potential targets is of key importance, therefore a correct classification can reduce reaction time and collateral damage. In medical applications, however, identification of tumors is a crucial skill which presents very strict requirements in terms of errors in the detection process.

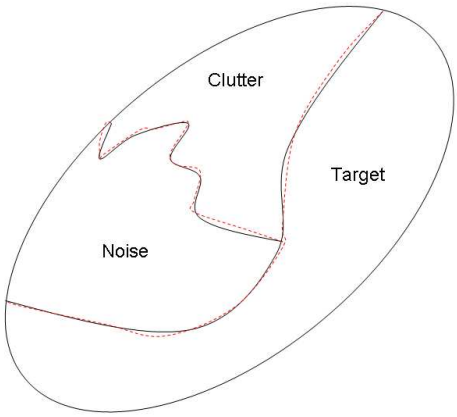


Figure 1.1: *Problem description: the main goal of an ATR system is to 'build' boundaries between equivalence classes of feature space, i.e. finding the partition of the feature space (solid black line). Unfortunately this is not possible, therefore a suboptimal solution has to be accepted (dashed red line). The main goal of this thesis is to understand how to asses the minimum error between borders (solid black line and dashed red line respectively), when several systems are compared.*

Generally speaking an ATR system can be considered as a sequence of tasks which processes information collected by instruments (usually a sensor or a set of sensors) in order to make a decision (as shown in Figure 1.1, i.e. finding the optimal partition of feature space and in Figure 1.2, i.e. finding the best mapping which is able to maximize the preserved information carried by sampled data). It is therefore important to understand what an ATR system should be able to do and what not. This thesis investigates the problem of performance analysis of ATR systems. Therefore our attention focuses on how information flow changes in an ATR signal processing chain.

1.1.1 History and principles of Radar systems

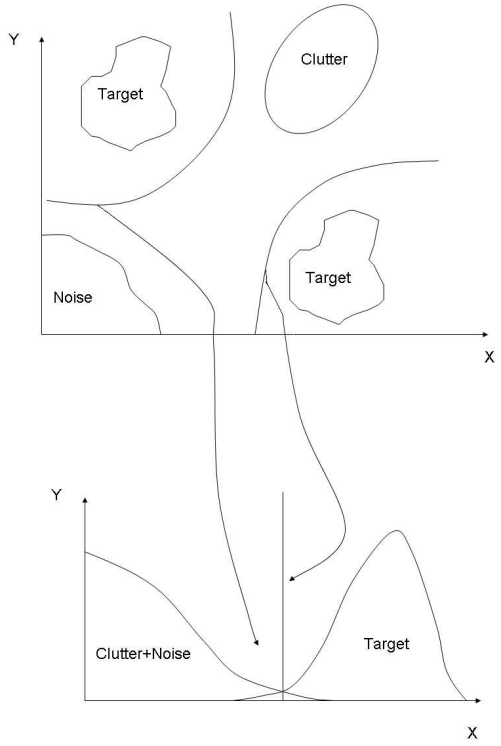


Figure 1.2: *ATR output: a generic ATR processing chain tends to reduce the dimensionality of the feature space in order to allow the classifier to perform a binary decision, thereby ATR system is theoretically able to map the feature space in 1-D Euclidean space. In practice however the 1-D mapping is not feasible because the optimal mapping is usually unknown.*

A RADAR ('*RAdio Detection And Ranging*') is an object-detection system which uses electromagnetic waves, specifically radio waves, to determine the range, altitude, direction, or speed of both moving and fixed objects such as aircraft, ships, spacecraft, guided missiles, motor vehicles, weather formations, and terrain. It has a transmitter that emits radio wave signals in predetermined directions. If an object is present in the scene, the electromagnetic wave is reflected and/or scattered in many directions (radar signals

are reflected especially well by materials of considerable electrical conductivity, especially: Most metals, sea water and land). The radar signals that are reflected back towards the transmitter are detected by the sensor. Due to the path attenuation, i.e. power interaction of radar signals with the illuminated scene objects, the received signal is usually very weak, therefore it is usually strengthened by the electronic amplifiers that all radar sets contain. The 'renewed' signal is hence processed in order to extract the information carried by the electromagnetic wave. The modern uses of radar are highly diverse, including air traffic control, radar astronomy, air-defense systems, antimissile systems; nautical radars to locate landmarks and other ships; aircraft anticollision systems; ocean-surveillance systems, outer-space surveillance and rendezvous systems; meteorological precipitation monitoring; altimetry and flight-control systems; guided-missile target-locating systems; and ground-penetrating radar for geological observations. Modern radar systems are associated with digital signal processing and are capable of extracting objects from very high noise levels.

Historically (Raymond C. Watson in '*Radar Origins Worldwide*' and Alan Dower Blumlein in '*The story of RADAR Development*') radar has been developed by several engineers, scientists and inventors and in particular the first studies had been performed by Heinrich Hertz in 1886 who showed that radio waves could be reflected from solid objects. Later the German Christian Huelsmeyer in 1904 was the first to use radio waves to detect 'the presence of distant metallic objects' and the first who received a patent for his detector. Despite this efforts, radar was not considered an important equipment for commercial and military application.

In August 1917 Nikola Tesla outlined a concept for primitive radar units, but his brilliant ideas were not considered by industry. As unlucky as Tesla were A. Hoyt Taylor and Leo C. Young, researchers working with the U.S. Navy, who in 1922 discovered that when radio waves were broadcast at 60 MHz it was possible to determine the range and bearing of nearby ships in the Potomac River. Despite Taylor's suggestion that this method could be used in darkness and low visibility, the Navy did not accept the idea.

From 1920 to 1940 the number of researchers and patents related to radar technologies increased, but the research had been performed independently and in great secrecy in several countries (e.g. Émile Girardeau in France, P.K.Oschepkov in USSR, Robert M. Page in USA, Rudolf K ühnhold in Germany and Robert A. Watson Watt in Great Britain). The British were the first to fully exploit radar as a defense against aircraft attack. This was spurred on by fears that Germany was developing 'death rays'. The Air Ministry asked British scientists in 1934 to investigate the possibility of propagating electromagnetic energy and the likely effect. Following a study, they concluded that a 'death ray' was impractical but that detection of aircraft appeared feasible and Robert A. Watson Watt was able to create a prototype (later patented) which was able to detect German aeroplanes.

The war expedited research to improve resolution, increase portability, and extend the utility of radar, including complementary navigation systems like Oboe used by the RAF's Pathfinder. After the second world war (**WWII**) have seen the use of radar in fields as diverse as air traffic control, weather monitoring, astronomy, and road speed control.

During the 1950's-1960's the research was focused on Moving Target Indica-

tion (**MTI**) radar which are able to discriminate the Doppler effects caused by moving targets to detect target motion, while suppressing echoes from stationary targets. Another important revolution in this period is the introduction of Phased Array Antenna technology in which a series of small transmitting elements operate collectively to enable dynamic formation of the antenna beam pattern.

In the 1970s radar systems were converted to imaging radar by the introduction of '*Synthetic Aperture Radar*' (**SAR**). With the introduction of digital processing techniques radars started having the ability of gathering high resolution ground images.

From the 1980s radar systems have improved their accuracy and they have been used in more and more commercial applications (velocity estimator for police forces, surveillance and recognition in military and non-military application for instance).

1.1.2 ATR system description

Attempts to add target discrimination features to the radar functionality have been made since the beginning of the radar era. The early experiments [1] were performed in 1937. Resonant dipoles were added to friendly aircraft so that their returns were distinctive from those of hostile aircraft. It was realized that such a system would have limited use when several aircraft flew in formation and focus shifted to using Secondary Surveillance Radar (**SSR**). By placing a transponder on targets to be observed by the radar they become 'co-operative targets' that transmit an enhanced version of the radar

signal when illuminated that contains encoded information about their identity. Similar technology is used in both civilian and military applications, although in military circles the technology is referred as Identify Friend or Foe (**IFF**), and both suffer from the weakness that they require the target to provide a truthful identity encoding.

Skolnik in [2] lists some radar principles and related phenomena that may be utilized by a radar to permit ATR: High Range Resolution (**HRR**) in which a one dimensional image of the target is produced and classified; Jet Engine Modulation (**JEM**) in which the characteristic frequency modulations induced in the echo signal by jet engines are identified; Radar Cross Section (**RCS**) fluctuation in which the angular variation of target's reflectivity is used as a discriminating feature; SAR whereby the motion of the radar platform is used to synthesize a large aperture antenna permeating the formation of a detailed terrain image in which targets can be recognized; Inverse Synthetic Aperture Radar (**ISAR**) which is the dual problem of SAR, where the target motion is used to synthesize the large antenna aperture.

Hence generally speaking, an ATR system consists of three main subsystems: a transducer (usually a sensor), which captures a finite set of object features, a signal processing unit, which processes the transducer signals in order to extrapolate object features and a classifier, which discriminates the targets of interest.

Unfortunately the described model is not feasible, therefore each described task consists of more than one subsystem depending on the nature of the ATR system (as shown in Figure 1.3 for a SAR/ATR system.). As for a SAR/ATR system, its processing model consists of several tasks as depicted

in Figure 1.4. It indeed performs a statistical test at each pixel; a test statistic with a threshold is applied and makes a preliminary target-declaration decision in order to reject most natural clutter. In practice Constant False Alarm Rate (**CFAR**) (i.e. detection) compares each pixel with a threshold in the following way:

$$\frac{(x-\mu_c)}{\sigma_c} > t \text{ if there is a target otherwise } x \text{ is considered clutter or noise} \tag{1.1}$$

(i.e. second order CFAR) where x is the intensity (power) under consideration (test pixel), μ_c and σ_c are the estimates of the local clutter mean and standard deviation respectively, and t is the threshold. The mean and the standard deviation are computed from a local square annular region, often surrounding the pixel of interest but far enough away from it to preclude the possibility of a target of interest occupying both the test and a portion of the annular region. Such a region is usually called Region Of Interest, (**ROI**) and usually it consists of about 100 pixels.

The output of detection step should be a set of potential targets, which represent the input of the discrimination stage. The main goal of the discrimination is to reject the residual natural clutter and most man-made clutter. Discrimination is performed by setting up a multidimensional feature space and the distance, in that space, between the feature vectors corresponding to the observed ROI and to an image of the target-type of interest is used to assess the likely class of the detected target. The most frequent features used in the discrimination step are:

- *Mass*, as the number of pixels in the principal-object region P.

Figure 1.3: *SAR/ATR scheme: The nature of the sensor affects its own structure. Indeed the transducer outputs (i.e. SAR images) are to be denoised firstly, and secondly a discrimination step is applied in order to separate the potential target features from the background and identifies Region Of Interest (**ROI**). As a consequence the block Denoising processing, corresponding to the signal processing unit of the ATR model consists of a denoising step and a discrimination task.*

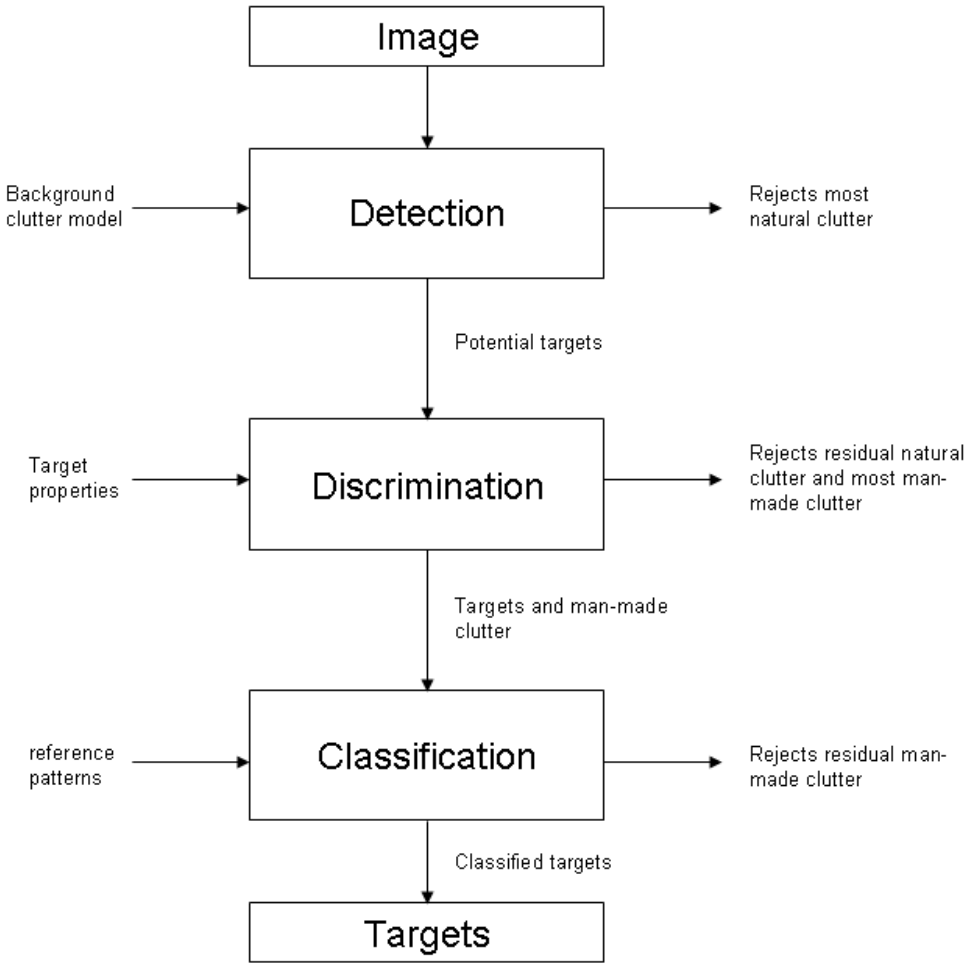


Figure 1.4: *SAR/ATR processing scheme. The model is also applied to many classification-recognition tasks.*

- *Diameter* is the length of the diagonal of the smallest rectangle that encloses P.
- *Rotational inertia* is the normalized second mechanical moment about the centre of the mass of P.
- *Peak CFAR* is the maximum value of the pixels within P.
- *Mean CFAR* is the average value of the pixels within P.
- *Percent bright CFAR* is the percentage of pixels in P that exceed a certain value.
- *Standard deviation* is the standard deviation of the pixel values in the Target-Sized Region T.
- *Ranked fill ratio* is the percentage of the power contained in the brightest 5% of the pixel in T.
- *Fractal dimension*, as described in [3]

The outcome of the discrimination stage results in a set of relatively small image regions, termed chips, which are likely to contain targets. As a consequence, they are fed into the classifiers, which perform the classification step. In the literature many classifiers have been extensively analysed, such as mean square error, (**MSE**) [4], or Hidden Markov Model, (**HMM**) [5], for instance. Classifiers indeed combine the potential target features and as result they label the object under investigation.

1.2 Overview of applications

1.2.1 Introduction

ATR systems are widely applied in research and in industry. In this section an introduction of the most important application of the classifier problem will be given. As mentioned, classification is a very common task, therefore a complete description of applications which use ATR algorithms is unfeasible. In the next subsections an introduction of the sectors, where classification algorithms can be applied, will be presented.

1.2.2 Defence systems

ATR is applied in many defence systems. Indeed surveillance and recognition which are the most important tasks for every defence critical system can be considered as classification problems between friends and foes and noise as well as clutter.

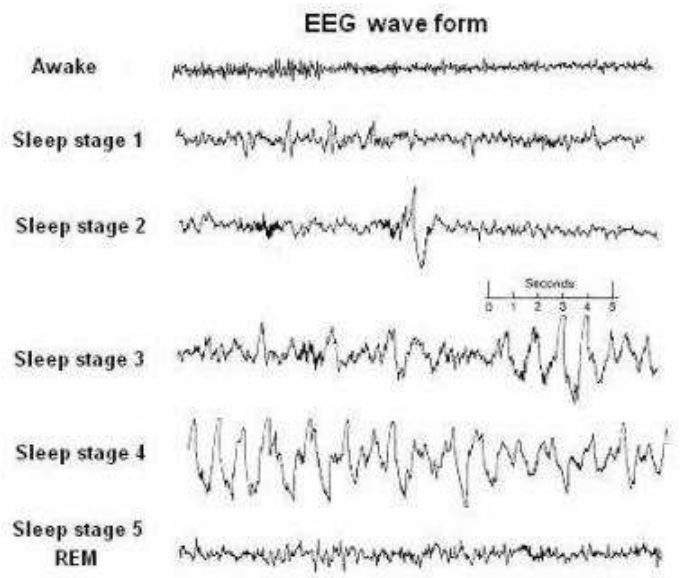
Historically target recognition has been a common task in military applications. Uniforms, flags, standard etc. are the most common signs used by armies to divide friends from foes. From WWII however with the introductions of new technological equipments (photography, Electro-Optic systems used for remote sensing) some specialized people were hired and trained by armies to detect what kind of targets were present in the Battle-space. Nowadays however the huge amount of data recorded by more and more sensors makes human-based target recognition unfeasible, therefore an automatic, software- or hardware-based, approach is necessary. Hence, generally

speaking, a sensor is adopted to acquire information from the scene and a electronic 'brain' (i.e. a processor) is 'trained' in order to understand if a threat is present in the observed scene. From a mathematical point of view, target recognition can be considered as follows: First a sensor acquires a finite numbers of feature of a potential target (e.g.Electro-Optical Cameras are able to detect properties of the objects such as colour, the movements etc., whereas radar systems are able to detect only the presence of a target and possibly velocity. Nevertheless both classes of sensors are not able to give information about the mass of the potential targets). Once that potential targets features are detected, they are processed in order to separate true targets from the rest (i.e. noise, man-made clutter, decoys, etc.) and finally a classification is performed (i.e. what kind of target is present in the observed scene). Compared with well trained personnel, automatic target recognition is thereby faster.

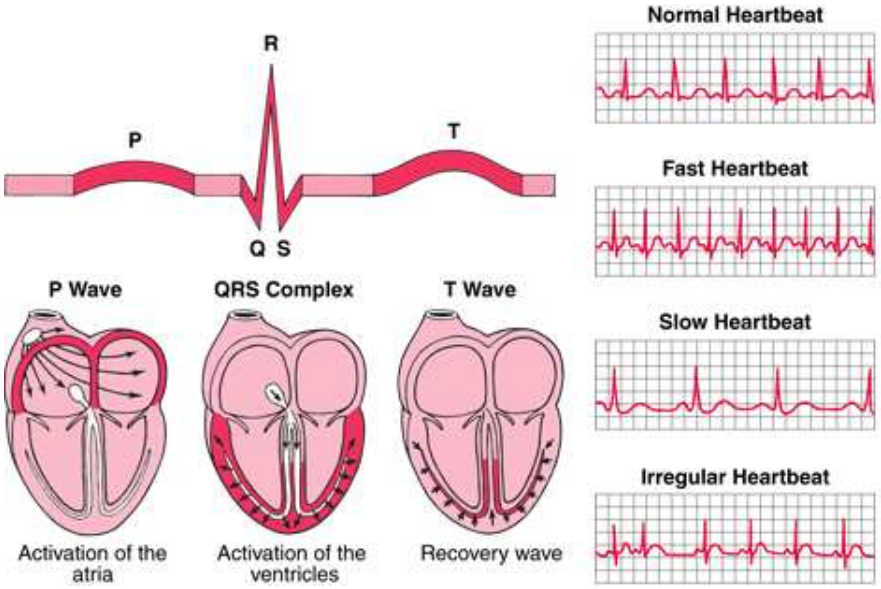
Synthetic Aperture Radar images are considered as the input of this thesis, nevertheless other kinds of sensors such as LIDAR, Electro-Optic systems, etc., can also be considered as input sensors.

1.2.3 Biomedical Application

Biomedical applications are another very interesting sector of application of Automatic target recognition. Lupo *et al.* in [6] summarized the proceedings of the workshop 'Tanks to Tumors' with the purpose of exploring means for exploiting the technological opportunities in the integration of image pro-



(a) EEG



(b) ECG

Figure 1.5: *Biomedical signals classifications*

cessing, data-base management and development as well as infrared sensor technology for early detection of breast cancer. Another example of the application of classifiers in biomedical applications can be related with classification of electroencephalograms (**EEG**), as depicted in Figure 1.5(a), and electrocardiogram (**ECG**) signals, shown in Figure 1.5(b), but also ultrasound signals as well as ecography (Figure 1.6), mammography (Figure 1.7), etc.. A very important difference between biomedical applications and military systems consists of the trade-off between probability of detection and probability of false alarm. As for military systems indeed strict requirements have to be adopted for both probabilities, whereas biomedical application, e.g. early detection of breast cancer, the requirements on P_{fa} (i.e. probability of False Alarm) are weaker, whereas specification on P_d (probability of detection) are more restrictive (i.e. if an error type II is performed, further analysis can clarify the presence of the cancer, whereas a mis-detection of a cancer has worse consequences).

1.2.4 Computer science

In computer science a scheduling algorithm (as shown in Figure 1.8) is the method by which threads, processes or data flows are given access to some resource systems. The scheduling usually performed for load balancing (i.e. the methodology to distribute workload across multiple computers) or achieve a target quality of service. An Operating System (**OS**) usually classifies a process according to some parameters such as priority, throughput (i.e. number of processes that complete their execution per time unit), latency (total time

Figure 1.6: E*cography example. Using ultrasound a non-ionizing and non-invasive investigation can be performed. and an image can be created. By detecting some image properties it is possible to determine the sex of the baby or some malformations.*

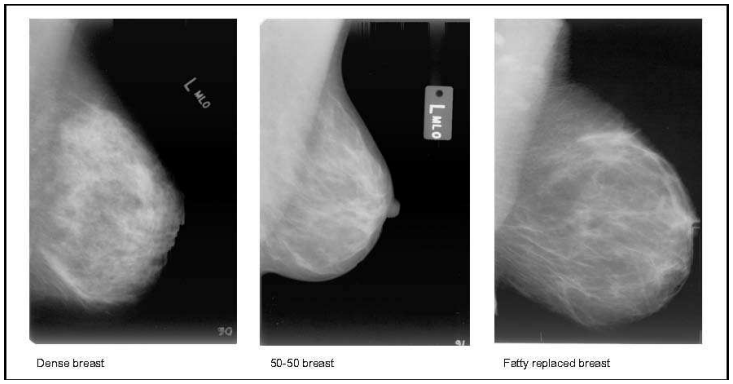


Figure 1.7: *Mammography examples. In this case a tumor is not present, nevertheless it is important to discriminate the presence of calcification from the background in a not so bright images.*

between submission of a process and its completion). In practice these goals often conflict, therefore a compromise is necessary. As a consequence of this there are three distinct types of schedulers: *Long-term*, *medium-term* and *short-term* and the names suggest the relative frequency with which these functions are performed.

The main purposes of scheduling algorithms are to minimize resource starvation (i.e. avoiding deadlocks: a process is perpetually denied necessary resources and it cannot be able to finish its task) and ensure fairness (i.e. equal CPU time to each process) among parties utilizing resources. Hence the scheduler deals with problem of deciding which of the outstanding requests is to be allocated resources.

There are several kind of scheduling algorithms and the most common can

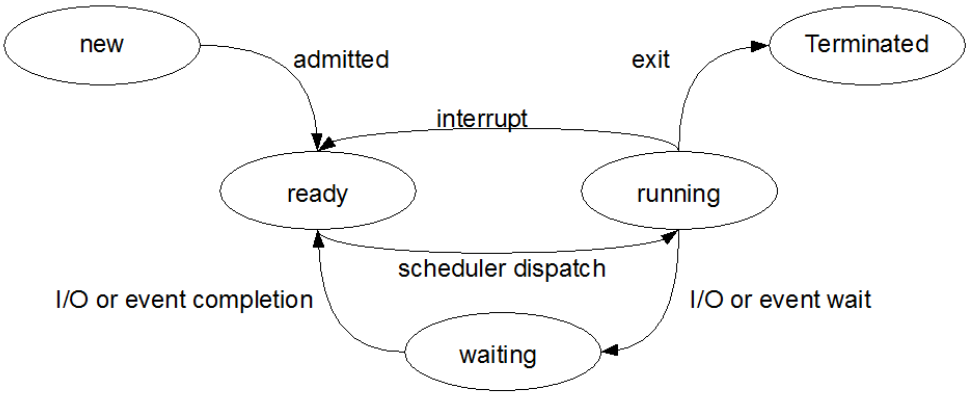


Figure 1.8: *Scheduler scheme*

be summarised as follows:

1. First In First Out (**FIFO**), it the simplest scheduling algorithm. As clear from name, it queues processes in order that they arrive in the

ready queue.

2. Shortest Remaining Time (**SRT** or **SJF** Short Job First), the scheduler arranges processes with the least estimated time remaining to be next in the queue. An advance knowledge or estimations about the time required for a process is necessary.

3. Fixed Priority Pre-emptive Scheduling (**FPPS**), the OS assigns a fixed priority rank to every process.

4. Round Robin. In this case a fixed time unit per process is fixed and the OS cycles the processes.

5. Multilevel Queue Scheduling. This is used for situations in which processes are easily divided into different groups.

How to choose a scheduling algorithm is one the most important issue in designing a OS. Unfortunately, similarly to SAR/ATR system, there is no universal criterion which is able to define which the best scheduling algorithm is. Indeed by using a statistical model for processes, it is possible to compare the performance of the schedulers in terms of ATR performance analysis.

1.2.5 Finance

Another result of WWII was *Operational Research*, which originated in the efforts of military planners. After the war, the techniques (which consist of statistics, optimization, probability theory, game theory, graph theory, decision analysis, mathematical modelling and simulation) began to be applied widely to problems in business, industry and society.

During WWII operational research was defined as 'a scientific method of providing departments with a quantitative basis for decisions regarding the operations under their control', and nowadays it widely used in following areas:

1. Critical path analysis in project management, identifying processes in a complex project which affect the overall duration of the project.

2. Network optimization in telecommunication (in order to maintain quality of service during outages).

3. Routing, find the best route for Internet packet or determining the routes of buses so that as few buses are needed as possible.

4. Automation/Robotics: Operational research is greatly important for Robotics system design process.

5. Search theory (microeconomics), it studies buyers or sellers who cannot instantly find a trading partner, and must therefore search for a partner prior to transacting.

Decision making is also crucial in business company strategies, in order to determine the most profitable set of actions. Quantitative analysis is a branch of finance which uses numerical or quantitative techniques; systematic empirical investigation of quantitative properties or phenomena and their relationships. Their target is to develop a mathematical model, a theory and/or hypothesis pertaining to observed phenomena. Quantitative analysis is often related to risk management, which is the identification, assessment and prioritization of risks and also to investment management as well as derivatives

pricing.

A method very common to measure the risk in finance is Value at Risk (**VaR**) (see Figure 1.9), defined in [7] and adopted in order to quantify the risk of loss on a specific portfolio of financial assets. In practice VaR defines the maximum potential loss of a portfolio of financial instruments with a given probability (i.e. confidence interval) over a certain horizon.

Despite the efforts of the researchers many issues on how to select the best

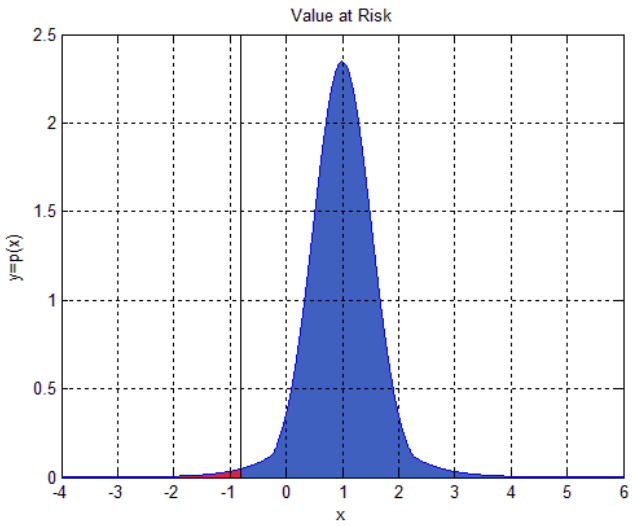


Figure 1.9: *VaR example. An interesting problem is when two similar portfolio have the same behaviour (i.e. they are overlapped) almost everywhere, and how to determine which portfolio produces the minimum risk?. The curve represents the hypothetical Profit-to-Loss probability (it has mean and standard deviation of unity). The blue area represents the* 95% *of total area of the curve, whreas the red one to the left of the black line represents the* 5% *of the total area under the curve. VaR is defined as a threshold value (i.e. the black line) such that the probability that the mark-to-market loss (i.e. A loss generated through an accounting entry rather than the actual sale of a security) on the portfolio over the given time horizon exceeds the threshold value.*

portfolio are still unsolved. An classical alternative to VaR is the Modern

Portfolio Theory [8], which models the return of an asset as an elliptically distributed random variable, therefore the selection of the optimal portfolio could be considered as a class-separability problem.

In conclusion all of the mentioned aspects of the financial markets can be considered as decision/classification problems and the results of this thesis could be helpful in order to estimate the performances of a financial product.

1.3 Hypothesis

The investigations have been performed in order to define a criterion and framework to allow the comparison of the performance of ATR systems. The studies focused on Synthetic Aperture Radar (SAR) ATR systems (**SAR/ ATR**), but the adopted criteria can be suitable for many different kinds of sensors.

For our purposes, the most important processing tasks of an SAR/ATR system have been simulated, in order to test and explore the comparison framework.

Despite ATR systems having been introduced several decades ago, several issues have not been resolved yet. Particularly the problem of predicting the separability of potential target classes at what degree (i.e. the percentage of correct classification of a class with respect to the others) is still unresolved. Class-separability and the degree of class-separability are furthermore strictly related to the information content variation through the ATR system processing chain, therefore it is important to understand how ATR subsystem parameter variation can affect the system performance in terms of class sep-

arability. Hence, summarizing, we are interested in *'how to quantify the information content and its variation through the ATR processing chain and to estimate the impact on the class-separability (i.e. assessing the degree of separability)'.*

1.4 Contribution of this study

As mentioned earlier the investigation is aimed at defining a criterion and framework to compare the performance of ATR systems. As a consequence several aspects of SAR/ATR systems have been investigated and the most important innovations will be reported in the next subsections.

1.4.1 Advances in class-separability analysis

This thesis considers the problem of information flow and ATR separability. Chapter 3 explains how the information content varies through the processing chain. In Chapters 6 and 8 several SAR subsystems have been considered and their parameters analysed. As a result the understanding of how the parameters affect the system outcomes has been obtained. Furthermore the optimal choice of subsystem parameters which maximize/preserve the information content was investigated and therefore lead to the best possible classifier performance. Some problems related to image quality metrics and their information content were considered, in particular the limits of the current most popular image quality metrics have been analysed and new denoising-metrics have been introduced. Moreover the importance of edge-preservation

in ATR systems has been also investigated. In this case some studies on the most popular edge preserving metrics have been performed and the results have been analysed.

As for class-separability, new contributions have been introduced such as a mathematical definition for class-separability and for the degree of separability. The metrics analysis for assessing the degree of separability between classes and the most important required properties of a metric for estimation of the degree of class-separability have been also performed. As consequence a new metric for assessing the degree of class-separability has been defined and tested (see Chapter 4 and 5).

1.4.2 Radar Cross Section data modelling simulation

Another important contribution of this thesis is SAR image generation testing despeckling algorithms. Data modelling in SAR systems plays a very important role, therefore the generation of synthetic SAR images could help in terms of estimating system analysis. In Chapter 6 the simulator is described and the most important limits of current simulation techniques for K-distributed Clutter modelling are reported as well as their possible solutions.

In particular Chapter 6 describes how Ward and Oliver's K-distributed clutter simulation techniques ([9] and [10]) fail when they are applied to simulate SAR images because they are not able to preserve the statistical properties of clutter RCS modelling (i.e. the mean value of image intensity is not equal to the mean power of the underlying RCS as expected) were investigated.

The results obtained by using Ward's modified technique for SAR simulated clutter images allowed us to generate an understanding of how to overcome the above-mentioned problems.

Please note also that the studies on clutter modelling have produced some useful considerations, well-known in radar community, about the asymptotic values of statistical clutter PDFs, which could be used for CFAR applications. In section 8.3 it will be reported that Weibull model can be handled easily compared with K-distribution clutter model because it can be considered as an upper/lower bound of K-distribution for sea and land clutter respectively.

1.4.3 Extensions of techniques for ATR parameters analysis

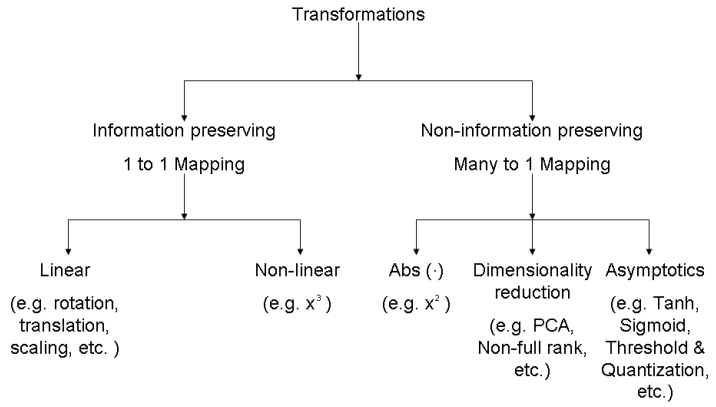


Figure 1.10: *Transformations classification: how operations in a processing chain can be classified. How the information is preserved, distorted or deleted by a processing subsystem.*

In order to determine the optimal choice of SAR/ATR parameters (for each SAR/ATR subsystem, classified as reported in Figure 1.10) and in order to understand how to quantify the information variation when sampled data pass through an ATR subsystem, several approaches have been considered and their limits analysed in order to determine the optimal procedure for assessing the information variation content in the SAR/ATR processing chain. In Chapter 3 the several methods for assessing how the information content changes through the processing chain will be described. In practice the most interesting approaches for analysing data transforms are based on differential geometry and Lie Groups and the Unscented Theory (**UT**)) (the UT reported in Section 3.10.1 can be considered a promising approximation of methods based on differential geometry, because it works properly with the sampled data used to analyse signal processing and classifier systems.
Moreover the experiments performed allowed us to create a taxonomy for the data transformations usually adopted in a SAR/ATR system. As shown in Figure 1.10 transformations can be divided into two groups: Information preserving and non-information preserving transformations. The first group consists mainly of linear and non-linear 1-to-1 mapping (e.g. rotation, translation, scaling functions as well as Fast Fourier Transform (**FFT**) or sign preserving function, i.e. y=x^3). The second group however consists of functions which may perform a many-to-1 mapping such as:

- non-sign preserving function, i.e. y=x^2, $abs(\cdot)$, etc.;
- Dimensionality reduction: PCA , ICA or non-full rank transformations, etc.;

- Asymptotic: tanh, thresholding/quantization (e.g. detection), etc.

In conclusion, a transformation can be considered information preserving if it is a member of the first group which consists of invertible transformations, whereas the second group are non-information preserving and consists of non-invertible functions, which lose information.

1.5 Compact thesis outline

In this section the structure of the thesis is reported.
In Chapter 2 a review ATR is be reported. In the chapter the most important SAR/ATR applications are discussed as well as the literature review is posted. The most significant experiments performed during the last decades are described in order to stress the most important unsolved problems in SAR/ATR technology. Moreover the literature review about ATR system performance analysis is discussed too.
In Chapter 3 the theories of the topics discussed in this thesis are described, in particular the principle of Bayes theory and pattern recognition theory will be described. Moreover the introduction of the principle of differential geometry, fractal geometry and information preserving analysis techniques are reported.
The comparison of the most popular class-separability metrics is the main topic of Chapter 4, where the most common methods adopted by researchers for assessing the degree of class-separability are reported. Moreover in the Chapter a new metric for the estimation of the degree of class-separability is

introduced. In practice we are interested in defining a global criterion which is able to predict if two different configurations of a SAR system produce an equivalent error of classification or not. By analysing the processing and classification chain, we would also understand which SAR/ATR system produces the best results.

Chapter 5 is focused on the performance of the new proposed metric for assessing the degree of class-separability when the sample size varies.

Chapter 6 is addressed to test and analyse a set of metrics commonly used to compare denoising algorithm performance in SAR systems in order to check if they are able to satisfy all SAR/ATR despeckling requirements (i.e. removing noise and preserve image features).

Chapter 7 is concerned with the analysis of despeckling algorithms when they are applied to ATR problems.

In Chapter 8 another case study is considered. A Clutter model and its information content is analysed in order to understand how the information changes at the detection step, which is a non-linear function.

In Chapter 9 A summary of previous Chapters as well as thesis conclusion are reported.

1.6 Chapter Summary

In this Chapter the following topics have been discussed: First an introduction to classification problems has been given, then the thesis structure has been introduced.

As for ATR system applications in the SAR domain, firstly a brief description of the most advanced systems present in literature has been reported, secondly a literature review of the most common methods adopted by researchers for estimating ATR performance has been developed. Beside an analysis of the most important deficiencies of the described methods has been proposed as well.

Other problems related with the ATR performance problem, such as the dimensionality reduction problem, feature extraction etc. have been introduced by describing several sectors where classification is widely applied.

The problems introduced have allowed us to determine some problems which need to be investigated. Indeed in the most popular ATR problems no global critera to access classifier performance is suitable, therefore there is no method to compare two systems in order to quantify their performances. In practice the problem can be stated in the following way: given two different systems (e.g two different SAR systems, two different threads or processes, two portfolios, two Mammography systems), which one produces the better results? Does the best method actually achieve the maximum performance that is possible given the sensor to provide the source data? In practice, hence, it is interesting to find, if there exits, a global metric which is able to predict the degree of separability of two classes.

Despite the researchers' efforts have increased during the last decades, few have investigated in this direction, therefore no global method for the ATR systems have been adopted and several times, naive and holistic/heuristic methods have been adopted which are non-efficient [11]. Moreover a set of problems/ambiguities is still present in the ATR community:

1. There is no definition of degree of separability, but there exist many methods which try to measure it.

2. Dimensionality reduction of the problem feature space, which introduces more ambiguity in the classification step (i.e. loss of information in practice).

3. How to perform the optimal feature extraction for a SAR/ATR system, i.e. which are the most important parameters which affect the classification performance.

These are the most important issues which this thesis try to answer in the next Chapters.

Chapter 2

Overview of ATR problems

2.1 Introduction

This chapter introduces the most recent research in ATR systems, in order to understand better the still open issues of the ATR problem (e.g. how to design the signal processing chain through to class decision; how to compare processing chain and classification stages solely; how to define maximum possible performance; etc.). First a literature survey on ATR technologies will be given, then a literature review on class-separability in ATR systems will be reported.

2.2 Overview of ATR technologies

As described in Section 1.1 ATR systems are widely adopted by researchers in order to make automated decisions. Face recognition, breast cancer early detection and Melanoma discrimination are just some examples of ATR ap-

plications. An interesting introduction on image classification and the most important techniques for improving classification performance are given by Lu and Weng in [12], whereas Zelnio in [13] gives a brief introduction on ATR systems and examines the issues associated with evaluation of complex decision-making systems by focusing on issues that surface in ATR systems. A brief description of still open issues and obtained results of research are given by Ratches in [14] (such as comparinson of classification algorithms, processors and evaluation techniques).

Novak *et al.* in [4] , in [15], [22], [16], [17], [18], [19], [20] and [21] introduced the main characteristics of SIAP metrics, which supports new sensor platforms gathering wide area SAR stripmap and SAR spotlight imagery. They evaluated the performance and summarized the results of several classifiers based on MSE (i.e. matching classifier) using imagery of 18 distinct targets contained in the MSTAR data set. Moreover Novak studied the effects of the processing on the performance of a classifier in a quantitative way also and no generalized method is defined for the estimation of the degree of class-separability. One important conclusion remarked by their initial studies was the ability to correct classify the independent tank targets (e.g. T72) depended strongly on the training set. Beside they demonstrated that interclass variability is a very important issue for matching classifier design. Novak compared classifier results in terms of a Confusion Matrix [23]. Similar considerations can be reported for Karl's work in [24] where the effects of feature enhancement of images with respect to several classifiers are reported (i.e. they presented an analysis demonstrating the impact of a non-quadratic optimization-based SAR image formation technique on feature enhancement

and ATR performance).

Novak was not the only researcher who investigated matching methods for object classification in SAR images. Hummel in [144] summarized the progress made in the course of the feature matching approach from the years 1995 to 2000 whereas Washburn et *al* in [145] focused on the description of the optimal design parameters of the search algorithms for model-based SAR/ATR. Bhanu ([25] and [26]) used however the model-base matching techniques for recognition of articulated objects and showed that articulation-invariant features can be used successfully for classifying articulated objects (actual and occluded).

Ettinger et *al.* in [27] (as well as Keydel in [28]) developed a probabilistic optimization approach, based on a matching algorithm, in order to solve a challenging issue in model-based approaches: The difficulty in generating accurate prediction of an electromagnetic signature and its variation in operating sensing conditions.

Wolfson in [29] described the advantages of using a matching method based on geometric hashing, especially in terms of data structures, because it is inherently parallel, therefore fast to process.

Chiang *et al.* in [30] and [31] presented a Bayesian formalism for model based classification and they investigated how parameters can affect classifier performance in terms of average probability of correct classification as well as average probability of error.

Fukunaga in [32], [33] and [34] reports several methods (i.e. bootstrap, hold-out, leave-one-out, resubstitution and nonparametric methods) which can be used to compare the classifiers performance and he also provided an analyt-

ical investigation of the afore-mentioned criteria.

Mitchell *et al.* in [146] also considered model based ATR classifiers, but they focused on the impact of various compression rates of model-catalog system, i.e. target/signature database, (via signal vector quantization) on the classification performance, also in the case of scenario variation (i.e. presence of noise and occlusion) for high-range resolution (**HRR**) and SAR data.

Daniell *et al.* in [35] introduced sub-band correlation filters (i.e. matched spatial filters) as a feature-matching classifier.

Brown however in [147] examined the aspect dependence of SAR target classification and developed a Bayesian classification approach that exploits multiple incoherent views of a target, based on a Maximum Likelihood Classifier.

Suvorova *et al.* in [36] considered the Karhunen-Loeve transform with invariance for an ATR classifier, Zhao *et al.* in [37] and Nilubol *et al.* [5] suggested to use a different approach for the classifer. Indeed Zhao *et al.* adopted the Support Vector Machine (**SVM**) for SAR/ATR, whereas Nilubol *et al.* decided to use the HMM for feature matching.

Finally Jain *et al.* in [38] evaluated a large number of algorithms for the selection of feature subset for SAR/ATR systems and they show that sequential forward floating selection [39] dominates the other algorithms which were tested (e.g. Deterministic Solution, Genetic Algorithms, Neural Networks etc.). Moreover they investigated on the danger of using feature selection in small sample size situations.

O'Sullivan *et al.* in [40] and [41] decided to model the SAR signal with a Gaussian model and used an approximated Bayes classifier.

As for Mahalanobis' approach, [42] a multi-class SAR/ATR has been per-

formed by using shift-invariant correlation filters.

Finally Bahnu *et al.* in [43] adopted an adaptive approach for ATR by performing a feedback loop depending on Probability of Correct Identification (**PCI**) and Probability of False Alarm (P_{fa}) requirements.

Please note that in this thesis synthetic data are used, as MSTAR data are collected by sensors which are classified, therefore they are not suitable for some experiments (e.g. despeckling algorithms analysis). Further information about simulated data will be reported in section 6.4.

2.3 ATR features review

As mentioned in Chapter 1 several features are usually adopted to perform the best classification, nevertheless this number is usually computed from some geometrical and statistical properties of objects present in the scene. This section gives a brief introduction of the most important investigations performed by researchers in order to compute and understand how the parameters such as RCS of potential targets can be modeled in the optimal way to maximize the performance of the classifier.

The studies can be grouped in four sets based on the idea which the researchers considered. The first set consists of investigations into RCS and scattering modeling; The second set is characterized by investigations performed on scale models of potential targets, whereas the third model consists of investigations on statistical target behaviour in SAR images. Finally the last set is concerning with physical properties of targets for high-resolution

radars.

As for the first group of investigations, Rosario in [44] investigated the effects of modifications in target signature in SAR images and he suggested considering the target in terms of a spatially decomposed mode instead of a single pixel because this approach provides some advantages in terms of feature/signature stability instead of single pixel RCS modeling approaches. Krogager in [45] however proposed a new decomposition of the complex radar target scattering matrix in three components that provides a clearer picture of the physical mechanisms behind the scattering and as a consequence a clearer picture of the target itself.

Turner and Gerry *et al.* in [46] and [47] respectively considered the modelling of RCS for its prediction. In the second group of articles the signature of scale model RCS was investigated by using scaled models of targets of interest as reported in [48], [49] and [50], whereas Blacknell (i.e. third group) in [51] analysed the statistical behaviour of targets in synthetic aperture radar images.

The most interesting approach for signature modelling and as a consequence for feature extraction was introduced by Potter in [52] and Bhanu *et al.* in [53]-[57], where the approach of scattering centers has been analysed. The main advantage of this kind of analyses are that the data generated in high frequency radar measurements by returns from isolated scattering centre such as corners and flat plates are almost invariant to articulations and to small changes of SAR parameters.

2.4 ATR performances review

An introduction of the most common methods adopted by researchers in order to estimate the classification accuracy is reported by Stehnam and Congalton in [58], [59] (e.g. Error Matrices (i.e. Confusion Matrices), Class-level accuracy measures etc.) and [60].

A very important aspect of the problem which has been partially considered as well as investigated is to define the information content carried by sensors in order to have good ATR performance and as a consequence defining the information loss in all of the steps of a signal processing scheme as well as in problem modelling. In order to solve the afore-mentioned problem, a classical information theory approach has been adopted, as used by Horne in [61] (i.e. adopting entropy, the amount of information obtained when the outcome is observed), Briles [62] and Zelnio *et al.* in [11] and [63] respectively. Briles extended the rate-distortion function, used to compare data compression algorithms, to the Bayes rate-distortion function in order to associate the Bayes risk to the distortion so that information-theoretic tools can be adopted for statistical identification problems. Zelnio however tends to divide the approach for estimating ATR system performances in two groups:

1. Techniques based on Pattern Space, which include the information theory approach;

2. Extrapolation Techniques.

As for the techniques of the first group, they tend to address the relationship of representations in object space to the corresponding representation in

pattern space or signal observation space. As a consequence, any 'change' in object space is related to a change 'in distance' between the observed signal patterns so that a probability density function for the 'distances' can be introduced to evaluate the performance of the system. In Techniques based on Pattern Space approach, a very important role is played by some parameters which affect the performance of the classifier heavily, termed *variants*. Hence, in Techniques based on Pattern Space each target exemplar is associated with a 'noise sphere' in pattern space with radius proportionally larger to account for ambiguities. Alternatively the same variations might be viewed as target variants as well so that each target occupies a larger pattern subspace, no longer spherical. Extrapolation techniques however are based on statistical inference. Zelnio reports that the most important difficulty in applying these techniques of ATR system prediction lies in determining the 'rate of growth' of pattern space as more targets are added to the set.

Finally Zelnio remarks that techniques based on Shannon's information theory do not capture the essence of the target recognition problem, i.e. how the SAR system parameters affect the classifiers.

Takkola in [64] used Mutual Information (**MI**), which describes the amount of information that one random variable gives about a second random variable. Assume a random variable Y, $y_i \in \mathbb{R}^d$ representing feature vectors, and a discrete-valued random variable C representing the class labels, with samples as pairs $\{y_i, c\}$. From a Shannon's theory point of view, drawing one sample of Y at random, the entropy or uncertainty of the class label, making

use of Shannon's definition, is expressed in terms of class prior probabilities:

$$H(C) = -\sum_{c} P(c)\log(P(c)). \tag{2.1}$$

After having observed a feature vector y, the uncertainty of the class identity is now the conditional entropy:

$$H(C|Y) = -\int_{y} p(y)\left(\sum_{c} p(c|y)\log(p(c|y))\right)dy. \tag{2.2}$$

The amount by which the class uncertainty is reduced, after having observed the feature vector y, is the MI , $I(C,Y) = H(C) - H(C|Y)$, which can be written as:

$$I(C,Y) = \sum_{c}\int_{y} p(c,y)\log\frac{p(c,y)}{P(c)p(y)}dy \tag{2.3}$$

after applying the identities $p(c,y) = p(c|y)p(y)$ and $P(c) = \int_y p(c,y)dy$. It equals zero when $p(c,y) = P(c)p(y)$, that is, when the joint density of C and Y can be factored as a product of marginal densities, which is the condition for independence.

The MI between class labels and transformed features is used as a criterion for a method of learning discriminative feature transforms. Miller also in [65] considered MI as information measurement, because independent of the recognition system, in order to quantify both information gain due to remote observation of the scene and the information loss due to signature variability, the model mismatch can be quantitatively examined using the Kullback-Leibler divergence. Kanaya however in [66] introduces a mathematical function which conjoins the two key concepts of mutual information

and Bayes risk and then some asymptotic theorems that verify an important implication in the context of practical Bayesian decision-making are proven. A different approach was considered by Miller in [67], who investigated the information theoretic bounds of ATR performance. He adopted a statistical approach because it provides a systematic framework for integrating prior knowledge about the scene and targets and for fusing information from multiple sensors by application of basic principles of statistical inference. He adopted *Chernoff* and *Kullback-Leibler* distances ([32],[68]) to quantify separability, because the Chernoff distance provides upper bounds and asymptotic expressions for the *probability of miss* (P_{miss}), *probability of false alarm*(P_{fa}) and *probability of error* (P_e) in detection problems, whereas Kullback-Leibler metric provides an expression for P_{miss} for a fixed, small value P_{fa}. This interesting approach does not satisfy same requirements such as the quantification of information carried by sensor signals (i.e. target features analysis) and does not provide any means to estimate information loss along a signal processing scheme.

Another method for predicting an upper bound of ATR performance was described and analysed by Boshra and Bhanu in [69] and [70]. It is also characterized by a statistical approach because the features obtained by processing sensor data are corrupted by distortion factors such as uncertainty, occlusion and clutter. Unlike Miller's investigation, Boshra was interested in determining an upper bound on the *probability of correct recognition*, **PRC**, of a given model view in the presence of data distortion. They thereby defined a two step method, which works as follows:

1. By defining a Data-Distortion Model (i.e. a PDF of distortion for oc-

clusion, clutter and uncertainty), Model Objects and a Transformation Class (i.e. rotation/translation) the object similarity coefficient is computed;

2. The computation of performance bound is computed by using object similarity coefficient and the Data-Distortion-Model.

Other researchers, such as Grenander in [72] presented a method to predict fundamental performance of template-based ATR for a given image noise model using the Hilbert-Schmidt bound, which is characterized by performing a lower bound on the error of an other estimator. However Irving in [73] described a formal method for predicting performance of SAR target detection, based on statistical modelling of both data distortion factors and model target views. Despite their results, both the approaches used in [72] and in [73] are not suitable for our purpose because they are focused only on classification algorithm performance and not on the parameters which affect the performance.

Bhanu *et al.* in [74] however introduce a new criterion of prediction of performance of an ATR system. It consists in adopting the Hausdorff distance measure in order to estimate correctly the classification region borders. It works as follows: first the system detects the scatter centers and the Major axis is computed, then a matching model with a simulated target (based on CAD examples) is performed and the outcome is sent to a Rank-order filter using Haussdorf distance Measure. As a result a set of Rank-Ordered Hypotheses (**ROH**) which are used for the recognition is obtained.

Finally other methods such as Receiver Operating Curve (**ROC**) [75] and its

area [1], Confusion matrix (the percentage over the sample size of misclassified objects) [22] have also been adopted. As for Smith's works [76] and [1] the evaluation of classification is measured in terms of:

- Probability of declaration $P_d(i)$ for a particular input is defined as:

$$P_d(i) = \sum_{k=1}^{nc} P(Out = k|In = i) \tag{2.4}$$

  (nc is the number of classes)therefore the overall probability of declaration is equal to $P_d = 1/nc \sum_k P_d(k)$.

- Probability of False Alarm: the rate at which declarations for known classes are made when an input of unknown class is presented. It is defined as:

$$P_{fa} = \frac{1}{nu} \sum_{k}^{nu} \sum_{h}^{nc} P_{fa}(k, h) \tag{2.5}$$

  where $P_{fa}(i, j)$ is the probability that class C_j will be declared when the input is of class U_i (nu is the number of test sample classes).

- Probability of Generalization: the ability of the classifier to label the target correctly even when different configuration of the reference class are considered. Therefore the overall generalization performance can be estimated as:

$$P_{gen} = \frac{1}{nc'} \sum_{i}^{nc'} P_{gen}(i) \tag{2.6}$$

  where nc' is the number of the classes, whereas is $P_{gen}(i) = \sum_p P(Out = C_i|In = C_i^{(p)})$.

- Confusion matrix, reporting the number of object misclassified as a different target;
- ROC and AUC analyses (as reported in [1]) which are the graph of P_d vs P_{fa} and its underlying area respectively.

As for the ATR performance metric, a very interesting approach has been introduced by Richards *et al.* in [77] base on the principle that confidence metric has to satisfy the following requirements:

- Provide quantitative scores that are intuitive and informative in an absolute sense;
- It should be adaptable to different ATR mission and provided in real time along with the ATR declaration.

Similar investigations have been performed by Ross in [78].
Other investigation such as Asymptotic performance analysis ([79]) based on Bayesian pattern-theoretic framework and Monte Carlo prediction [80], which showed the trade-off between ATR performances and SAR resolution. Unfortunately all above mentioned methods are unable to quantify how separated two or more classes are.

2.5 ATR literature review for the general problem

In this section other ATR systems are considered and a brief literature review is reported.

2.5.1 Computational performance

A very important issue, which has not been investigated fully yet, is the problem related to computational system performance for ATR systems. DeVore *et al.* in [81] and [82] has investigated the necessary requirements of computational system performance for recognition systems performance. Moreover they have also evaluated the performance of recognition systems in terms of consumption of system resources.

2.5.2 Available ATR dataset overview

As for the generation/collection of the data which have been used in aforementioned works, two approaches can be defined: Using synthetic data or real data. As for synthetic data, there are many programs which are able to create synthetic data from CAD models. Among them, the most important are XPATCH, as reported in [56], and RESPECT, as defined in [46] and validated by Blacknell *et al.* in [83] which usually use a CAD model (as depicted in Figure 2.1) in order to predict the electromagnetic scattering of an object. As for XPATCH, it uses a CAD model in order to generate image chips of objects (articulated/non-articulated) at 360 azimuth angle (at 15° depression angle) such as tanks (e.g. T72, T80 and M1a1), SCUD missile launcher. It is able to generate models with number of facets ranging from 5,345 to 32,954. As for RESPECT, it employs a shooting-bouncing-ray high frequency physics model in order to determine the electromagnetic scattering calculations from large complex bodies.

(a) CAD model of a T72 (b) CAD model of a M1A1 (c) CAD model of a SCUD LAUNCHER

Figure 2.1: *Examples of model for XPATCH and RESPECT*

As for the dataset from real SAR images, the most important one which is available for researchers is the Moving and Stationary Target Acquisition and Recognition (**MSTAR**) database, which is described in [84]-[88]. Despite it is the biggest available database, simulated data have been preferred for two reasons: Firstly, MSTAR parameters are classified (i.e. the number of looks of the system is not available for the researchers, therefore any statistical analysis for clutter models is impossible); Secondly, no certainty on the RCS model (i.e. Gamma distribution) distribution is given.

Other authors used some real data for non-SAR images. Alexandrov, for instance, considered in [89] a set of data collected with an on-shore surveillance radar 'NAYADA-3' over the port Varna. Many images (no number is reported) of three targets were collected by the sensor. In particular the database consists of a 5,000t tanker, 13,800t and 38,000t cargo ships. Qiang in [90] however used a set of radar returns from 8 classes of ships, but no information is reported about them.

2.5.3 Other subsystems and Multi-sensor data fusion

In this section a concise introduction of the research in multi-sensor applications in ATR systems and an overview of the research of ATR techniques in HRR systems is explained.

Miller in [71] shown the simultaneous detection, tracking and recognition of objects via data fused from multiple sensors. The variability of the infinity of pose is accommodated via the actions of matrix Lie groups extending the templates to individual instances. The variability of target number and target identity is accommodated via the representation of scenes as unions of templates of varying types, with the associated group transformations of varying dimension. The remote sensing data is organized around both the coarse scale associated with detection as provided by tracking and range radars, along with fine scale associated with pose and identity supported by high-resolution optical, forward looking infrared and delay-Doppler radar images. A Bayesian approach is adopted in which prior distributions on target scenarios are constructed via dynamical models of the targets of interest. These are combined with physics-based sensor models which define conditional likelihoods for the coarse/fine scale sensor data given the underlying scene. Inference via the Bayes posterior is organized around a random sampling algorithm based on a jump-diffusion process. New objects are detected and object identities are recognized through discrete jump moves through parameter space, the algorithm exploring scenes of varying complexity as it proceeds. Between jumps, the scale and rotation group transformations are generated via continuous diffusions in order to smoothly deform templates

into individual instances of objects.

Mishra in [91] describes his first investigation on the application of bistatic SAR systems in ATR problems. In particular Mishra focuses his work on the development of algorithms for bistatic-SAR image classification of ground based targets and also examined the potential of bistatic radar for SAR classification.

As for other radar systems used for ATR, an interesting role is played by HRR systems. Williams in [92] gives a summarized introduction of the methodologies about data and algorithm, simulated performance results and recommendations for the classification of 1D HRR radar systems images.

Vespe *et al.* in [93] analysed the information content of the target signature for HRR/ATR. The classification performance is evaluated using full-scale 2D inverse SAR images obtained from a stepped-frequency chirp modulation radar system and corresponding sub-spectra of the target reflectivity function forming lower resolution images at difference centre frequencies. The classification performance as given by different combinations of RF frequencies are also evaluated and compared with the coherent reconstruction from the full bandwidth. Finally, the classification results are also computed using multiple aspects to sense the target. In this way, classification performance as function of diversity space is examined. As for the metrics adopted in order to evaluate the ATR performance, Vespe *et al.* adopted the probability of correct classification, the probability of missing a target and as a consequence the probability of declaration as well as probability of false alarm (2.5) and probability of generalization (2.6). Additionally a new measure of performance is introduced and it is termed as reliability. Reliability is defined as

the degree of trustworthiness when the declaration is made for a particular class j having input i. Thus the reliability of declaring a particular class k is given by the degree of trustworthiness when the declaration is made for a particular class k having input k ($R(k)$), hence the reliability of a classifier R_{avg} is the overall degree of the classifier trustworthiness:

$$R_{avg} = \frac{1}{n_c}\sum_{k=1}^{n_c} R(k) = \frac{1}{n_c}\sum_{k=1}^{n_c} \frac{P(out = k|in = k)}{P(out = k|in = l)} R(k) \tag{2.7}$$

where n_c is the number of classes.

The classification process typically requires a high probability of correct classification and reliability.

2.6 Summary

In this chapter an introduction of the current state of the art for ATR systems has been given. The result of the introduction of the previous works is that the information content has been usually evaluated by considering the end-to-end ATR system. Moreover no investigation on what the degree of class-separability is and how to assess it have been performed. As a consequence most of the investigations performed in order to understand how the parameters of a system can affect the performance of the final system classifiers are accurate because they are effected by the ambiguities introduced by the classifier itself. Moreover no investigation on the variation of the information content has been performed, therefore the proposed parameter settings could be unsuitable for our purposes of investigation.

Chapter 3

Theory background

3.1 Introduction

ATR systems have been studied for several decades and many different algorithms have been developed. Despite these efforts, many issues are still open such as information flow through the sensor, the definition of class-separability and the corresponding degree of class-separability, for instance. In this chapter an overview of the theoretical background of the thesis will be reported. Firstly, an introduction about the Bayes theory and Bayes classifiers is necessary in order to describe some assumptions about the investigation reported in the next chapters, then the important issues about the information flow model will be introduced and the most important problem regarding class-separability will be analysed. Secondly some methods for the analysis of information preservation will be described. In particular, the chapter is concerned with the dependency of the information flow on the sensor structure, describing and comparing some techniques which allow

assessing the measure of the variation of information content through the processing chain (i.e. how the sample set changes through the processing chain) as well as how to measure the variation of content in terms of class-separability (i.e. how to assess the degree of class-separability).

3.2 Bayes theory and Bayes classifier

Before analysing other aspects of this Ph.D problem, it is important to give an introduction of Bayes theory, as mentioned in previous sections. As described in [32] and [68], Bayes decision theory is one statistical approaches to the problem of pattern recognition. Its most important assumption is that the decision problem is posed in probabilistic terms. In a 2−class problem, for instance, the most important parameters of Bayes decision theory can be summarised as follows:

- ω denotes *the state of nature* (i.e. the tag which the samples can be labelled with) and it can assume a two values ω_1 or ω_2.
- $P(\omega_i)$ $(i = 1, 2)$ is the *a priori probability* which reflect our prior knowledge about the state of nature before it actually appears.
- $p(x|\omega_i)$ $(i = 1, 2)$ is termed *state-conditional probability density* function for x, i.e. the probability of a data value x given a state of nature.

Suppose the above-mentioned quantities are known, by using the Bayes theorem one can determine the value of another important parameter: *a posteriori*

probability, $P(\omega_j|x)$. Indeed this quantity can be defined as:

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)} \tag{3.1}$$

where:

$$p(x) = \sum_{j=1}^{2} p(x|\omega_j)P(\omega_j) \tag{3.2}$$

The corresponding decision rule is defined as:

$$P(\omega_1|x) \overset{\omega_1}{\gtrless} P(\omega_2|x) = p(x|\omega_1)P(\omega_1) \overset{\omega_1}{\gtrless} p(x|\omega_2)P(\omega_2) \tag{3.3}$$

which labels data x as ω_1 if $P(\omega_1|x) > P(\omega_2|x)$ otherwise ω_2.

Generally speaking, the decision rule 3.3 or other decision rules, produce an error of misclassification (i.e. a data x is labeled as ω_1 when it belongs to class ω_2). In order to evaluate the performance of a decision rule, it is usual to calculate the *probability of error*, i.e. the probability that a sample is assigned to the wrong class. As for a $2-$class problem it is possible to introduce an a posteriori error probability defined as:

$$P(error|x) = \begin{cases} P(\omega_1|x) & \text{if it is decided } \omega_2 \\ P(\omega_2|x) & \text{if it is decided } \omega_1 \end{cases} \tag{3.4}$$

As a consequence the average probability of error is given by:

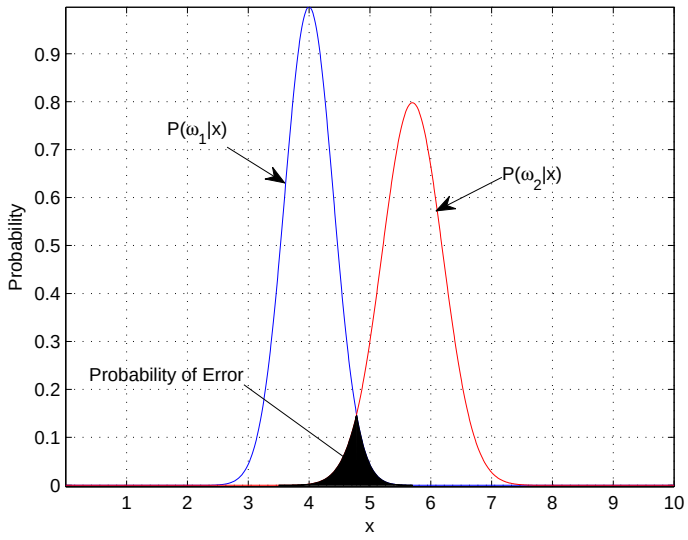
$$P(error) = \int_{-\infty}^{+\infty} P(error, x)dx = \int_{-\infty}^{+\infty} P(error|x)p(x)dx \tag{3.5}$$

therefore if for every x, $P(error|x)$ is as small as possible, the integral must be as small as possible. Bayes rule, hence, gives the smallest probability of error (as shown in Figure 3.1(a)). Indeed, if Figure 3.1(b) is considered (where $x = t$ is the value of x which satisfies the condition $p(x|\omega_1)P(\omega_1) = p(x|\omega_2)P(\omega_2)$ and t' is a bigger value), the value of error probability for $x > t'$ is equal to C, whereas for $x \leq t'$ is equal to $A + B + D$, therefore the total error area is equal to $A + B + C + D$, which is bigger than case t by D. The same conclusion are valid if the threshold is shifted to the left.

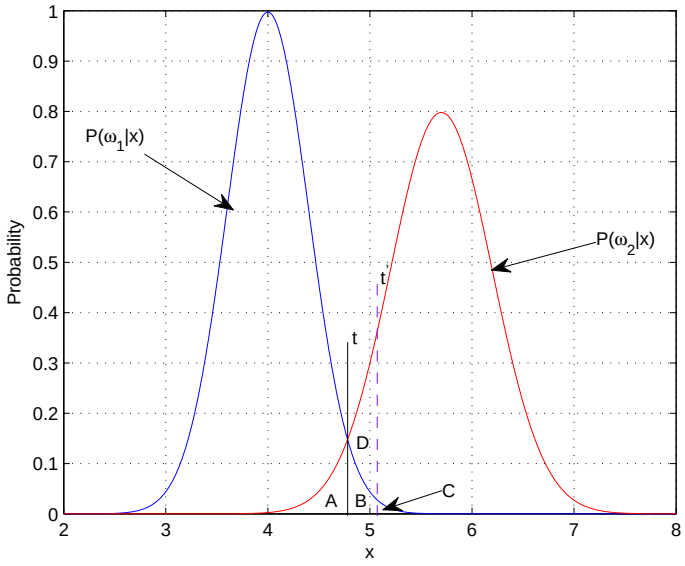
In conclusion one can state that the Bayes decision is the best rule which can be adopted because it minimizes the error of misclassification. Unfortunately this kind of classifier is not feasible in practice because it needs the a priori knowledge of classes (i.e. $P(\omega_i)$ $(i = 1, 2)$) which usually are not available, for this reason sub-optimal solutions are usually adopted. As a consequence thereby we adopted the following assumption: *'The class-separability has to be assessed before classification'* (termed *Before Classification analysis*, i.e. *BC analysis*). Despite this assumption can be considered a disputed point, we argue that the class-separability has to be performed before classification because:

- classifier performance analysis has been already successfully investigated and applied (e.g. ROC and AUC analysis, confusion matrix etc. as described in the next sections), therefore class-separability investigation would have been useless.

- The BC analysis is related only to the sensor parameters, therefore it does not consider the ambiguity introduced by a sub-optimal classifier.



(a) Error probability area



(b) Error probability area changes with two different thresholds

Figure 3.1: *Error probability area*

- BC analysis can be interesting for two reasons: first it uses a single and big set of samples (i.e. it is unnecessary to divide the samples into training and testing sets); secondly, it helps classifier performance analysis because it emphasizes the deficiencies of classifiers, helps researchers in defining better training and testing sets (i.e. in terms of feature skewness).

3.3 Information flow

Before describing the model it is important to consider the fact that a crucial problem is related to the structure of the sensor that is gathering the potential target features. Since this thesis is focused on SAR/ATR systems, the information flow is modeled on SAR systems. Some key functions (e.g. denoising, detection, classification etc.) are common to several ATR systems. As discussed in the introduction, the hypothesis of this Ph.D. thesis is related to two issues: information flow definition and class-separability in ATR systems.

For information flow, in literature ([32] and [68]) a general model of statistical pattern recognition systems can be represented as depicted in Figure 3.2. The data, gathered by a sensor, are processed in order to reduce the noise, to enhance some region of interest etc. Processed data are then fed up to the next step which extracts the most important features and reduces the samples dimensionality, so that the transformed data are finally elaborated by a classifier which separates data in two or more classes.

This model is suitable for a very wide range of pattern recognition topics of

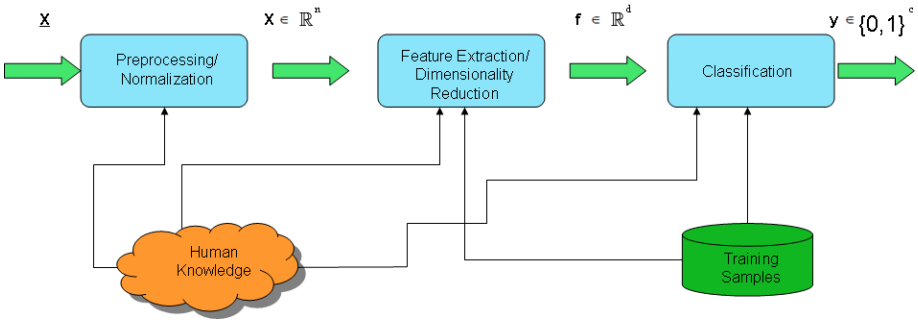


Figure 3.2: *Statistical pattern recognition model. It is a general model for ATR systems and it is also independent of the structure of the sensor.It represents an ATR systems as a dimensionality reduction problem (by using some human knowledge), which performs a binary decision at the last step.*

research, i.e. it is able to manage the most important issues of recognition systems such as denoising, dimensionality reduction and definition of classification algorithms as well as their performance analysis separately, some deficiencies are present. Firstly the generalization of the stages in Figure 3.2 tends to hide some characteristics of the analysed system which affect its performance, i.e. SAR and mammography systems have two different digital signal processing chains. Secondly, the reliability of features extracted from a set of data gathered by a sensor depends on the sensor structures self, i.e. the sensor architecture affects the information content of data and as a consequence the information content of the potential target features. Moreover some steps depicted in Figure 3.2 consist of several independent subsystems, i.e. in SAR systems, before feature extraction, detection and discrimination are necessary, as reported in Chapter 1. Hence the structure of the sensor plays an important role in defining the statistical pattern recognition model, therefore it has to be included as backbone into the information flow model.

As for the definition of the information flow model for a SAR/ATR system,

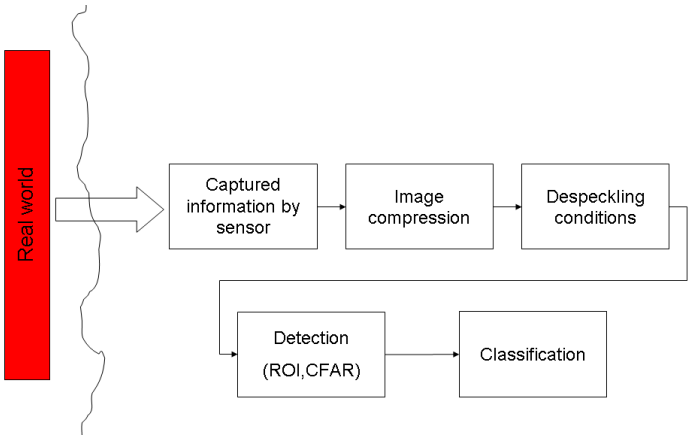


Figure 3.3: *Analysis model: a potential target can be considered as an $\infty-$dimension vector which contains all the information regarding it. Unfortunately a generic sensor is able to collect just a small number of dimensions containing information (i.e. SAR systems are able to collect information regarding length, width for instance, but they are not able to detect information such as targets colour, their temperature, etc.) therefore many object features are lost. Moreover the amount of information lost could be increased by each step reported in this picture. The goal of this model is to understand what kind of outcome is obtained when the parameters change and to find the optimal parameter set for the family of sensors under investigation.*

it has to be modelled as shown in Figure 3.3. As known indeed a SAR sensor is able to gather a finite number of the potential target features, therefore a SAR sensor selects a small number of the target characteristics which have to be also discriminated among noise, natural and man-made clutter. Unfortunately a complete feature space is generally of infinite dimensions, therefore a SAR sensor performs an unwanted dimension reduction which produces ambiguity and affects performance analysis.

Once the target signals have been gathered, they are processed and an image is obtained. In order to discriminate a potential target from noise, natural and man made clutter, the image will be processed by three algorithms: Despeckling algorithms which tend to remove the noise introduced by coherent radar signal processing; Detection and discrimination which separate the potential target features from natural and man-made clutter respectively. Finally a dimensionality reduction process is performed and a 2-class classifier is used.

The model depicted in Figure 3.3 has several important advantages. First, it emphasizes the most important steps of a SAR/ATR system. Particularly it is able to identify all the 'bottlenecks' of the information flow (i.e. points where the information content is reduced, such as at the sensor step for instance). Secondly it aids the researchers in understanding which inputs are the most suitable for each step (i.e. the optimal results that each step needs in order to produce the optimal results). The latter characteristic of the model helps researchers in understanding which information is deleted/removed during each step of processing chain.

The proposed scheme hence has been adopted for assessing performances of

SAR/ATR systems. It has been used for understanding the end-to-end information flow (*forward direction*) and to select the optimal requirements in the *backward direction*, i.e. in the forward direction it is possible to describe how the information content varies whereas in backward direction the model is able to stress what the optimal SAR/ATR parameters are necessary for a correct target classification.

3.4 ATR separability problem

Prediction of separability of potential targets classes is another issue in ATR systems, very important especially for safety critical systems such as SAR/ATR systems.

Unfortunately despite ATR systems having been adopted since the 1960's, the definition of class-separability is still an open question. In the next sections therefore the analysis of all aspects of this issue will be analysed and a literature review of the previous studies on the matter will be given.

3.4.1 Problem description

As mentioned in the previous section, the main purpose of this chapter is to investigate what class-separability and degree of separability are and how to measure the degree of separability (i.e. given that for classification applications, the probability of misclassification must be controlled). Before analysing the criteria adopted for assessing the degree of separability, it is important to emphasize the conditions which ensure class-separability. Two classes are separated if they have no intersection, thereby a necessary con-

dition for class-separability is that the error area (i.e. the overlapping area between two classes) is equal to zero. In practice the above-mentioned requirement is never satisfied (i.e. PDFs have usually a $(-\infty, +\infty)$ domain), therefore a high degree of separability when a false alarm rate associated with classification problems is considered can be reached by minimizing the Type II error area (i.e. fixed false alarm rate of a misclassified target-class with respect to another one, the highest degree of separability is reached if the maximum misclassification error of the considered class is less or equal to a fixed false alarm rate).

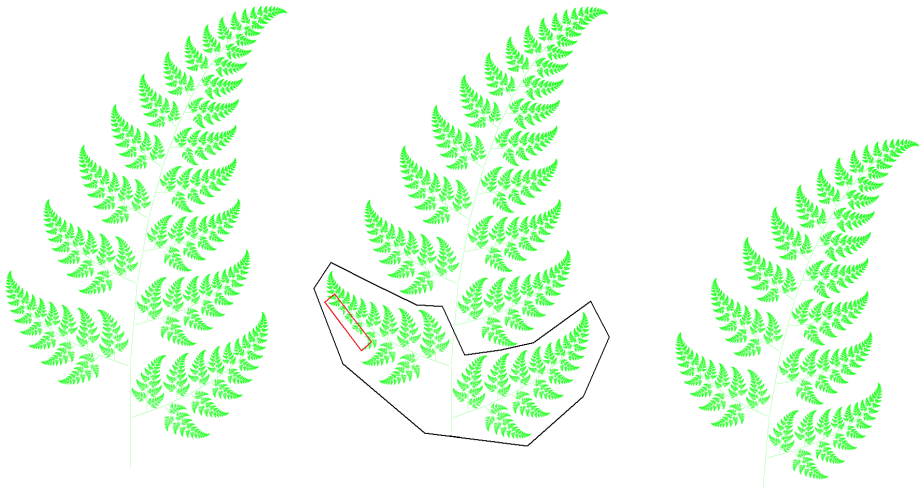
3.5 Mathematical separability

Zelnio *et al.* in [11] described several problems in ATR systems. Among them a crucial one is the proper interpretation of the variation of a particular target. They suggest to view the variations as target variants with the interpretation that each target occupies a pattern subspace, i.e. each target exemplar could be associated with a 'noise sphere' in pattern space. The mentioned model is described mathematically by considering target pattern subspaces as statistical self-affine fractal sets (i.e. each subset of a self-affine set is a perfect copy of the whole, eventually translated, scaled and rotated). In statistics, however, given a Gaussian distribution $N(\mu, \sigma)$, with mean value μ and standard deviation σ for example, if two sets of samples from this distribution are drawn and their sample mean and standard deviation are computed, the two sets have to give the same values of the original distribution, i.e. μ and σ, so that the statistical self-affininity is preserved.

Fractal assumptions indeed allow us to manage properly some ATR issues which have not been addressed yet: The definition of class-separability and degree of separability.

A fractal is an object that is self-similar, i.e. it exhibits not-exactly the same structure at all scales, but the same 'type' of structure must appear on all scales. Fractals moreover are very common objects present in nature. The fern depicted in Figure 3.4(a) is a classical example of a self-affine fractal, because each part of its leaf is a scaled, translated and rotated copy of the whole.

The world fractal has origin from Latin word *fractus*, meaning broken, and



(a) Fractal example (b) Non-separable example (c) Degree of separability

Figure 3.4: *Fractal model for classification problem: Example of fractal geometry Figure 3.4(a); Non-separable case Figure 3.4(b): a non-empty intersection can be considered as a violation of self-similarity (small rectangle); Degree of similarity Figure 3.4(c): maximum percentage of preserving self-similarity. The difference between Figure 3.4(a) and Figure 3.4(c) is the number of sub-leaves (less in the latter), i.e. the leaves in the large polygon in Figure 3.4(b) have been removed.*

was adopted mathematically by Mandelbrot [94] to try to give a definition of sets whose Hausdorff dimension is strictly greater than their topological dimension [95]. Given indeed a n-dimensional Euclidean space, $\mathbb{R}^n$, the following limit is defined as *s-dimensional Hausdorff measure of a set* $F \subset \mathbb{R}^n$:

$$H^s(F) = \lim_{\delta \to 0} H^s_\delta(F). \tag{3.6}$$

where:

$$H^s_\delta(F) = \inf \left\{ \sum_{i=1}^{+\infty} |U_i|^s : \{U_i\} \text{ is a cover of F} \right\} \text{ with } s, \delta > 0. \tag{3.7}$$

($|U| = \sup \{|x - y| : x, y \in U\}$ is the *diameter* of U). Hence one looks at the cover of F which minimizes the sum in (3.7) when $\delta \to 0$. As a consequence the Hausdorff dimension is defined formally as:

$$\dim_H F = \inf \{s : H^s(F) = 0\} = \sup \{s : H^s(F) = \infty\}. \tag{3.8}$$

In practice the Hausdorff dimension which is invariant for *bi-Lischiptz* transformation [95] generalizes the notion of dimension of a real vector space, i.e. the necessary number of independent parameters to pick out a point inside the object, and relates the dimension to the concept of a metric. Hausdorff dimension is very useful for describing fractals because this kind of object usually has an integer topological dimension, but in terms of occupied space it behaves as a higher dimensional space.

As described by Falconer in [96], a self-affine set E, which is by definition necessary self-similar (but not vice-versa), has to satisfy the following condition:

'E is self-similar if and only if the intersection $H^s(\psi_i(E) \cap \psi_j(E)) = 0$', where s is the Hausdorff dimension of E and H^s is the Hausdorff measure, whereas $\psi_i : \mathbb{R}^n \to \mathbb{R}^n$, $i = 1, \ldots, m$ are contractive mappings on $\mathbb{R}^n$ with contraction constant $r_i < 1$. Hence, if two different target classes have some samples in common, then the condition of statistical self-similarity is no-more satisfied, thereby one can state that two different target classes are fully separated if and only if they can be considered as two self affine fractals individually.

Graphically one can say that a class is separated from another if, as de-

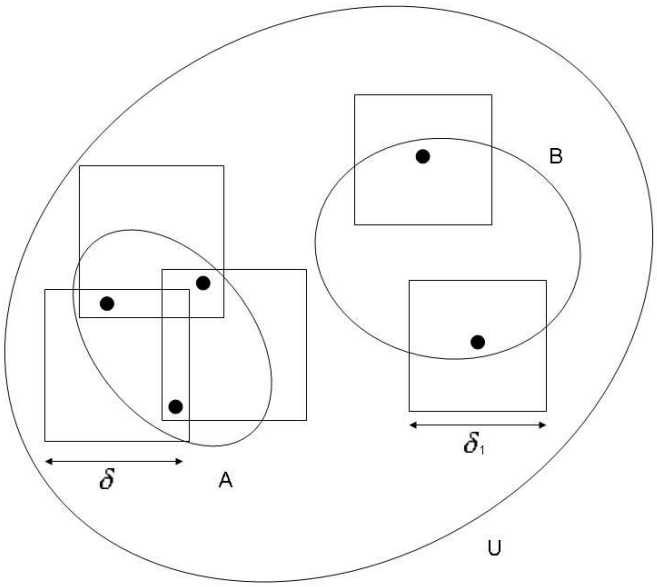


Figure 3.5: *Box counting limit example*

picted in Figure 3.4(a), all its 'leaves' are whole. However, if two classes have some common samples, the situation can be visualised as in Figure 3.4(b). Indeed by removing the common points of two classes, one can see that the necessary condition of self-affinity is no more ensured. As a consequence

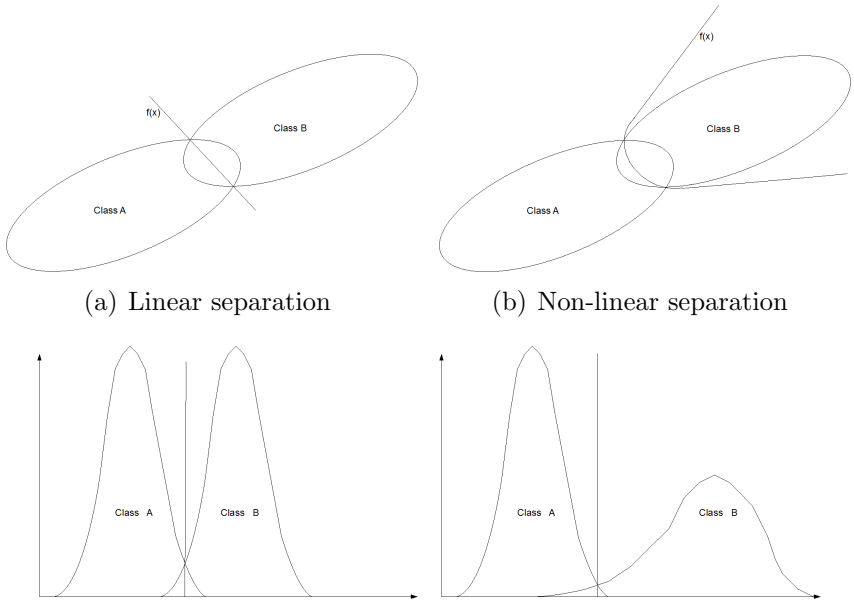
the degree of separability, as depicted in Figure 3.4(c), can be qualitatively defined as the larger set containing self affine subsets (i.e. the percentage of self-affine sub-leaves of the fern in Figure 3.4(c)). Mathematically, thereby, the degree of separability of a class with respect to another is the largest subset of samples which can be classified correctly. Hausdorff dimension hence is a promising tool which allow to estimate the maximum number of statistically self-affine subsets. By removing the common points of classes and computing the Hausdorff dimension, the no-more statistically self-affine subsets will have a different Hausdorff dimension value, therefore the degree of separability can be defined as the percentage of elements which belong to subsets with the same Hausdorff dimension, i.e. the maximum percentage of correctly classified elements.

Box counting D is indeed a practical approach adopted for computing the upper limit of Hausdorff dimension. Box counting dimensions of a compact metric space X is a real number such that if $n(\epsilon)$ denotes the minimum number of the open sets of diameter less than or equal to ϵ, then $n(\epsilon)$ is proportional to ϵ^{-D} as $\epsilon \to 0$, therefore:

$$D = \lim_{\epsilon \to 0^+} \frac{\ln N}{\ln \epsilon} \tag{3.9}$$

Nevertheless this approach has two main deficiencies: Firstly it is difficult to be implemented; secondly it does not work with the sampled data, i.e. if two subsets with a finite number of sample of the same class (as depicted in Figure 3.5) are considered, the value of (3.9) is null, therefore it has no sense. In conclusion, Hausdorff dimension has the advantage of being related

indirectly to the error area (i.e. the intersection of two separated classes has a null Hausdorff dimension), nevertheless it is not a reliable method.
In conclusion, the separability and the degree of separability are concepts which have to be related to the misclassification area. Indeed, as reported in Figure 3.6, two different projection functions (i.e. a linear and a non-linear $f(x)$, for instance) can transform the sample distributions in two different ways, which can produce two different error areas, therefore the degree of separability has to be able to detect the difference in error areas.



(a) Linear separation (b) Non-linear separation

(c) Linear separation projection (d) Non-linear separation projection

Figure 3.6: *Separability problem: two different transformations (i.e. linear and non-linear projection) can produce two different distributions, which have two different error areas. The degree of separability therefore should be able to assess the error area in order to detect the properties of a projection function as reported in this picture.*

3.6 Introduction of the analysis of information preservation

In Section 3.3 an information model flow for SAR/ATR systems is defined. We also mentioned the fact that the model can be used in a forward direction to understand how information content varies through the digital signal processing chain and in backward direction in order to select which parameters maximize/preserve information content in the digital signal procesing chain. In practice the forward procedure is useful in order to understand how a set of samples are mapped in feature space through the digital signal processing chain, whereas the backward procedure is useful to understand, given a known samples set, which information has been lost/distorted by the processing chain.

Understanding how to estimate the information content variation through the system processing chain is one of the main topic of this Chapter. Several mathematical approaches thereby will be considered. First a general introduction of the variation of information content problem will be given and then an appropriate description of the most powerful methods will be reported. Finally the analysis of the advantages/disadvantages and the limits of each of them will be also discussed.

3.7 Variation information content problem

As described in section 3.3, the information flow model is a promising tool in order to understand how information content varies through the process-

ing chain. Unfortunately its structure is strongly related to the considered structure, i.e. a SAR sensor in our case.

The information flow model allows us to understand how the forward procedure can provide a set of techniques which are able to determine how the information content varies in processing chain. This set of techniques can be applied on sampled data (drawn from a known/unknown distribution) or on continuous targets PDFs . The main goal is to emphasize which information is lost (e.g. some taget properties are lost), if unwanted information is introduced (e.g. presence of artifacts in a denoised image) and how system parameters can affect the information content (e.g. Signal-to-Clutter Ratio (**SCR**) gives information on how spread is the sample distribution).

The backward procedure however consists of a set of techniques which are able to determine which parameters are the most important in terms of preserving information content/minimize information distortion. As in the case of forward procedure a backward procedure can be applied to sampled data sets or continuous PDFs , but with the main difference that many ATR subsystems transfer functions are non-linear or do not admit an inverse function, therefore some further assumptions on the data are necessary (i.e. ideal output answer for instance).

So far a theoretical description has been given, of how the information flow can affect the information content investigation, now the description of some common techniques which can be useful tools in order to estimate the information content and its variation through the processing chain are reported. In the following sections Principal Component Analysis, Bayesian Approaches and geometrical methods will be described and their suitability for use within

the forward/backward procedures will be discussed.

3.8 Principal Component Analysis

Principal Component Analysis (**PCA**) [97] is perhaps one of the oldest and best known techniques in multivariate analysis and data mining. The main purpose of PCA is to convert a set of observations of possibly correlated variables into a set of uncorrelated variables called principal components (**PC**s). PCA is performed by an orthogonal linear transformation which transforms data to a new coordinate system such that the greatest variance by any projection of data comes to lie on the first coordinate, then the second greatest variance on the second coordinate etc.
The most important objectives of PCs can be summarized as follows:

- dimensionality reduction;
- feature selection;
- identification of groups of objects or outliers.

Because of its objectives, PCA is often used in solving problems related to data compression, feature extraction, noise filtering, signal restoration and classification. In image processing moreover PCs have been adopted for solving problems such as face and object recognition, tracking, detection and background modeling [97]. PCA thereby could be useful in case of target features selection for determining which potential target parameters are less sensitive to the SAR/ATR processing chain and therefore selected as 'information carrier' through the processing chain.

As known, an orthogonal linear transformation is a linear combination of orthogonal vectors, which preserves a symmetric inner product. In particular an orthogonal transformation preserves the length of vectors and angles between vectors. A very interesting application of orthogonal linear transformation is the linear map diagonalization which consists in finding a basis of vectors V, if exists, with respect to which a linear map is represented by a diagonal matrix. If diagonal matrix is also ordered, i.e. the largest value is placed in the first column, the second greatest value is placed in the second column etc., as reported in 3.10.

$$\begin{pmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & \sigma_n \end{pmatrix} \tag{3.10}$$

PCA is suitable for application to sampled data. Indeed PCA consists in computing Eigenvectors and Eigenvalues of samples covariance matrix. A widely used method to compute Eigenvalues and Eigenvectors is the Singular Value Decomposition (**SVD**) [98]. SVD is a factorization of a matrix, defined as:

$$M = U\Sigma V^{-1} \tag{3.11}$$

where U, V^{-1} are unitary matrices, whereas Σ is a diagonal matrix. Eigenvalues and Eigenvectors can be considered as a special case of SVD when matrices are square.

In practice the PCA consists of following steps:

1. Data organized in data matrix;
2. Data should be normalized before running PCA;
3. Compute the Eigenvectors and Eigenvalues;
4. Ordering the Eigenvectors by ordering the corresponding Eigenvalues from the largest value to the smallest one.

Despite the large use of PCA, it presents some limits. Firstly PCA will perform a linear transformation that makes the data as uncorrelated as possible and it works well for Gaussian data where the mean and covariance matrix provide a complete description of measurements. Secondly, for non-Gaussian distribution or multiple cluster distributions, a covariance matrix is not sufficient to describe the distribution spread. Beside data could have extreme outlying points that bias the PCA analysis. As for the information content variation estimation through PCA, therefore, it is suitable if the samples are drawn from a Gaussian distribution, in case of non-Gaussian distribution PCA is not able to describe the shape of the sample distribution properly (in practice PCA is not suitable because sample distributions in SAR/ATR are usually non-Gaussian). Moreover in case of non-linear processing subsystem (i.e. CFAR procedure), PCA could fail. Hence PCA is not suitable neither as forward nor backward techniques. Nevertheless PCA may be used within a signal processing chain especially if dimensionality reduction is used. Beside if covariance matrix is non-singular, PCA is an invertible procedure,

i.e. PCA is an information preserving procedure when it is using during the processing chain, as long as nonte of the smaller eigenvectors are removed.

3.8.1 Independent component analysis

As described in [97], another interesting technique is Independent Component Analysis (**ICA**). ICA is a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals.
ICA finds the independent components (latent variables or sources) by maximizing the statistical independence of the estimated components. One may choose one of many ways to define independence, and this choice governs the form of the ICA algorithms. The two broadest definitions of independence for ICA are [99]:

1. Minimization of Mutual Information (MI)

2. Maximization of non-Gaussianity

Typical algorithms for ICA use centering, whitening (usually with the eigenvalue decomposition), and dimensionality reduction as preprocessing steps in order to simplify and reduce the complexity of the problem for the actual iterative algorithm.
In general, ICA cannot identify the actual number of source signals, a uniquely correct ordering of the source signals, nor the proper scaling (including sign) of the source signals [100].
ICA is important to blind signal separation and has many practical applications. It is closely related to (or even a special case of) the search for a

factorial code of the data, i.e., a new vector-valued representation of each data vector such that it gets uniquely encoded by the resulting code vector (loss-free coding), but the code components are statistically independent ([99] and [100]).

3.9 Bayesian approach

Since more than a century, science has abandoned Laplace's deterministic vision and has fully accepted to use a random variable analysis to describe system models especially in the case of incomplete knowledge of reality and of the lack of information which forbids a perfect prediction of the future [101]. This statistical approach consists in modelling collected data with proper PDFs. As a consequence the variation of information content can be evaluated by considering how the sample PDFs of different classes vary through the information flow model.

A random variable transformation is a common task to perform an estimation of the information content variation by using PDFs models. Suppose the unidimensional case. Given a generic transformation $\in C^1(\mathbb{R})$, i.e. the set of continuous differentiable function in $\mathbb{R}$:

$$y = g(x) : \mathbb{R} \to \mathbb{R} \tag{3.12}$$

and let x be a random variable with PDF $p_x(x)$, we want to compute the $p_y(y)$, i.e. the PDF of random variable $y = g(x)$. As known, any random

variable transformation has to satisfy the condition:

$$p_y(y)dy = p_x(x)dx \tag{3.13}$$

i.e. the infinitesimal probability has to be the same for both random variables. From hypothesis of differentiable function for $g(x)$, one can state:

$$dy = g^{'}(x)dx \tag{3.14}$$

therefore (3.13) can be written as:

$$p_y(y)g^{'}(x)dx = p_x(x)dx \Leftrightarrow p_y(y) = \frac{p_x(x)}{g^{'}(x)} \tag{3.15}$$

if $g(x)$ is a monotonic decreasing function, (3.15) becomes:

$$p_y(y) = p_x(x)\left|\frac{dx}{dy}\right|. \tag{3.16}$$

Suppose to have an exponential random variable, for example:

$$p_x(x) = e^{-x} \ (x \geq 0) \tag{3.17}$$

it is possible to compute the distribution when x is transformed by $y = \sqrt{x}$. From (3.15) one can write:

$$\begin{aligned} p_y(y) = \left.\frac{p_x(x)}{g^{'}(x)}\right|_{y=\sqrt{x}} = \left.\frac{e^{-x}}{\frac{1}{2\sqrt{x}}}\right|_{y=\sqrt{x}} = \\ = \left. 2\sqrt{x}e^{-x}\right|_{y=\sqrt{x}} = 2ye^{-y^2},\ y \geq 0 \end{aligned} \tag{3.18}$$

where $g^{'}(x) = \frac{1}{2\sqrt{x}}$.

One of the most important deficiencies of the Bayesian approach is due the fact some processing operations are not differentiable (e.g. detection), therefore a Bayesian approach can not handle it easily. Because of this moreover information theory tools (Kullback-Leibler [102], MI metrics [103] etc.) for assessing the information content variation and systems performance comparison fail, because non-differentiable transformation do not allow to compute the transformed PDF . Moreover non-one-to-one maps make the transforms non-reversible, i.e. non information-preserving, the process of PDF -transformation. Hence a Bayesian approach cannot be used as forward/backward procedure for all processing sections.

3.10 Geometric methods

Geometric methods are a set of procedures based on the geometric properties of distribution which tries to overcome the problems encountered with PCA analysis and Bayesian approaches.

Differential geometry is a mathematical branch which uses techniques of differential and integral calculus as well as linear and multi-linear algebra to study problems in geometry and it is the basis of concepts such as differentiable manifold widely used in Information Geometry [102]. Intuitively a manifold is a 'set with a coordinate system' and it usually represents a generalization of geometric objects such as smooth curves or surfaces in a n-dimensional space. An example of a manifold is a set whose points are probability distributions which is provided with a coordinate system.

Let S be a manifold and $\varphi : S \to \mathbb{R}^n$ be a coordinate system for S. Since φ maps each point p in S to n real numbers $\varphi(p) : [\xi^1(p), \ldots, \xi^n(p)]$ then each $\xi^i(p),\ i = 1, \ldots, n$ can be considered as a function $p \to \xi^i(p),\ i = 1, \ldots, n$ which map a point p in the i-th coordinate, therefore the maps $n\ \xi^i(p) : S \to \mathbb{R},\ i = 1, \ldots, n$ are defined as the coordinate functions.

If another coordinate system $\psi = [\rho^i]\ \ i = 1, \ldots, n$ for S is considered, the coordinate transformation from $\varphi : [\xi^i]$ to $\psi = [\rho^i]\ \ i = 1, \ldots, n$ is the application defined as:

$$\psi \circ \varphi^{-1} : \left[\xi^1, \ldots, \xi^n\right] \to \left[\rho^1, \ldots, \rho^n\right]. \tag{3.19}$$

Formally the previous concepts are defined as follows: Let S be a set. If there exists a set of coordinate systems A for S which satisfy:

1. Each element φ of A is a one-to-one mapping from S to some open subset of $\mathbb{R}$;

2. For all $\varphi \in A$, given any one-to-one mapping ψ from S to $\mathbb{R}^n$, the following holds:

$$\psi \in A \Leftrightarrow \psi \circ \varphi^{-1}$$
$$\text{is a } C^\infty \text{ diffeomorfism (i.e. } (\psi \circ \varphi^{-1})^{-1} \text{ is still } C^\infty) \tag{3.20}$$

S is termed an $n-$ dimensional (C^∞-differentiable) manifold.

Usually φ is called '*chart*', whereas φ^{-1} is defined parametrization and the maximal set containing all the charts of a manifold is termed '*atlas*'.

Differential Geometry is useful also for defining another powerful mathemat-

ical tool for our purpose: Lie groups [104], [105], [106] and [107]. Lie groups are often used in physics as a group of transformations acting on a manifold S. Let $\{x^i\}$ be a chart of S, an r-parameter Lie group q of transformation is a group of transformations of S defined as:

$$x^{'(i)} = f^i(q^1, \ldots, q^r; x^1, \ldots, x^n) \ (x^{'} = f(\vec{q}; \vec{x})) \tag{3.21}$$

for which the function $f^i,\ (i = 1, \ldots, n)$ are smooth function of r-parameters $q^k,\ (k = 1, \ldots, r)$ (assumed to be essential to determine the transformation). Consider for instance the one-dimensional transformation:

$$x^{'} = a \cdot x,\ a \neq 0 \tag{3.22}$$

if the product of two such operations is defined, i.e. $x^{''} = b \cdot x^{'}$ and $x^{'} = a \cdot x$ $(a, b \neq 0)$, the following result is obtained:

$$x^{''} = a \cdot x^{'} = ab \cdot x \tag{3.23}$$

By rewriting $x^{''} = c \cdot x$, one can write:

$$c = a \cdot b \tag{3.24}$$

so the multiplication of two transformations is described by a function which have the same form of (3.23). This operation is an Abelian group [104], because the product of transformation corresponds to the multiplication of real numbers, therefore the transformation defined in (3.23) is a one-parameter

Abelian Lie group. As a consequence of the definition of Lie groups and differential geometry, one can consider set of samples and process it in order to understand how their coordinates change in a feature space. It is clear that Geometrical methods are suitable for analysis of both the forward and backward procedure. Indeed by mapping the samples we can understand how the information content changes through the processing chain, whereas comparing the actual output with the ideal one, one can understand which parameters convey the information. In practice Geometrical method can be used to map sample into the feature space in order to analyse how the sample sets change their topology into the feature space.

3.10.1 Approximation using Unscented Theory

Handling charts and atlas cannot be easy especially in a high dimensional space because of complexity of the problem of managing the coordinates transformations in a high dimensional features space.

A more convenient approach can be the unscented transformation. As described by Julier and Uhlmann in [108], unscented theory is based on the concept that it is easier to approximate a Gaussian distribution than it is to approximate an arbitrary non-linear function or transformation.

As illustrated in Figure 3.7, a set of points, termed *sigma points*, are chosen so that their mean and covariance matrix are $\bar{x}$ and P_x respectively. The non-linear function is applied to each point in turn to yield a cloud of transformed points and P_{yy} are the statistics of transformed points. The main difference with respect to Monte Carlo simulation consists in the fact that

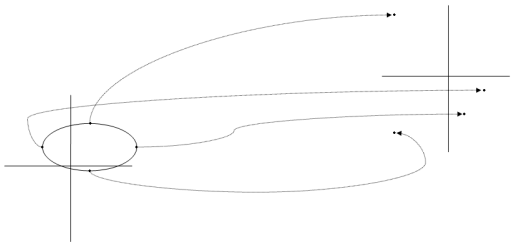


Figure 3.7: *Unscented Theory model*

the samples are not drawn at random but rather according to a deterministic algorithm. In practice the n-dimensional random variable x with mean $\bar{x}$ and covariance P_{xx} is approximated by $2n+1$ weight points defined as:

$$\begin{aligned} X_0 &= \bar{x} & w_0 &= k/(n+k) \\ X_i &= \bar{x} + (\sqrt{(n+k)P_{xx}})_i & w_i &= 1/(2(n+k)) \\ X_{i+n} &= \bar{x} - (\sqrt{(n+k)P_{xx}})_i & w_i &= 1/(2(n+k)) \end{aligned} \tag{3.25}$$

where $k \in \mathbb{R}$, $(\sqrt{(n+k)P_{xx}})_i$ is the i-th row or column of the matrix square root of $(n+k)P_{xx}$ and w_i is the weight associated with the i-th point. The algorithm can be summarized as follows:

1. $y_i = f[X_i]$ the transformation of sigma points;

2. computation of:

$$\bar{y} = \sum_{i=0}^{2n} w_i y_i; \tag{3.26}$$

3. computation of covariance matrix P_{yy}:

$$P_{yy} = \sum_{i=0}^{2n} w_i \{y_i - \bar{y}\} \{y_i - \bar{y}\}^T \tag{3.27}$$

The most important properties of this approach are:

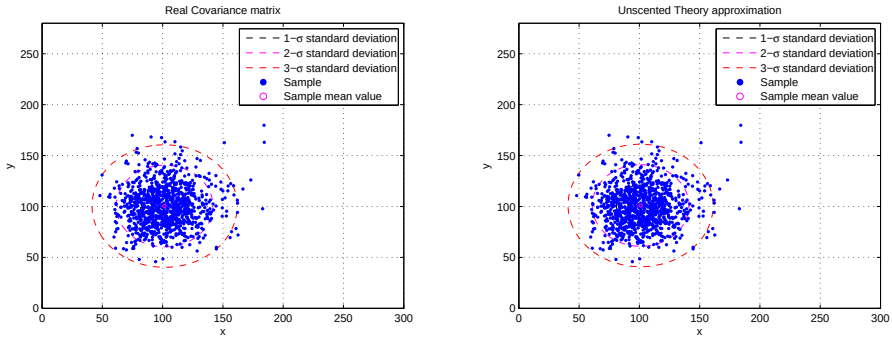
- Mean and covariance matrix of y are correct to the second order, because the corresponding values of x are correct.
- The parameter k provides an extra degree of freedom to 'fine tune' the higher order moments of the approximation.

As for experiments, unscented transformation was compared with the linearization of the function [109]:

$$x^2 \approx c_0 + 2c_1(x - x_0) \tag{3.28}$$

If the input is a bi-dimensional Gaussian distribution with mean value $\underline{\mu} = [10, 10]$ and an identity matrix as covariance matrix we can determine the corresponding output value of the linearized function (3.28). If the Taylor approximation is stopped at first order, we obtain as result a Gaussian with unitary mean value and standard deviation.

By applying the mentioned algorithm and considering the $\pm 3 \times \lambda_i$, where λ_i $(i = 1, 2)$ is the eigenvalue of matrix P_{yy}. As for the considered parameter (sigma points ($k = 1,\ n = 2$ in (3.25)): $X_0 = [10, 10]$ $X_1 = [11.7321, 10]$ $X_2 = [10, 11.7321]$ $X_3 = [8.2679, 10]$ $X_4 = [10, 8.2679]$), whereas the corresponding weights are: $w_0 = 0.3333$, $w_1 = w_2 = w_3 = w_4 = 0.1667$), the percentage



(a) Real Covariance Matrix (b) Unscented Theory approximation

Figure 3.8: *Unscented Transform examples*

Method	1σ	2σ	3σ
Real Covariance Matrix	0.3920	0.8640	0.9870
Unscented approximation	0.4001	0.8720	0.98

Table 3.1: *Percentage of samples within $n - \sigma_i$ ($n = 1, 2, 3$, σ_i standard deviation of i-th component, i.e. $i = 1, 2$) for real covariance matrix and unscented theory approximation.*

of sample which are inside the range of $3 - \sigma$ is reported in Table 3.1. The mean value and the covariance matrix of the real transformed sample and the approximation one are reported in (3.29) and (3.30) respectively and depicted in Figure 3.8. Unscented Theory, in conclusion, is able to approximate the Geometrical methods and it has the advantage of being faster (i.e. less time consuming) than the application of the mathematical formalism.

$$Cov = \begin{pmatrix} 402.4125 & -0.7128 \\ -0.7128 & 401.8285 \end{pmatrix}$$
$$Mean = [101.0117, 101.0070] \tag{3.29}$$

$$Cov = \begin{pmatrix} 402 & -1 \\ -1 & 402 \end{pmatrix}$$
$$Mean = [101, 101] \tag{3.30}$$

As for the linearization of non-linear transform (3.28) (i.e. $y \approx 2 \cdot x$), the results can be summarized as a Gaussian distribution with mean value of $[20, 20]$ and a covariance matrix equal to $4 \cdot I$, where I is the Identity matrix, which represents an huge error of approximation of the non-linear problem.

3.11 Summary

In Chapter 3 an overview of the theoretical background of the thesis has been reported. Firstly an introduction to Bayes theory and Bayes classifier has been given, then the main advantages of 'before classification analysis' have been reported. In order to analyse the variation of information content through the processing chain an information flow model for ATR systems has been described. The definition of an information flow model allowed us to describe (generally and mathematically) the problem of class-separability in ATR systems. Another important consequence of the definition of the information flow model is the classification of the procedure for assessing the ATR performances in forward and backward procedures. Finally a set of mathematical techniques have been analysed in order to check their suitability as method for assessing the variation of information content in an ATR system. The most important techniques can be summarized as follows:

- Principal Component Analysis;
- Independent Component Analysis;
- Bayesian Approach (i.e. random variable transformation);
- Geometric methods (Differential Geometry, Lie Groups and Unscented Theory method).

Chapter 4

Metrics for degree of separability

4.1 Introduction

The comparison of the most popular class-separability metrics is the main topic of this chapter and the most common methods adopted by researchers for assessing the degree of class-separability will be analysed. Firstly the suitability of metrics which are based on the Covariance matrix will be analysed, then the criteria based on information theory (i.e. Entropy/Kullback-Leibler divergence), the methods based on thresholding and the graphical methods based on thresholding (i.e. ROC and AUC) will be studied. Finally a new metric will be introduced.

4.2 Covariance based methods

4.2.1 Linear Discriminant Analysis

Linear Discriminant Analysis [32] (**LDA**) is a very common method for estimating the degree of separability by considering scatter matrices of samples. A *within-class scatter matrix* shows the scatter of samples around their class mean values and it is defined as:

$$S_w = \sum_{i=1}^{L} P_i E\left\{(X - M_i)(X - M_i)^T | \omega_i\right\} = \sum_{i=1}^{L} P_i C_i \tag{4.1}$$

where L is the number of classes, P_i is the *a priori* class probability, $E\{\cdot\}$ is the expectation operator, M_i the class mean value, ω_i is the class label and C_i the class covariance matrix. On the other hand a *between class scatter matrix* is equal to:

$$S_b = \sum_{i=1}^{L} P_i E\left\{(M_i - M_0)(M_i - M_0)^T | \omega_i\right\} \tag{4.2}$$

where $M_0 = E\{X\} = \sum_{i=1}^{L} P_i M_i$. Finally *mixture scatter matrix* is the covariance matrix of all samples regardless their class assignments:

$$S_m = E\left\{(X - M_0)(X - M_0)^T\right\} = S_w + S_b \tag{4.3}$$

As for the LDA as separability metrics it should be able to reach a large value when the between class scatter is larger or the within-class is smaller, therefore some common criteria for assessing the degree of separability are:

1. $J_1 = \mathrm{tr}(S_2^{-1}S_1)$;

2. $J_2 = \ln\left|S_2^{-1}S_1\right| = \ln|S_1| - \ln|S_2|$;

3. $J_3 = \frac{\mathrm{tr}S_1}{\mathrm{tr}S_2}$.

where S_1 and S_2 are one of S_b, S_w and S_m, whereas $\mathrm{tr}(\cdot)$ represents the matrix trace.

The criterion is powerful for many reasons, such as it is suitable for being computed on sampled data and it is indirectly related to error area. Its main disadvantages are: 1) it depends on the *a priori* probability of ω_i (i.e. P_i) which is usually unknown; 2) it is only guaranteed for Gaussian distributed samples; 3) it does not give any information on the error area, i.e. no guarantee that requirements on false alarm rate are satisfied.

4.2.2 Bhattaccharrya and Chernoff bounds

The *Bhattaccharrya* bound is a special case of the *Chernoff* bound ([32], [68]) defined in equations (4.5) and (4.4) respectively (P_j is a priori probability, whereas $p_j(X)$ is the likelihood functions, $j = 1, 2$). Both of them define an upper bound of error probability. For Normally distributed classes, the corresponding distances (used for measuring the separability of classes) are defined as in equations (4.6) and (4.7).

$$\epsilon_u^t = P_1^t P_2^{1-t} \int p_1^t(X) p_2^{1-t}(X) dX \quad \text{for } 0 \leq t < 1 \tag{4.4}$$

$$\epsilon_u^{1/2} = P_1^{1/2} P_2^{1/2} \int p_1^{1/2}(X) p_2^{1/2}(X) dX \tag{4.5}$$

$$\mu(t) = \frac{t(1-t)}{2}(M_1 - M_2)^T \left[t\Sigma_1 + (1-t)\Sigma_2\right]^{-1}(M_1 - M_2) + \\ + \frac{1}{2}\ln\frac{|t\Sigma_1 + (1-t)\Sigma_2|}{|\Sigma_1|^t\,|\Sigma_2|^{1-t}} \tag{4.6}$$

$$\mu(1/2) = \frac{1}{2}(M_1 - M_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}(M_1 - M_2) + \\ + \frac{1}{2}\ln\frac{\left|\frac{\Sigma_1+\Sigma_2}{2}\right|}{\sqrt{|\Sigma_1|\,|\Sigma_2|}} \tag{4.7}$$

where M_j and Σ_j (for $j = 1, 2$) are the mean value and the covariance matrices of the distributions.

4.2.3 Mahalanobis distance

Mahalanobis distance however is defined as:

$$d(\vec{x}, M) = \sqrt{(\vec{x} - M)^T \Sigma^{-1} (\vec{x} - M)} \tag{4.8}$$

and it is used for determining the distance from the class distribution mean value M weighted by the inverse of covariance matrix Σ. As a consequence Mahalanobis distance is not symmetric. Moreover Mahalanobis distance indicates how distant a sample from a mean value is, but it does not give any information about if two classes are separated or not.

4.3 Kullback-Leibler divergence

Another common measurement of the difference of two classes is performed by Kullback-Leibler divergence [102] (or relative entropy) defined for continuous

Probability Density Function (PDF) as:

$$D_{KL}(P_1||P_2) = \int_{-\infty}^{+\infty} p_1(X) \log \frac{p_1(X)}{p_2(X)} dX. \tag{4.9}$$

where $p_1(X)$ and $p_2(X)$ are the two classes PDFs .
Moreover relative entropy can be easily computed only for exponential family distributions (i.e. Gaussian, Exponential, Gamma etc. distributions).

4.4 Thresholding criteria description

In the previous section the concept of class-separability and the degree of separability have been introduced. Now a set of potential metrics for assessing the degree of separability can be analysed.

4.4.1 K-S threshold

The Kolmogorov Smirnov (K-S) test [110] is usually adopted in statistics in order to detect if two data populations have been drawn from the same distribution. The test consists of determining the maximal distance in probability, D_n (n, number of population samples as depicted in Figure 4.1(a)), between the populations CDFs and comparing the selected distance with a threshold K_α (found from the Kolmogorov distribution where α is the level of confidence): if $\sqrt{n}D_n > K_\alpha$, then the null hypothesis of the sample being drawn from the same distribution is rejected, otherwise it is accepted. Graphically, as depicted in Figure 4.1(a) and Figure 4.1(c), K-S threshold (T_{KS}) is hence

the value of random value X such that the error of misclassification is minimized in the Bayesian sense. Beside $\vec{T}_{KS}$ can be mathematically defined as :

$$\begin{aligned} \vec{T}_{KS} &= \max_{\vec{x}\in\mathbb{R}^n} \left\{ \int_{\vec{x}\leq f(\vec{x})\subseteq\mathbb{R}^n} p_1(\vec{x})d\vec{x} - \int_{\vec{x}\leq f(\vec{x})\subseteq\mathbb{R}^n} p_2(\vec{x})d\vec{x} \right\} \\ &\Leftrightarrow \vec{T}_{KS} = \left\{ \vec{x} : \frac{\partial}{\partial x_i} F_1(x) = \frac{\partial}{\partial x_i} F_2(x) \right\}, \text{ i=1,. . . ,n.} \end{aligned} \tag{4.10}$$

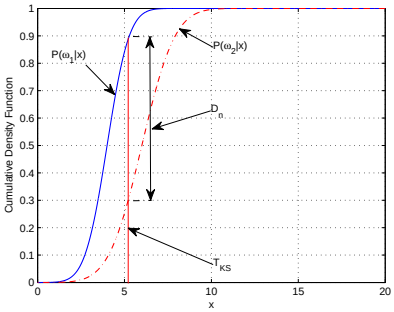
where $F_j(\cdot)$, $j = 1, 2$ are the CDFs of the populations, whereas $f(\cdot)$ is the generic surface which generalized the T_{KS} in $\mathbb{R}^n$. As for the last step in (4.10), it is the result of first derivative of the argument of the $\max\{\cdot\}$ condition.

4.4.2 Support Vector Threshold

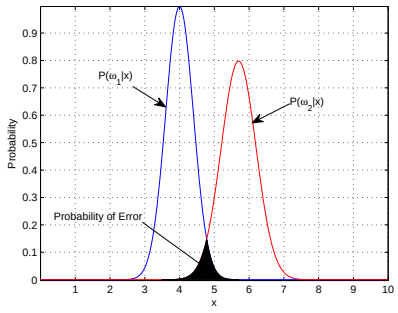
A related measure is the Support Vector Threshold (T_{SV}, adopted from SVM [111] and as depicted in Figure 4.2), which is the hyperplane which ensures a fixed value of the generalized error by maximizing the functional margin. In a linear unidimensional case (very common in ATR/SAR systems) the optimal T_{SV} is equal to T_{KS}. Indeed, as known for the linear case, the problem of the optimal hyperplane is described as:

$$y_i(\mathbf{w}\cdot\mathbf{x}_i - b) \geq 1 - \xi_i \quad 1 \leq i \leq n \tag{4.11}$$

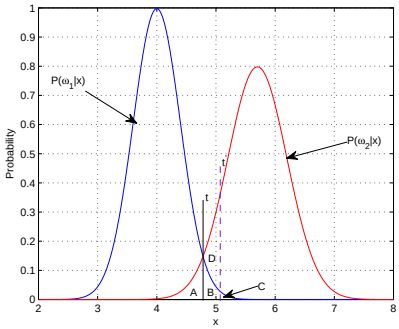
where $y_i \in \{-1, 1\}_{i=1}^n$ is the class label, $\mathbf{x}_i$ is the $i-$th sample, $\mathbf{w}$ and b are the parameter of the hyperplane which has to be estimated, C is an arbitrary constant, whereas $\xi_i > 0$ is a slack variable which is a measure of



(a) Kolmogorov Smirnov distance and threshold



(b) Error probability area



(c) Error probability area changes with two different thresholds

Figure 4.1: *Figure 4.1(a) represents the distance D_n and T_{KS} respectively. Error probability area: Definition Figure 4.1(b), whereas Figure 4.1(c) represents the fact that the value of error probability for $x > t^{'}$ is equal to C, whereas for $x \leq t^{'}$ is equal to $A + B + D$, therefore the total error area is equal to $A + B + C + D$, which is bigger than case t by D.*

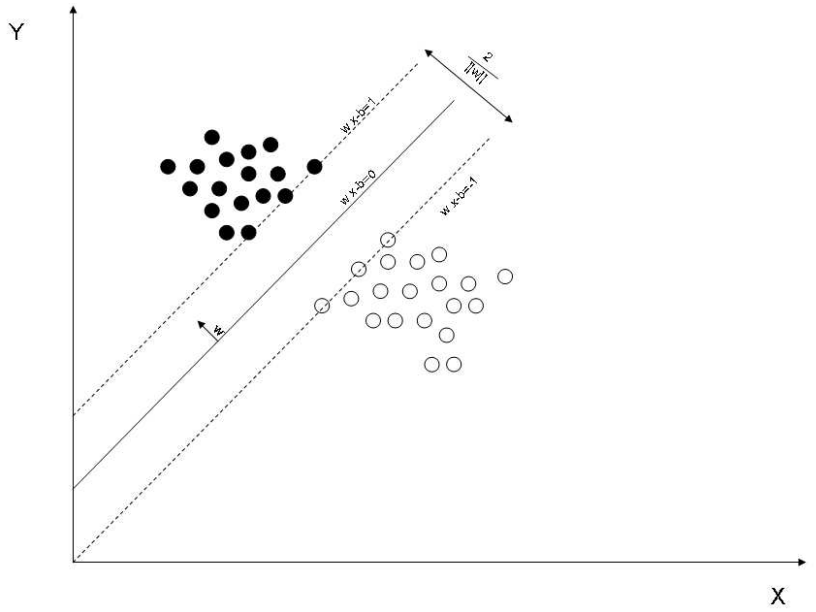


Figure 4.2: *SVM hyperplane example: SVM classifiers consists in computing a hyperplane which maximizes the margins (dotted lines). The Support Vector Threshold (T_{SV}) is practically the same hyperplane, but its meaning is different (i.e. SVM is related to the classification, whereas it is the hyperplane which maximizes the class-separability).*

misclassification of data x_i. The optimization problem is:

$$\min_{\mathbf{w},\xi} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{n} \xi_i \right\} \tag{4.12}$$

In the one dimension case, equation (4.12) becomes:

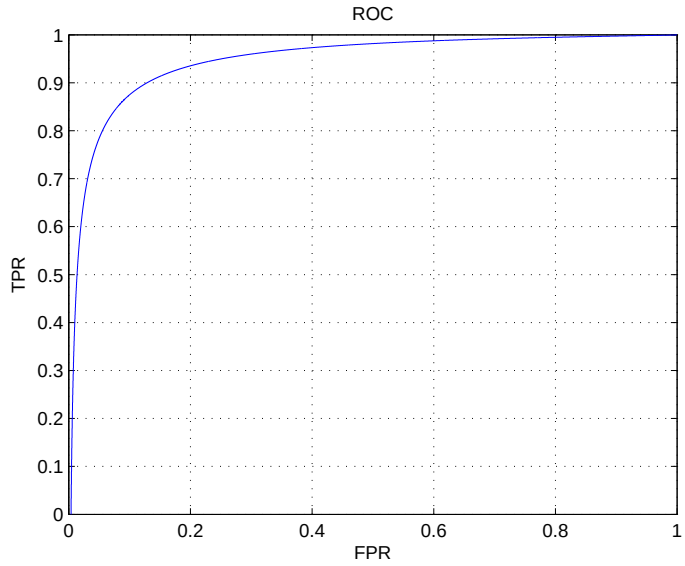
$$\min_{\xi} \left\{ \sum_{i=1}^{n} \xi_i \right\} \tag{4.13}$$

i.e. b is the threshold which minimizes the error of misclassification. In many cases the linear SV hyperplane is an approximation which can have as good performance as or worse than the real hyperplane.

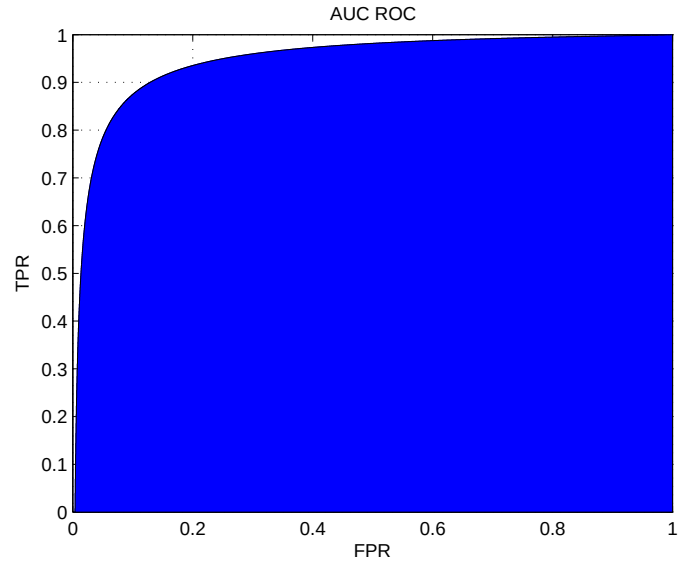
4.4.3 ROC analysis

A ROC graph is a technique for visualizing, organizing and selecting classifiers based on their performance. Fawcett in [112] gave several examples of successful uses of this kind of techniques for visualizing the performance of a classifier. Smith in *et al.* [76] considered ROC analysis for summarizing the information gathered through probabilities of correct classification, Unknown detection, declaration, false alarm and generalization. Smith *et al.* in [76] considered the area under the ROC curve (AUC) as a criterion for maximizing the detection of land-mines. In some applications AUC is preferred because it is able to produce a single number which is easier to be handled (an example of ROC and AUC is reported in Figure 4.3).

By thresholding the generated sample of (4.17), (4.18) and (4.19) the be-



(a) ROC curve



(b) AUC, Area Under the Curve

Figure 4.3: *Figure 4.3(a) represents ROC curve. Figure 4.3(b) represents however the Area Under the Curve ROC.*

haviour of the ROC curve and corresponding AUC as class-separability metrics was analysed.

As for ROC analysis, the following parameters have been computed:

$$\text{True positive rate} = \frac{\text{Positive correctly classified}}{\text{Total positives}} = \frac{\text{TP}}{\text{TP+FN}} \tag{4.14}$$

$$\text{False positive rate} = \frac{\text{Negative correctly classified}}{\text{Total negatives}} = \frac{\text{FP}}{\text{FP+TN}} \tag{4.15}$$

where the *TP*, *FN*, *FP* and *TN* are depicted in Figure 4.4. However as for the

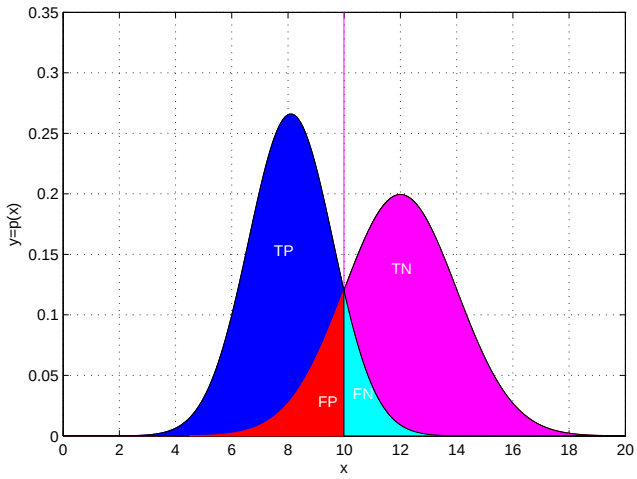


Figure 4.4: *Parameter definition of ROC curve.*

computing of AUC of a ROC the trapezoidal rule has been adopted, which is defined as:

$$\int_a^b f(x)dx \approx \frac{b-a}{2N}(y_0 + 2y_1 + \ldots + 2y_{N-1} + y_N) \tag{4.16}$$

where N is the number of subintervals and y_i, $i = 1 \ldots N$ the number of the point of the function over which the integral is computed.

4.5 Experiments and results

Several experiments have been performed in order to verify the condition under which the separability is ensured and at which degree, therefore three classes of samples (10^7 samples for each class) have been generated, drawn from Exponential, Rayleigh and Gaussian distributions.

As for samples drawn from the Exponential distribution, they satisfy the following conditions:

$$\begin{aligned} p_E(x) &= \lambda e^{-\lambda x},\ \lambda = 1 \\ \text{Mean value: } \langle x \rangle &= \lambda^{-1} = 1 \\ \text{Variance: } \left\langle x^2 \right\rangle - \langle x \rangle^2 &= \lambda^{-2} = 1 \end{aligned} \tag{4.17}$$

As for the Rayleigh distribution however, the sample parameters are:

$$\begin{aligned} p_R(x) &= \frac{x}{\sigma_R^2} e^{\frac{x^2}{2\sigma_R^2}}\ \sigma_R = 1 \\ \text{Mean value: } \langle x \rangle &= \sigma_R \sqrt{\frac{\pi}{2}} = \sqrt{\frac{\pi}{2}} \\ \text{Variance: } \left\langle x^2 \right\rangle - \langle x \rangle^2 &= \frac{4-\pi}{2}\sigma_R^2 = \frac{4-\pi}{2} \end{aligned} \tag{4.18}$$

Finally, the Gaussian distribution is defined as:

$$p_G(x) = \frac{1}{\sqrt{2\pi\sigma_G^2}} e^{-\frac{(x-\mu_G)^2}{2\sigma_G^2}} \tag{4.19}$$

where $\mu_G = 5$ is the mean value and $\sigma_G = 1$ is the standard deviation. In the next subsections all the described techniques are analysed.

4.5.1 Kolmogorov Smirnov Threshold

As for T_{KS}, it is the value of x corresponding at the intersection between PDFs and it minimizes the error of overall misclassification. Indeed, as depicted in Figure 4.5 and Figure 4.6, if the T_{KS} and its values $T_{KS}\pm\epsilon$, $\epsilon = 10\%$ of the T_{KS} true value, are considered and the corresponding error of misclassification as reported in Table 4.1 for Exponential-Gaussian case (E-G) are computed, one can state that T_{KS} minimizes the error of misclassification. As for the separability, the same previous experiments by comparing the Rayleigh distributed class with the Gaussian distributed one were considered (as reported in Table 4.1, R-G case). By considering the two experiments, one can affirm that a necessary condition for the separability of the two classes is to ensure that error probability is minimized. Indeed the experiments confirm that the smaller the error probability area is the smaller the error of misclassification is obtained. In conclusion one can affirm that T_{KS}

	T_{KS}	$T_{KS}+\epsilon$	$T_{KS}-\epsilon$
Error of misclassification E-G	0.0726	0.0799	0.0789
Error of misclassification R-G	0.0328	0.0418	0.0411

Table 4.1: *Error of misclassification for T_{KS} (T_{KS}^{E-G}=2.9730 and T_{KS}^{R-G}=2.8964): The true value and error affected ones ($\epsilon = \pm 10\%$) are used in order to estimate the error of misclassification. The true value of T_{KS} preduces the smallest error of misclassification.*

is not suitable for our purposes because the error area cannot assessed.

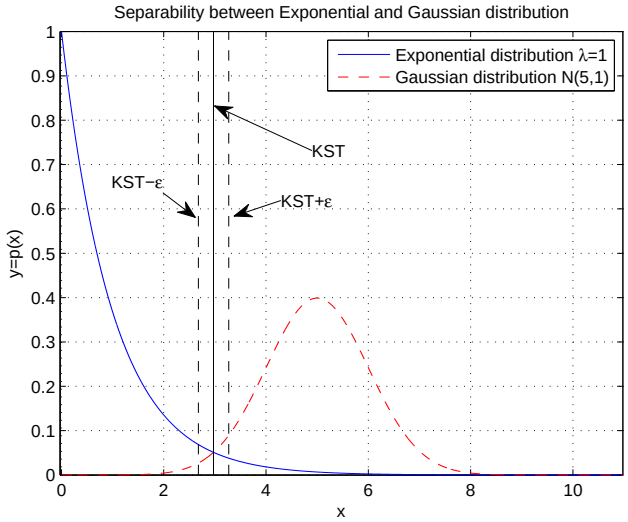


Figure 4.5: *T_{KS} experiment: comparison of the error of misclassification of the true value and a threshold value affected by an error $\epsilon = \pm 10\%$ if the true value.*

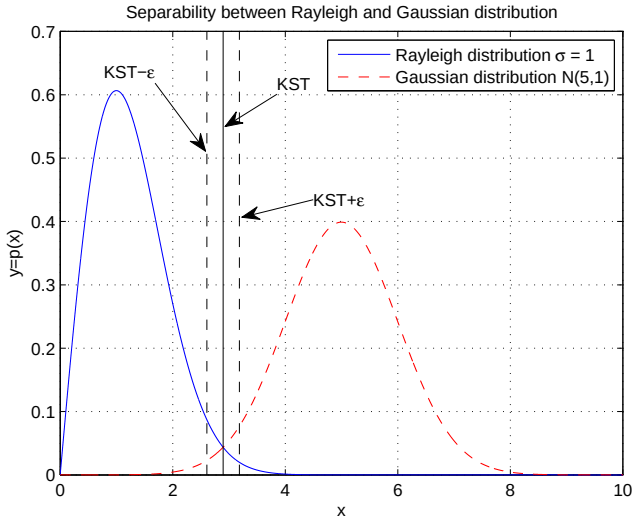


Figure 4.6: *T_{KS} experiment: comparison of the error of misclassification of the true value and a threshold value affected by an error $\epsilon = \pm 10\%$ of the true value.*

4.5.2 Kolmogorov Appropriate Prediction of Separability: KAPS

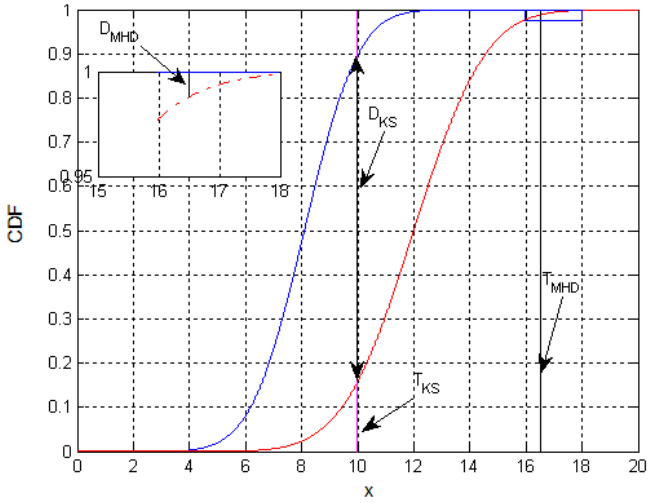


Figure 4.7: *KAPS parameters description.*

Since PDFs usually have an infinite domain (i.e. $x \in (-\infty, +\infty)$), there is properly no condition of separability between classes, therefore a new quantity, termed 'Threshold of Marino-Hughes Threshold' (T_{MHD}), is defined which determines the tolerated maximum value of the type I error (i.e. false negative which occurs when null hypothesis is rejected when it is in fact true):

$$T_{MHD} = \min_{x \in \mathbb{R}} \left\{ \int_{x}^{+\infty} p(\xi) d\xi \leq \alpha \right\}, \text{ with } \alpha \ll 1. \tag{4.20}$$

In our experiment $\alpha = 10^{-6}$ was set (as described in Table 4.2, which reports the theoretical value and the corresponding number of false positive samples).

As depicted in Figure 4.7 the T_{KS} represents the value of the threshold

Distribution	Theoretical T_{MHD}	Simulated Type I Error
Exponential	13.8155	8×10^{-7}
Rayleigh	5.2565	10^{-6}
Gaussian	9.7534	8×10^{-7}

Table 4.2: *Marino-Hughes Distance for Exponential, Rayleigh and Gaussian distributions: Simulated Type I Error is the percentage of* 10^7 *samples which overcomes the value of* T_{MHD}.

which maximizes the correct classification of the ω_1-class and at the same time minimizes the misclassification of ω_2-class, i.e. it maximizes the value of the TPR and at the same time minimizes the largest value of FPR. Moreover the closer the T_{MHD} and T_{KS} are, the smaller the error area is, as a consequence the distances corresponding to T_{MHD} and T_{KS} tend to be equal, when $T_{MHD} = T_{KS}$.

In order to determine the value of degree of separability the following metric was adopted [172]:

$$KAPS = \frac{D_{KTS}}{D_{MHD}} \in (0, +\infty) \tag{4.21}$$

where D_{KTS} is the distance between two classes CDFs corresponding at T_{KS} (i.e. the maximum one) and D_{MHD} is the analogous distance corresponding at T_{MHD}. As a consequence of the definition, the separable case is proved when KAPS is equal to 1.

4.5.3 ROC analysis

As for the experiments the results related to ROC analysis are reported in Figure 4.8.

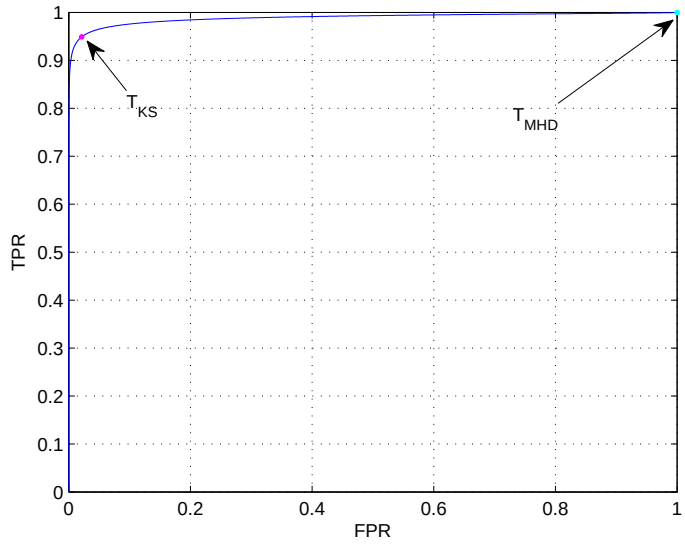
AUC analysis was related to the previous experiments and also to considering a new experiment has been described where the Exponential distribution parameters are chosen in order to have the same error area of R-G case, i.e. $\lambda = 1.4116$ in (4.17). In this case, the ROC curves are reported in Figure 4.9, whereas the AUCs values are reported in Table 4.3. It is clear that the ROC analysis is difficult to interpret, whereas AUC produces ambiguities in terms of AUC values interpretation, i.e. 'bigger is better' is not true (as demonstrated comparing second and third line of Table 4.3), and because an area under the curve could be the same for two different ROC curves which represent two different class-separability cases. Indeed ROC analysis and AUC are not able to assess the relative position between T_{KS} and T_{MHD}, which is an indirect estimation of error area. Moreover distribution shapes can affect the performance of ROC analysis and AUC, especially the latter, because one can obtain good results in terms of these metrics, despite the real scenario is absolutely different, as clarified later.

Experiment	AUC
E-G	0.9889
R-G	0.9986
E*-G	0.9977

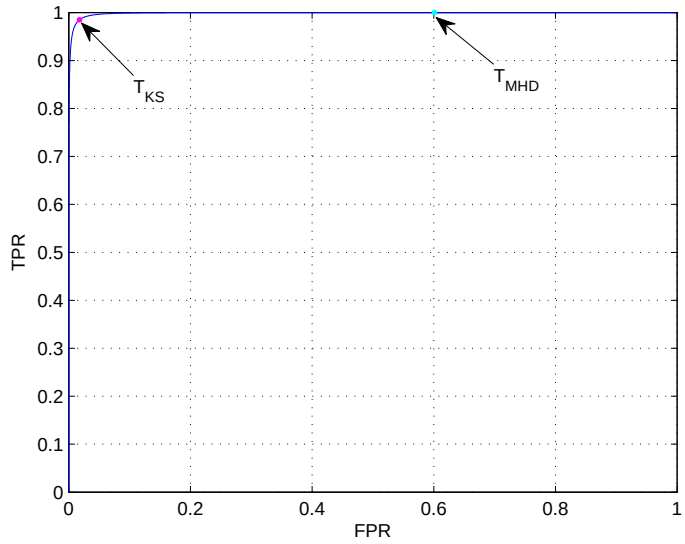
Table 4.3: *Values of AUC*

4.5.4 Bhattaccharrya, Chernoff distances and Kullback-Leibler divergence

As for the Chernoff distance in equation (4.22) and (4.24) as well as the Bhattaccharrya distance in equation (4.23) and (4.25), the investigations have



(a) ROC curve E-G case



(b) ROC curve E-G case

Figure 4.8: *Figure 4.8(a) represents ROC curve for class-separability in case of Exponential and Gaussian classes distribution. Figure 4.8(b) represents ROC curve for class-separability in case of Rayleigh and Gaussian classes distribution.*

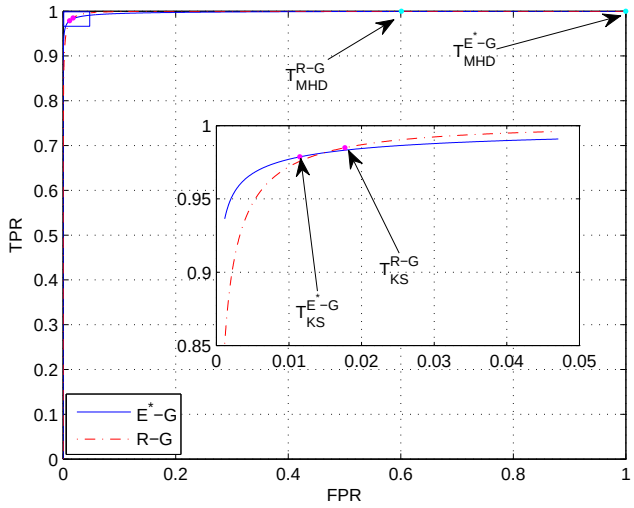


Figure 4.9: *ROC analysis for same area distribution cases.*

been performed by using $Mathematica^{TM}$ in order to find a closed form for the integrals (4.4) and (4.5). As results the following formulas (Exponential-Gaussian case and Rayleigh-Gaussian case respectively with parameters of equations (4.17), (4.18) and (4.19)) were obtained:

$$\int_0^{+\infty} \frac{e^{-xs}}{\sqrt{2\pi}} e^{-\frac{(x-5)^2(1-s)}{2}} dx = \left[e^{-\frac{1}{2}s\left(\frac{s}{s-1}+10\right)} \text{erfi}\left(\frac{s(x-6)-x+5}{\sqrt{(2)}\sqrt{s-1}} \right) \frac{1}{2\sqrt{s-1}} \right]_0^{+\infty} \tag{4.22}$$

$$\int_0^{+\infty} \frac{e^{-\frac{x}{2}}}{\sqrt{2\pi}} e^{-\frac{(x-5)^2}{4}} dx = [0.0745285\text{erf}(0.5x-2)]_0^{+\infty} \tag{4.23}$$

$$\int_0^{+\infty} x \frac{e^{-\frac{x^2 s}{2}}}{\sqrt{2\pi}} e^{-\frac{(x-5)^2(1-s)}{2}} dx = \left[\frac{1}{2\sqrt{2\pi}} e^{s(\frac{25}{2}-5x)-\frac{1}{2}(x-5)} \left(5\sqrt{2\pi}(1-s) e^{\frac{1}{2}(5s+x-5)^2} \text{erf}\left(\frac{5s+x-5}{\sqrt{2}} \right) + 2 \right) \right]_0^{+\infty} \tag{4.24}$$

$$\int_0^{+\infty} x \frac{e^{-\frac{x^2}{4}}}{\sqrt{2\pi}} e^{-\frac{(x-5)^2}{4}} dx = \left[0.055\text{erf}(0.707x - 1.767) - 0.0008 e^{x(2.5-0.5x)} \right]_0^{+\infty} \tag{4.25}$$

Where erfi$(\cdot)$ is the complex error function, whereas erf$(\cdot)$ is the classical error function. Equations (4.4) and (4.5) for the non-Gaussian mixture cannot be computed easily and the obtained results are not reliable (i.e. integrals produce either undetermined expressions or values which are difficult to interpret). Moreover these kind of metrics are not easy to compute numerically when a finite number of samples drawn from an unknown distribution are available. A similar conclusion is valid for Kullback-Leibler divergence. As for the case of T_{KS}, the previous methods fail in being able to take into account of a restriction of the maximum tolerable error area.

4.5.5 Mahalanobis threshold

As defined in equation (4.8), Mahalanobis classifiers measure the Mahalanobis distance for each class and then select the minimum one. As a consequence, the Mahalanobis thresholds, i.e. the value of x where equation (4.8) for two classes (i.e. two different mean values and covariance matrix

respectively) assume the same value, is obtained by:

$$d^2(\vec{x}, M_1) = (\vec{x} - \vec{M_1})^T \Sigma_1^{-1} (\vec{x} - \vec{M_1}) = (\vec{x} - \vec{M_2})^T \Sigma_2^{-1} (\vec{x} - \vec{M_2}) = d^2(\vec{x}, M_2) \tag{4.26}$$

The threshold for the unidimensional case for two population with the same variance is equal to:

$$T_{MD} = \frac{M_2 + M_1}{2} \tag{4.27}$$

whereas in case of different variances the threshold is equal to:

$$T_{MD} = \frac{(\frac{M_1}{\sigma_1} - \frac{M_2}{\sigma_2}) \pm \sqrt{(\frac{M_1}{\sigma_1} - \frac{M_2}{\sigma_2})^2 - (\sigma_1^{-1} - \sigma_2^{-1})(\frac{M_1}{\sigma_1} - \frac{M_2}{\sigma_2})}}{(\sigma_1^{-1} - \sigma_2^{-1})} \tag{4.28}$$

where σ_i and M_i, for $i = 1, 2$ are the standard deviations and mean values respectively.

The first experiment has considered three Gaussian distributions $N(2, 1)$, $N(5, 1)$ and $N(5, 3)$ whose T_{KS} between population 1 and population 2 as well as population 1 and population 3 are $T_{KS}^{1-2} = 3.5$ and $T_{KS}^{1-3} = 3.5582$ respectively. The corresponding Mahalanobis distances are $T_{MD}^{1-2} = 3.5$ and $T_{MD}^{1-3} = 3.9019$ respectively. The results confirm that the $T_{KS} = T_{MD}$ just in the case where the two Gaussian distributions have the same variance. Indeed from (4.10) one can state that:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-M_1)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-M_2)^2}{2\sigma^2}} \Leftrightarrow x = \frac{M_1 + M_2}{2} \tag{4.29}$$

which is equal to (4.27).

the second experiment is related to understanding how the distribution shapes

affect the performances of the Mahalanobis thresholds. In this experiment Exponential and Gaussian distributions with parameters described in (4.17) and (4.19) and Rayleigh as well as Gaussian with parameters described in (4.18) and (4.19) have been considered in order to check the behaviour of Mahalanobis distance in the case of different sample populations with the same variance values (E-G case) and with different variance values (R-G case). Comparing the results with values of Table 4.4, one can state that the Mahalanobis threshold is less efficient (i.e. it produces a bigger misclassification error) and it depends more on the standard deviation of populations than their mean values. Mahalanobis distance fails also in assessing the error.

	$\epsilon_{T_{MD}}$	$\epsilon_{T_{KS}}$
$T_{MD}^{E-G} = 3$	0.0726203	0.0725578
$T_{MD}^{R-G} = 2.8235$	0.0333	0.0328

Table 4.4: *Results of experiment between two populations drawn from different probabilities with the same variance values (E-G case) and with different variance values (R-G case). The error of misclassification is always smaller for T_{KS} (as reported in Table 4.1) than T_{MD}.*

4.5.6 KAPS

As for the experiments to validate the KAPS metrics, the experiments of section 4.5.3 are considered and the results are reported in Table 4.5. In

Experiment	KAPS
$E - G$	10^6
$R - G$	2.4253
$E^* - G$	10^6

Table 4.5: *Values of KAPS*

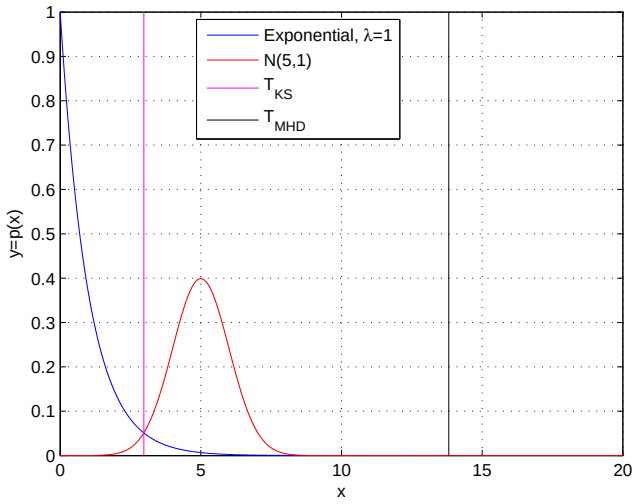


Figure 4.10: *Experiment Exponential* $\lambda = 1$ - *Gaussian* $N(5, 1)$

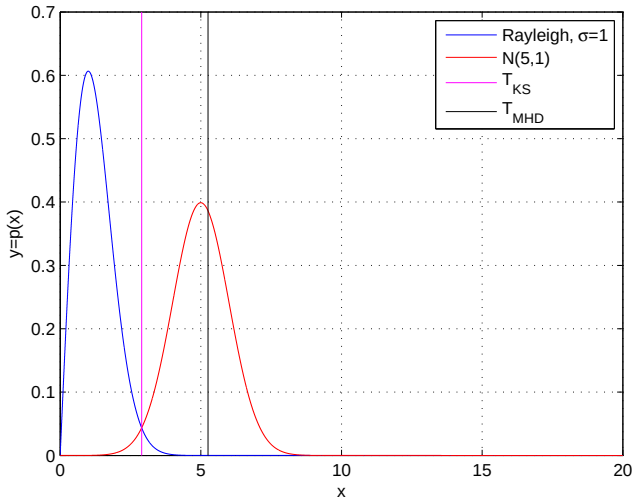


Figure 4.11: *Experiment Rayleigh* $\sigma = 1$ - *Gaussian* $N(5, 1)$

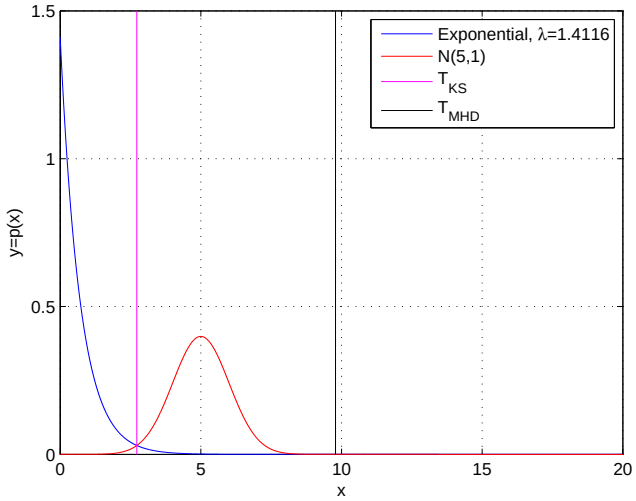


Figure 4.12: *Experiment Exponential* $\lambda = 1.4116$ - *Gaussian* $N(5, 1)$

Figure 4.10, Figure 4.11 and Figure 4.12 the three experiments are depicted. As shown, despite experiments R-G and E*-G have the same error areas, the KAPS value emphasizes that in case of exponential distributions, the desired maximum tolerated Type I error is reached if and only if almost the whole ω_2-class is misclassified as ω_1-class, whereas the AUC values are almost equal.

4.6 Summary

Chapter 4 focused on the comparison of the most popular class-separability metrics. Firstly the suitability of metrics which are based on the Covariance matrix have been analysed, then the criteria based on information theory (i.e. Entropy/Kullback-Leibler divergence, Bhattaccharrya and Chernoff distances), the methods based on thresholding (i.e. K-S threshold, Support

Vector Threshold and Mahalanobis threshold) and the graphical methods based on thresholding (i.e. ROC and AUC) have been studied. Finally a new metric has been introduced (i.e. KAPS) which presents many advantages compared with previous metrics in order to assess the degree of separability.

Chapter 5

Sample size effects

5.1 Introduction

In Chapter 4 the most common criteria adopted in pattern recognition for predicting the degree of class-separability were analysed. We also introduced a new criterion based on the Kolmogorov-Smirnov test which is important because it is able to manage the separability when a constraint about the maximum tolerated Type II error value is introduced. Kolmogorov appropriate prediction of separability is a criterion which is able to predict if two classes are separated or not and at what degree. It is based on the idea that the smaller the error area is, the more separated the classes are. As reported KAPS is defined as in equation (4.21), i.e. the distance between classes CDFs at T_{KS} and T_{MHD} threshold values.

The topic of this chapter is to understand how the value of the KAPS metric varies as function of sample size. The methods used for assessing the performance of metrics are order statistics analysis [113] and Monte Carlo [101]

simulations.

5.2 Order Statistics analysis

Order statistics analysis [113] is a fundamental tool in non-parametric statistics. It consists usually of computing the distribution of a sorted (in increasing order) set of the realizations of a random variable. A classical example of order statistics can be described as follows: Drawn n-samples independent and identically distributed $\{x_1, \ldots, x_n\}$ from a distribution with PDF $f(x)$ and CDF $F(x)$, find the distribution of the following random variables:

$$U = \max\{x_1, \ldots, x_n\} \tag{5.1}$$

$$V = \min\{x_1, \ldots, x_n\} \tag{5.2}$$

The problem of (5.1) can be solved easly by defining set of sample values as follows:

$$\begin{aligned} A_U &= \{U \leq u\} \text{ (the maximum is at most)}u \\ &= \{x_1 \leq u\} \cap \{x_2 \leq u\} \cap \ldots \cap \{x_n \leq u\} \end{aligned} \tag{5.3}$$

Therefore we obtain:

$$F(U) = \Pr\{U \leq u\} = \prod_{i=1}^{n} \Pr\{x_i \leq u\} = [F(u)]^n \tag{5.4}$$

$$f(U) = \frac{d}{du}[F(u)]^n = n f_x(u)[F(u)]^{n-1}. \tag{5.5}$$

The problem defined in (5.2) can be solved in a similar way.

Our problem however is different. We are indeed interested in computing the statistics of the j-th smallest value of the sample. Suppose that our x_i, $i = 1, \ldots, n$ is drawn from a continuous distribution with CDF $F_x(x)$ and PDF $f_x(x)$. The j-th order statistics is hence equal to:

$$\begin{aligned} F_{x_{(j)}}(x) = \Pr\left\{x_{(j)} \leq \tilde{x}\right\} = \{\text{Exactly } j \text{ samples of } x_i \leq \tilde{x}\} \cup \\ \cup \{\text{Exactly } j+1 \text{ samples of } x_i \leq \tilde{x}\} \cup \ldots \cup \\ \cup \{\text{Exactly } n \text{ samples of } x_i \leq \tilde{x}\} = \sum_{k=j}^{n} \binom{n}{k} [F_x(x)]^j [1 - F_x(x)]^{n-j} \quad (5.6) \end{aligned}$$

$$f_{x_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_x(x) [F_x(x)]^{j-1} [1 - F_x(x)]^{n-j} \quad (5.7)$$

i.e. (5.6) gives us the probability that a sample is placed at position j when a set of n-samples is ordered.

5.3 Monte Carlo simulation

The term 'Monte Carlo' was apparently first used by Ulam and von Neumann [101] as a Los Alamos code word for the stochastic simulations they applied to building better atomic bombs. The Monte Carlo method is an application of the laws of probability and statistics to the natural sciences. The essence of the method is to use various distributions of random numbers, each distribution reflecting a particular process in a sequence of processes such as the diffusion of neutrons in various materials, to calculate samples that ap-

proximate the real diffusion history. Statistical sampling had been known for some time, but without computers the process of making the calculations was so laborious that the method was seldom used unless the need was compelling. The computer made the approach extremely useful for many physics problems.

The spirit of Monte Carlo [101] is best conveyed by the example discussed in a letter of von Neumann to Richtmyer. Consider a spherical core of fissionable material surrounded by a shell of moderator material. Assume some initial distribution of neutrons in space and in velocity but ignore radiative and hydrodynamic effects. The idea is to now follow the development of a large number of individual neutron chains as a consequence of scattering, absorption, fission, and escape.

At each stage a sequence of decisions has to be made based on statistical probabilities appropriate to the physical and geometric factors. The first two decisions occur at time $t = 0$, when a neutron is selected to have a certain velocity and a certain spatial position. The next decisions are the position of the first collision and the nature of that collision. If it is determined that a fission occurs, the number of emerging neutrons must be decided upon, and each of these neutrons is eventually followed in the same fashion as the first. If the collision is decreed to be a scattering, appropriate statistics are invoked to determine the new momentum of the neutron. When the neutron crosses a material boundary, the parameters and characteristics of the new medium are taken into account. Thus, a genealogical history of an individual neutron is developed. The process is repeated for other neutrons until a statistically valid picture is generated.

As for Monte Carlo algorithms however, there exist many methods and they are useful for simulating complex systems, especially systems with many coupled degrees of freedom.

As for Monte Carlo method properties, they consists of the following pattern:

- Define a domain of possible inputs;
- Generate inputs randomly from a probability distribution over the domain;
- Perform a deterministic computation on the inputs;
- Aggregate the results.

From a more mathematical point view, consider a (possibly multidimensional) random variable X having probability mass function or probability density function $f_X(x)$ which is greater than zero on a set of values $\{X\}$ Then the expected value of a function g of X is:

$$\mathbb{E}(g(X)) = \sum_{x \in \{X\}} g(x) f_X(x) \tag{5.8}$$

Now, if we were to take an n-sample of X's, $(x_1, \ldots, x_n)$, and we computed the mean of $g(x)$ over the sample, then we would have the Monte Carlo estimate of the (5.8):

$$\tilde{g}_n(x) = \frac{1}{n} \sum_{i=1}^{n} g(x_i) \tag{5.9}$$

We could, alternatively, speak of the random variable

$$\tilde{g}_n(X) = \frac{1}{n} \sum_{i=1}^{n} g(X) \tag{5.10}$$

which we call the Monte Carlo estimator of (5.8).

If $\mathbb{E}(g(X))$, exists, then the weak law of large numbers tells us that for any arbitrarily small ϵ:

$$\lim_{n\to\infty} P(|\tilde{g}_n(X) - \mathbb{E}(g(X))| \geq \epsilon) = 0 \tag{5.11}$$

This tells us that as n gets large, then there is small probability that $\tilde{g}(X)$ deviates much from $\mathbb{E}(g(X))$. For our purposes, the strong law of large numbers says the same thing (the important part being that as long as n is large enough, $\tilde{g}_n(x)$) arising from a Monte Carlo experiment shall be close to E(g(X)), as desired. Moreover $\tilde{g}_n(x)$ is unbiased for $\mathbb{E}(g(X))$:

$$\mathbb{E}(\tilde{g}_n(x)) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(g(X_i) = \mathbb{E}(g(X)). \tag{5.12}$$

Therefore a Monte Carlo simulation becomes useful when one realizes that very many quantities of interest may be cast as expectations.

5.4 Experiment description

The main goal of this Chapter is to determine the performance of the KAPS metric as a function of sample size. Unfortunately two kinds of problems have arisen: first, the computation of order statistics (5.6) and (5.7) is quite hard in many cases, whereas Monte Carlo requires a high number of samples, which is very time consuming. In order to solve both problems the following criteria have been adopted: As for assessing values of T_{KS} and D_{KS} as function of samples, Monte Carlo analysis has been adopted, whereas for T_{MHD}

and D_{MHD} an analytical study has been performed.

The experiment considered consists in computing the value of KAPS when a set of m samples is considered ($m = 10,\ 10^2$ and 10^3). Firstly we computed the order statistics PDF for i.i.d. samples drawn from an exponential distribution:

$$\frac{n!}{(j-1)!(n-j)!}(1-e^{-x})^{j-1}(e^{-x})^{n-j+1} \tag{5.13}$$

Since for $x = MHD$, $e^{-x} \ll 1$, therefore (5.13) can be approximated as:

$$\frac{n-j+1}{n}\left[\frac{(n+1)!}{(j-1)!(n-j+1)!}x^{j-1}(1-x)^{n-j+1}\right] \tag{5.14}$$

by considering the Taylor approximation, stopped at first order (i.e. $f(x) = f(a) + f^{'}(a)(x-a)$), of e^{-x}. The term between bracket in (5.14) is a $Beta(j, n-j+2)$ distribution whose mean value and variance are respectively:

$$E_x = \frac{n-j+1}{n}\frac{j}{n+2} \tag{5.15}$$

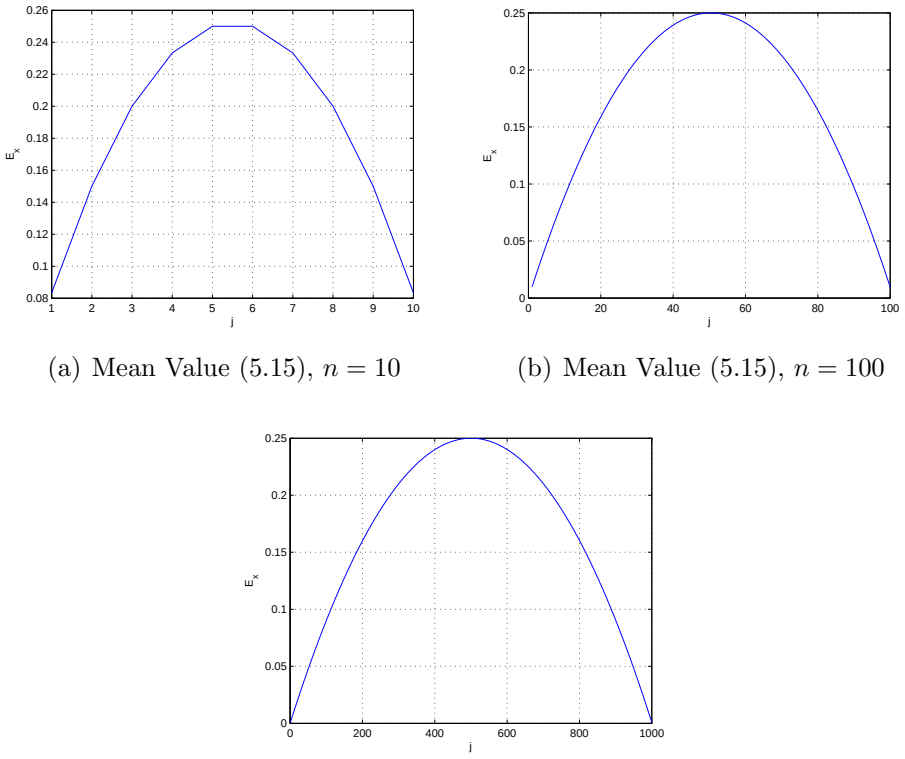
$$Var_x \leq \frac{n-j+1}{n}\frac{j(n-j+2)}{(n+1)^2(n+3)} \tag{5.16}$$

$$\tag{5.17}$$

the inequality of (5.16) is valid because $\frac{n-j+1}{n} \leq 1$ in (5.14). The behaviour of the described quantities is reported in Figure 5.1, where the mean value of (5.15) is reported as function of the order j, and Figure 5.2, where the variance value of (5.16) is reported as function of the order j. Because of the linear approximation, we are interested in the median value of j (i.e. we are interested in computing the order statistics at the approximation point which

is the median value of the approximated Beta distribution) which gives the highest outcomes, as expected, therefore we can assess for $j = n/2$ the value of the (5.16) (results for $n = 10,\ 100,\ 1000$): $2.7{\cdot}10^{-3}$, $2.5{\cdot}10^{-5}$ and $2.5{\cdot}10^{-7}$. Hence the good estimation of T_{MHD} with sampled data is given by a number of samples $n \geq 1000$ because the corresponding variance value of the order statistics is $\approx 10^{-7}$. Moreover the corresponding variation of distance D_{MHD} is negligible (i.e. simulated value of D_{MHD} for $n = 10,\ 100,\ 1000$: $\approx 10^{-6}$).

As for T_{KS} value, the (5.14) is no more valid, therefore only a numerical



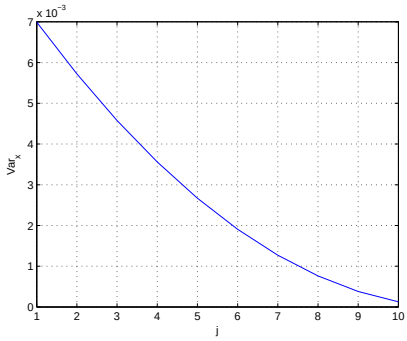
(a) Mean Value (5.15), $n = 10$

(b) Mean Value (5.15), $n = 100$

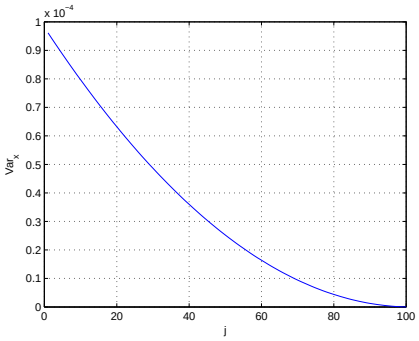
(c) Mean Value (5.15), $n = 1000$

Figure 5.1: *Mean value behaviour of (5.15) as function of j and for $n =$ 10, 100, 1000 number of samples. The linearized approximation of the order statistics presents a local maximum at the median value as expected.*

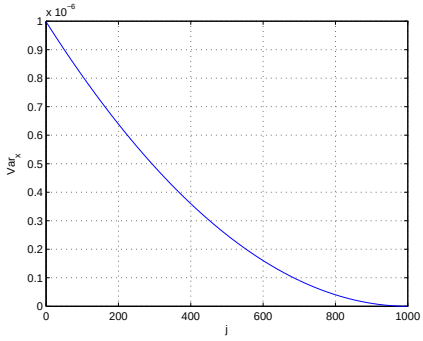
Monte Carlo simulation can be performed. As a consequence of 10^3 iterations



(a) Variance Value (5.16), $n = 10$



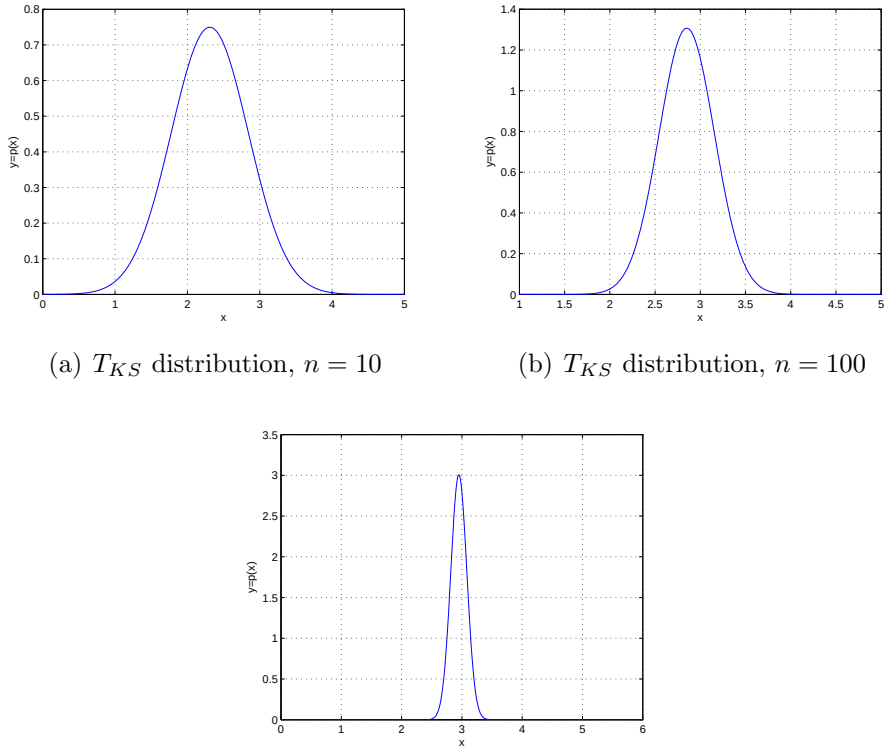
(b) Variance Value (5.16), $n = 100$



(c) Variance Value (5.16), $n = 1000$

Figure 5.2: *Variance value behaviour of (5.16)as function of j and for $n =$ 10, 100, 1000 number of samples. The value of variance is computed for $j = 5,\ 50,\ 500$ in order to satisfy the linearized approximation assumptions, i.e. the range over which the approximation is valid is centered around the threshold value.*

of our simulations (approximated with a Gaussian) the results are reported in Figure 5.3. Because the values of T_{MHD} are predominant, the KAPS measure



(a) T_{KS} distribution, $n = 10$

(b) T_{KS} distribution, $n = 100$

(c) T_{KS} distribution, $n = 1000$

Figure 5.3: *T_{KS} distribution as function of the number of samples used to compute the threshold. As expected, increasing the number of samples reduces the variance of the measurements (i.e. $\sigma_{meas.} = 1.5,\ 0.8,\ 0.04$, for $n = 10,\ 100,\ 1000$ number of samples).*

is reliable when at least 1000 samples are used.

5.5 Summary

Performances of KAPS metric have been investigated in Chapter 5. The analysis consisted in understanding how many samples are necessary in order to

get a reliable value of KAPS metric. The methods used for assessing the KAPS reliability have been the order statistics and Monte Carlo algorithm. Firstly a description of order statistics has been given and then an introduction of Monte Carlo methods have been described. Finally the results of the simulations have been reported.

Chapter 6

Metrics in ATR/SAR systems

6.1 Introduction

In ATR/SAR systems a crucial step is due to denoising algorithms which perform the task of removing noise from images and preserving the most important features (e.g. geometrical and statistical) of potential targets (Bhanu *et al.* in [56]). Besides it is also important to avoid introducing distortion which can affect the classifiers dramatically (i.e. artifacts), therefore it is necessary to evaluate algorithm performance by using a 'global' criterion: metric (i.e. a tool to assess the preservation of information during signal processing).

Several methods have been introduced by researchers to evaluate SAR images ([2], [9], [114], [115]), but no complete investigation of their suitability have been performed.

This chapter is addressed to test and analyse a set of metrics commonly used to compare SAR system performance in order to check if they are able to

satisfy all SAR/ATR despckling requirements (i.e. removing noise and preserve image features). Moreover the adopted algorithms have been chosen in order to remark what kind of information is lost and how to restore it, as the 'forward-backward' procedures stress.
To facilitate experimentation we have used synthetic data which are described in section 6.4

6.2 SAR Image Quality Metrics

In SAR systems several methods have been adopted by researchers during the last decades as *Image Quality Metrics* (**IQMs**), but the most important can be summarized as follows:

- PSF: Point Spread Function ([2], [9], [114])
- SNR: Signal-to-Noise Ratio ([2], [114])
- PSNR: Peak Signal-to-Noise Ratio ([2])
- ISLR: Integrated SideLobe Ratio ([2])
- MNR: Multiplicative Noise Ratio ([2], [116])
- CR: Contrast Ratio ([2])
- MSE: Mean Squared Error ([115])
- ENL : Equivalent Number of Looks ([9])

The first six methods are unitary metrics. but they may be computed for a single image, whereas *MSE* and *ENL* must always be calculated with respect

to a reference image. Furthermore, as described by Skolnik in [2], PSF, SNR, PSNR, ISLR are related to system performances (i.e. resolution, range etc. etc.) and they can be used if the signal is known or modeled.
Two IQMs introduced by Skolnik for direct image comparison [2] are MNR and CR. Multiplicative noise is proportional to the average scene intensity of the system whose main contributors are:

1. Integrated SideLobe energy;

2. Range and azimuth ambiguity;

3. Digital noise.

As a consequence MNR is defined as the ratio between *Non-Return-Area* (**NRA**) (i.e. noise-free parts of images such as calm lakes, metal slabs etc.) and the image intensity in a relatively bright surrounding area. The suitability of this criterion is related to the opportunity of finding such kinds of areas. A similar consideration can be argued for CR which is defined by Skolnik in [2] as the ratio between the average intensity of a typical bright region and the intensity of an NRA (if thermal noise is small compared with the signal, then one can write: $CR = 1/MNR$). Another problem we encountered is the difficulty of simulating these image areas because they produce nonsense results.
Mean Squared Error (MSE) is one of the commonly used performance measures in image and signal processing. For an image on size $N \times M$ pixels, it

can be defined as:

$$MSE = \frac{1}{MN} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} (x[n,m] - \hat{x}[n,m])^2 \tag{6.1}$$

where $x[n,m]$ is the original image pixel, whereas $\hat{x}[n,m]$ represents the processed one. This metric assumes that the distortion is caused by only additive, image independent noise. Unfortunately in SAR images a multiplicative model is used [9].

As described in [9], besides, because of the multiplicative nature of SAR image noise, a natural proposal for a metric is to adopt the *Equivalent Number of Looks* (**ENL**), defined as:

$$ENL = \frac{(mean)^2}{variance} \tag{6.2}$$

In the first generations of SAR systems, data were captured at low resolution so that objects [9], such as trees and houses, were much smaller than a resolution cell. Contributions from RCS fluctuations were averaged out so that no spatial variations or correlation effects were visible. The resultant complex field PDF was then Gaussian, with Rayleigh amplitude and negative exponential intensity PDFs, respectively. As a consequence in low resolution data, the L-looks average intensity (average of L antenna sub-apertures or pixels) obey a Gamma distribution with order parameter L, given by:

$$p_I(I) = \frac{1}{\Gamma(L)} \left(\frac{L}{\sigma_0} \right)^L I^{L-1} e^{LI/\sigma_0}, \quad I \geq 0 \tag{6.3}$$

(where I is the image intensity, Γ is the Gamma function and σ_0 is the mean value of RCS) which has moments:

$$\langle I^m \rangle = \frac{\Gamma(m+L)}{\Gamma(L)} \left(\frac{\sigma_0}{L}\right)^m \tag{6.4}$$

Note that equation (6.2) is derived directly by (6.4). Indeed L is the number of looks of SAR images, therefore if the second moment of intensity and its variance are computed by considering $\langle I \rangle = \sigma_0$ as mean value of image intensity, one can write (by considering properties of the Gamma function: $\Gamma(z+1) = z\Gamma(z)$):

$$\left\langle I^2 \right\rangle = \frac{L+1}{L}\sigma_0^2 \tag{6.5}$$

$$\left\langle I^2 \right\rangle - \langle I \rangle^2 = \frac{L+1}{L}\sigma_0^2 - \sigma_0^2 = \frac{\sigma_0^2}{L} \tag{6.6}$$

the ratio defined in equation (6.2) the number of look L is obtained. This is why ENL means *Equivalent Number of Look*.

An important characteristic of this metric is that it is equivalent to the number of independent intensity values averaged per pixel. Furthermore the averages of the ENL ratio are carried out in intensity over a uniformly distributed target, otherwise the condition of Gamma distributed intensity is not satisfied.

6.3 NVM : Normalized Variance Metric

Using high resolution SAR, the exponential distribution does not fit the intensity distribution, therefore researchers have introduced several new dis-

tribution models such as Weibull, Log-normal and K, because of their ability to fit real data [9] (especially the last one). Indeed, under the assumption of Gamma-distributed RCS (order parameter ν) and multilook speckle (order parameter L), the intensity, I, has a K-distribution PDF, as defined in [9]:

$$P(I) = \frac{2}{\Gamma(L)\Gamma(\nu)} \left(\frac{L\nu}{\langle I \rangle}\right)^{(L+\nu)/2} \times I^{(L+\nu-2)/2} K_{\nu-L}\left[2\left(\frac{\nu L I}{\langle I \rangle}\right)^{1/2}\right] \tag{6.7}$$

where L is the number of the looks, ν is the order parameter of RCS gamma distribution, Γ is the Gamma function, K is the modified Bessel function of second kind and $\langle I \rangle$ is the mean value of image intensity over an homogeneous local region of the image.

By rewriting normalized variance of equation (6.7) (i.e. $var(I)/\langle I \rangle^2 = 1/L + 1/\nu + 1/L\nu$, as defined in [9]) one can get straightforward a metric as follows (termed *Normalized Variance Metric*, (**NVM**)):

$$L' = \frac{\nu' + 1}{\frac{var(I)}{\langle I \rangle^2}\nu' - 1} \tag{6.8}$$

where L' and ν' are the estimated values of the numbers of look and the order parameter of RCS Gamma distribution respectively.

The most important advantage of this metric is to relate a parameter of the SAR system (i.e. the number of looks L), which is known by the designer, with statistical parameters that can be acquired by image processing. In an ideal case a despeckling algorithm should be able to suppress noise and preserve the image RCS values. As known indeed, despeckling and RCS are

both approximated by a Gamma distribution as described in [9], thus it is necessary to estimate the value of ν (given by $var\ \sigma/\langle\sigma\rangle^2 = 1/\nu$) and by the equation (6.8) compare it with the system number of looks. The more accurate is the estimation of ν, the smaller should be the error between the actual number of looks and the estimated one.

A very important problem is to simulate a set of 10^6 K-distributed SAR images (as the MSTAR data are classified and therefore they are not available for assessing the suitability of despeckling metrics) in order to compute the PDF distributions of the analysed metrics. Several methods are present in the literature, such as [9] and [10] for instance, but they are not able to simulate all the statistical properties of a K-distribution (e.g. Oliver's method in [9] is not able to preserve the property that the mean value of RCS is equal to the mean value of the K-distribution, whereas Ward's algorithm in [10] uses independent samples of unit power Rayleigh distributed noise). A very interesting alternative is to formulate a K-distribution by what is termed the *product model*, as described in [117].

By modelling RCS (σ) as a Gamma-distributed random variable σ with a probability density function $p_\sigma(\sigma)$ as in (6.9) (where $\Gamma(\cdot)$ is the gamma function, $\mu = \langle\sigma\rangle = \sigma_0$ is the mean, and ν is the order parameter as well as the variance of σ is equal to μ^2/ν), whereas $L-$look speckle random variable of mean $\langle z\rangle = 1$ with probability density function given by (6.10), the product of random variables $x = \sigma z$ is K-distributed with a probability density function given by (6.7).

$$p_\sigma(\sigma) = \frac{1}{\sigma}\left(\frac{\nu\sigma}{\mu}\right)^{\nu}\frac{1}{\Gamma(\nu)}\exp\left(-\frac{\nu\sigma}{\mu}\right) \tag{6.9}$$

$$p_z(z) = \frac{L^L z^{L-1} e^{Lz}}{\Gamma(L)} \tag{6.10}$$

As for our simulations, a single look radar system with a SAR Point Spread Function (**PSF**), defined by 2-D $sinc(x)$ filter, which produces a correlated exponential distributed speckle and as a consequence a correlated K-distributed set of images was considered. However the correlated Gamma-distributed RCS was simulated by using the method described by Ward in [10]. Indeed by generating a set of correlated Gaussian samples $\{y^1\}$ which are non-linearly mapped into a Gamma distributed sample by solving the following equation:

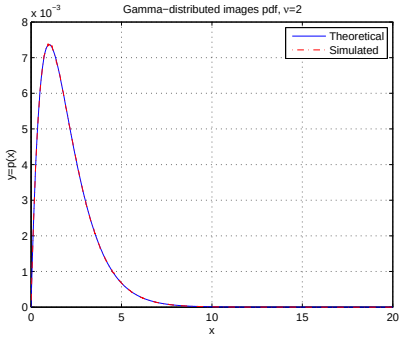
$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y^1}\exp\left[-\frac{y^2}{2}\right]dy \equiv \int_{-\infty}^{\sigma^1} b^\nu \frac{\sigma^{\nu-1}}{\Gamma(\nu)}\exp\left[-b\sigma\right] \tag{6.11}$$

($b = \nu/\mu$ is the scale parameter of the Gamma PDF of (6.9)) it is possible to generate a Gamma-distributed image.

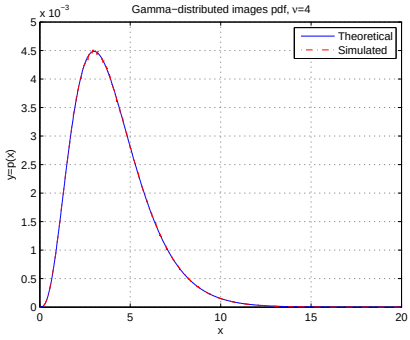
Finally a K-distributed image is created by multiplying the two variables in (6.9) and (6.10) was abtained.

By considering a scale parameter $b = \nu/\mu$ equal to 1 and using $MATLAB^{(TM)}$ tools, we computed a set of 1000, 100 by 100 pixels, correlated spatially Gamma-distributed images for each $\nu = 2, 4, 6, 8, 10$ (as depicted in Figure 6.1).

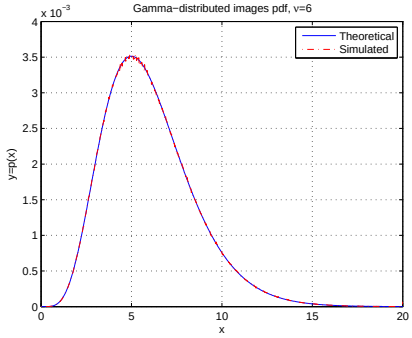
As described in Table 6.1 the Gamma-distributed images are simulated with an accuracy of 10^{-10} which is appreciable for our targets. Similar re-



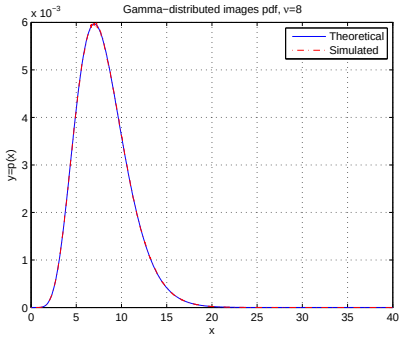
(a) Gamma-distributed images PDF, $\nu = 2$



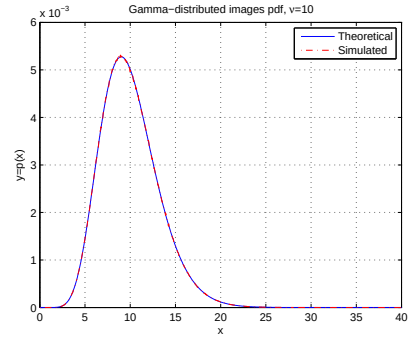
(b) Gamma-distributed images PDF, $\nu = 4$



(c) Gamma-distributed images PDF, $\nu = 6$

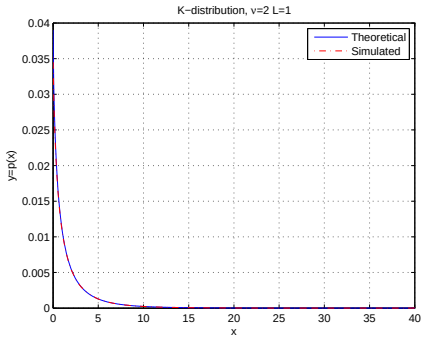


(d) Gamma-distributed images PDF, $\nu = 8$

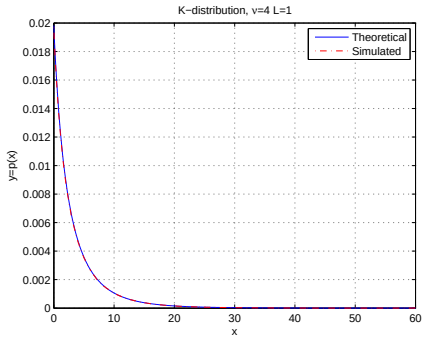


(e) Gamma-distributed images PDF, $\nu = 10$

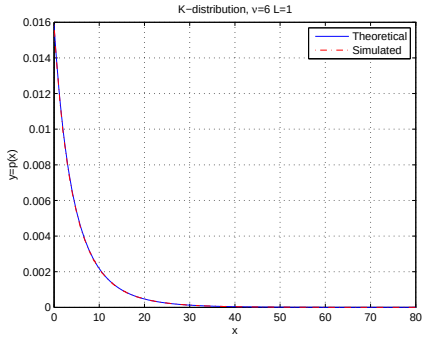
Figure 6.1: *Gamma-distributed images PDF: Simulated and theoretical. Comparison of PDFs of simulated Gamma-distributed images and theoretical Gamma-PDF for* $\nu = 2, 4, 6, 8, 10$.



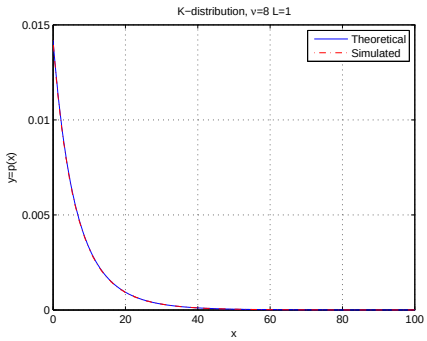
(a) K-distributed images PDF, $\nu = 2$



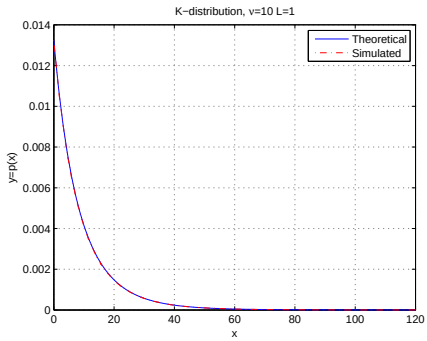
(b) K-distributed images PDF, $\nu = 4$



(c) K-distributed images PDF, $\nu = 6$



(d) K-distributed images PDF, $\nu = 8$



(e) K-distributed images PDF, $\nu = 10$

Figure 6.2: *K-distributed images PDF: Simulated and theoretical. Comparison of PDFs of simulated K-distributed images and theoretical K-PDF for* $\nu = 2, 4, 6, 8, 10$ *and single look* ($L = 1$).

ν	2	4	6	8	10
MSE	$9.15\ 10^{-11}$	$1.15\ 10^{-10}$	$1.048\ 10^{-10}$	$1.11\ 10^{-10}$	$1.1\ 10^{-10}$

Table 6.1: *Mean squared error of simulated Gamma-distribution with theoretical one (ν is the order parameter).*

Order parameter	$\nu = 2$	$\nu = 4$	$\nu = 6$	$\nu = 8$	$\nu = 10$
$\langle\sigma\rangle$	2.0002	3.9988	6.0033	8.0028	9.9991
$\langle\sigma\rangle^2 / var(\sigma)$	2.0029	4.0011	5.9869	8.0011	9.9893

Table 6.2: *Mean value and normalized variance of simulated Gamma-distributions. Because the shape parameters was set equal to one a mean value equal to the order parameter as well as the inverse-normalized variance should be obtained.*

ν	2	4	6	8	10
MSE	$4.27\ 10^{-9}$	$1.72\ 10^{-9}$	$6.45\ 10^{-10}$	$5.5\ 10^{-10}$	$4.3\ 10^{-10}$

Table 6.3: *Mean squared error of simulated K-distribution with theoretical one. Accuracy increases with order parameter (ν is the order parameter, whereas the Number of Look $L = 1$)*

ν	2	4	6	8	10
$\langle I\rangle = \langle\sigma\rangle$	2.0009	4.0001	6.0043	8.0036	9.9991
$var(I)/\langle I\rangle$	2.0125	1.5107	1.3450	1.2604	1.2111

Table 6.4: *Mean value and normalized variance of simulated K-distribution. The normalized variance for a single look K-distributed is equal to $1 + 2/\nu$, where ν is the order parameter of Gamma-distributed RCS (ν is the order parameter, whereas the Number of Look $L = 1$)*

sults are shown in Table 6.3 for K-distributions (depicted in Figure 6.2). As for Table 6.4, it represents the mean value of K-distribution which has to be equal to the mean value of the Gamma-distributed RCS and the normalized variance, defined as $1 + 2/\nu$ for single look SAR images (where ν is the order parameter of Gamma-distributed RCS). Comparing the results of Table 6.4 with results shown in Table 6.2, one can state that our simulation errors are smaller than 1% (an example of K-distributed image is reported in Figure 6.3(a)).

6.4 Nature of experiments

The experiments are implemented in order to determine properties of the following metrics: ENL , MSE and NVM . In practice they were implemented in order to check if the described metrics are able to discriminate the accuracy in RCS reconstruction of algorithms used for despeckling.

6.4.1 Algorithm description

Three algorithms were compared in order to estimate their performances in terms of RCS reconstruction and edge preserving properties. Three algorithms were analysed and their performances compared: Median filter, Beltrami flow and Gaussian filter. As for metrics comparison, only the Gaussian filter and Beltrami flow were taken into account.

Median filter

Median filter is a nonlinear digital filter technique for removing noise and it is often used because, under certain conditions, it preserves edges while removing noise [115]. The algorithm of the median filter consists of replacing the pixel which is being analysed with the median of the neighboring entries. For example, if one has the following 1-D vector $x = [2.5\ 80.4\ 6.3\ 3.1]$ and we use three cells window size, the median filter works as follows:

- $y[1] = Median[2.5\ 2.5\ 80.4] = 2.5$
- $y[2] = Median[2.5\ 80.4\ 6.3] = Median[2.5\ 6.3\ 80.4] = 6.3$
- $y[3] = Median[80.4\ 6.3\ 3.1] = Median[3.1\ 6.3\ 80.4] = 6.3$
- $y[3] = Median[6.3\ 3.1\ 3.1] = 3.1$

thus the output is $y = [2.5\ 6.3\ 6.3\ 3.1]$. Because there is no entry preceding the first value, the latter is repeated, as with the last value, to obtain enough entries to fill the window.

As for a 2-D algorithm, the median filter acts as follows: a $n-by-n$ matrix is converted in an n^2-cells array sorted in increasing order and then the previous algorithm is applied. In our simulation the median filter is implemented by $MATLAB^{TM}$ with the command *medfilt2*, with a window of $3-by-3$ pixels, was used (an example of a filtered image is reported in Figure 6.3(d)).

Gaussian filter

As described in [115], a Gaussian filter is defined as:

$$h(m,n) = \frac{1}{2\pi s^2} e^{-(m^2+n^2)/2s^2} \tag{6.12}$$

Equation (6.12) represent a 2-D convolution operator that is used to '*blur*' images and remove details and noise, which uses a kernel that represents the shape of a Gaussian ((an example of a filtered image is reported in Figure 6.3(b))).

The Gaussian filter output is a 'weighted average' of the central pixel of the filter window with its neighborhood (the filter window size depends on the standard deviation s in equation (6.12)). Because of weighted spatial averaging, the Gaussian filter provides a gentler smoothing and preserves edges better than a similarly sized mean filter, where the window has an uniform weighting.

Moreover researchers usually adopt the Gaussian as a smoothing filter because of its frequency response. Indeed Gaussian filter represents a bell-shape response in the frequency domain, thereby it does not exhibit oscillations in its frequency response, hence it is often used in the denoising step of edge detection techniques (i.e. it does not introduce artefacts due to ripplesin the frequency domain).

Beltrami flow

Kimmel *et al.* have introduced in [118], [119] and [120] a set of new denoising algorithms, termed Beltrami flow, which are capable of excellent smoothing

of images while preserving their visually important features (an example of a filtered image is reported in Figure 6.3(c)).

An interesting approach consists in introducing a Gaussian kernel to implement vector-valued image regularization. This kernel enables the implementation of Beltrami flow by convolving the image with the kernel, as defined in [124]:

$$K(u^1, u^2, \tilde{u}^1, \tilde{u}^2; t) = \frac{H_0}{t} \exp\left(-\frac{d_g^2((u^1, u^2), (\tilde{u}^1, \tilde{u}^2))}{4t}\right) \tag{6.13}$$

where (u^1, u^2) and $(\tilde{u}^1, \tilde{u}^2)$ are coordinates of two points on a manifold, $d_g^2(\cdot, \cdot)$ represents the geodesic distance between two points on image manifold, H_0 is a constant and t is the iteration step. As a consequence the update step for jointly smoothing the manifold and the image painted on it is:

$$X^i(u^1, u^2; t_0 + t) = \int\int_{(\tilde{u}^1, \tilde{u}^2) \in N(u^1, u^2)} X^i(\tilde{u}^1, \hat{u}^2; t_0) \times K(u^1, u^2, \tilde{u}^1, \tilde{u}^2; t) d\tilde{u}^1 d\tilde{u}^2 \tag{6.14}$$

where $X^i \in \left\{X^1, X^2, \ldots, X^N\right\}$ are the components of the images (i.e. Red, Green and Blue for colour image), whereas $N(u^1, u^2)$ is the neighborhood of the point $(\tilde{u}^1, \tilde{u}^2)$.

The difference between equation (6.12) and (6.14) consists in two different systems of reference. As for equation (6.12) the output is the sum of the neighboring points' amplitude weighted according to their distance along the coordinate axis. Beltrami flow however tends to smooth regular and flat regions, where $d_g^2(\cdot, \cdot)$ tends to be equal to zero and as a consequence the kernel

value is almost one , whereas it penalizes dramatically edge updates, where the value of the kernel tends to zeros because $d_g^2(\cdot,\cdot) >> 1$. Moreover, because of the nonlinearity of Beltrami flow (it depends on the original image), it is impossible to use a global kernel, therefore Spira *et al.* in [124] suggest to use the short-time kernel iteratively, therefore the number of iterations has been set at 2. Beltrami has never been adopted as despeckling algorithm by researchers, therefore it will be analysed as RCS filter reconstruction and as edge preserving denoising algorithm.

6.5 Results

The experimental process can be described as follows:

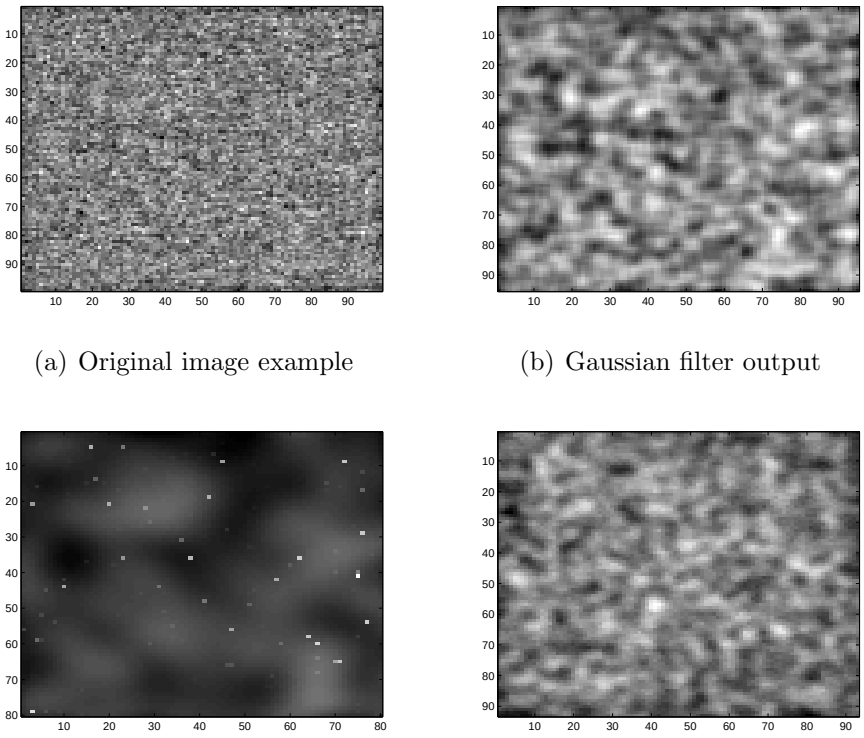
1. A set of 10^6 100×100 pixels spatially correlated K-distributed images have been generated;

2. The images are filtered using either the Gaussian and Beltrami methods in turn and the metric values are estimated (set the number of looks equal to 1, for our case);

3. The histograms of the metrics are computed;

A set of example images are depicted in Figure 6.3: First the original images is shown where others represent filtered examples (Gaussian filter and Beltrami flow outcomes respectively).

As for metric performances, results are depicted in Figures 6.4, 6.5, 6.6, 6.7 and 6.8, which represent the PDFs of metrics ENL , NVM and MSE respectively. Since the test images sets consist of homogeneous images, the



(a) Original image example

(b) Gaussian filter output

(c) Beltrami flow output

(d) Median filter output

Figure 6.3: *K-distributed images examples. A spatially correlated K-distributed image (originated from a spatially correlated Gamma distributed RCS, i.e. the underlying RCS consists of exponentially correlated samples, as Ward's MNTL model is reported in [10]), order parameter $\nu = 8$ and single look ($L = 1$) is generated (a) and then processed with algorithm the Gaussian filter (b), Beltrami flow (c) and Median filter (d).*

results are interesting in order to demonstrate which algorithm removes more speckle/reconstructs RCS in the best way. Reconstruction is important for two reasons: First it is important to remove as much speckle as possible while preserving images features (i.e. understanding which algorithm performs the best RCS reconstruction and how the denoising/RCS reconstruction process affects the Constant False Alarm (**CFAR**) detection performance for a SAR/ATR system); Second it is also crucial to estimate how much the output images are distorted by the denoising algorithm (i.e. quantifying the presence of artifacts). From these perspectives, ENL and NVM can be considered as dual. ENL indeed tends to quantify how much speckle is removed by the denoising filters (i.e. how similar the distribution of the ratio images to the order parameter of the Gamma-distributed speckle model is. Note that for our simulation the number of looks L is equal to 1, therefore equation (6.10) of the ratio image set becomes an exponential distribution). NVM numbers can be interpreted however as an indirect estimation of the presence of artifacts.

By inspection of Figures 6.4, 6.5, 6.6, 6.7 and 6.8 it is also clear that ENL produces a better separation of filter performance, whereas NVM and MSE introduce a superposition which produces performance ambiguities (nevertheless the $Mann - Whitney - Wilcoxon$ test (**MWW**) [133] confirms the results depicted in Figure 6.4(b), 6.5(b), 6.6(b), 6.7(b) and 6.8(b), as reported in Table 6.5: the Beltrami NVM PDFs are shifted to the right with respect to the Gaussian NVM PDFs. Note that only values of NVM distributions which are $\geq 10^{-6}$ have been considered). Moreover MSE computed in linear space cannot give useful information because it is not related to any

Order parameter	$NVM_{Belt}<NVM_{Gauss}$	$NVM_{Belt}>NVM_{Gauss}$
$\nu = 2$	0.9999	$1.3652e-04$
$\nu = 4$	0.9973	0.0027
$\nu = 6$	0.9945	0.0055
$\nu = 8$	0.9952	0.0048
$\nu = 10$	0.9642	0.0358

Table 6.5: *MWW test results, the p-values for two cases (Population A < Population B and Population A > Population B) of significance test (the null hypothesis is rejected when the p-value is less than 0.05 or 0.01, corresponding respectively to a* 5% *or* 1% *chance of rejecting the null hypothesis when it is true - Type I error) are reported. No equal values in two columns means that the distributions are different*

model parameter (i.e. number of looks of the SAR system/other parameters of speckle model for NVM and ENL respectively). A large value of MSE indicates a large difference between images, but it may not be related to reduction of speckle.

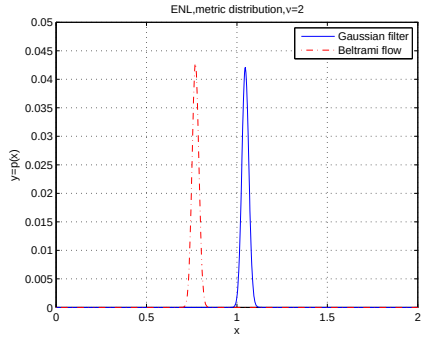
$$\lim_{\nu \to \nu_{true}} NVM = L \tag{6.15}$$

As defined in (6.2) and (6.15), an ideal algorithm should give the same result for ENL and NVM (i.e. the actual number of looks $L = 1$ in our case), therefore one can confirm that neither Beltrami flow nor the Gaussian filter are able to remove correctly the speckle and to reconstruct the underlying RCS in the samples images. Indeed by increasing the order parameter ν, ENL tells us that for very noise-like images ($\nu = 2$ which is very similar to woodland distribution as reported in [9]) the Gaussian filter removes speckle correctly and it reconstructs RCS PDF in the best way. However for the other experiments, Beltrami flow seems to remove more correctly the speckle

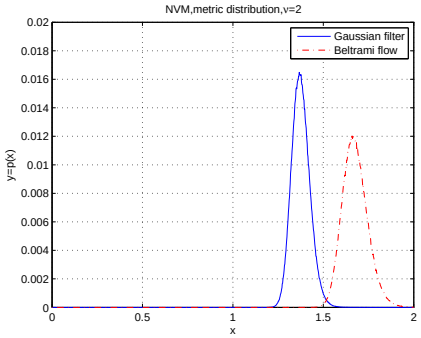
than the Gaussian filter, because the latter tends to smooth too much the images, as expected.

In terms of NVM, the Gaussian filter introduce less distortion in RCS reconstruction than an iterative algorithm such as Beltrami flow.

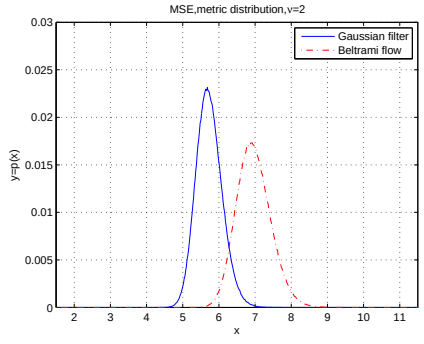
In conclusion it is suggested to use both NVM and ENL metrics to compare algorithm performance and to estimate the presence of artifacts in the images which could affect the detection and classification steps of an SAR/ATR system. Indeed ENL allows us to quantify how good the Gamma-distributed speckle model is estimated, whereas NVM gives us the same information for Gamma-distributed RCS.



(a) ENL metric distribution, $\nu = 2$

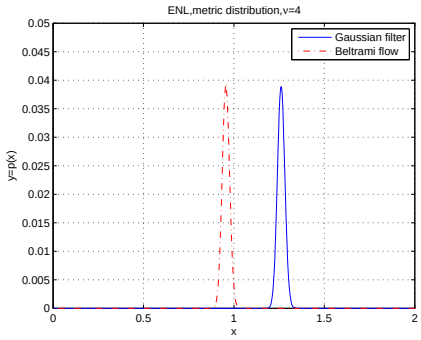


(b) NVM metric distribution, $\nu = 2$

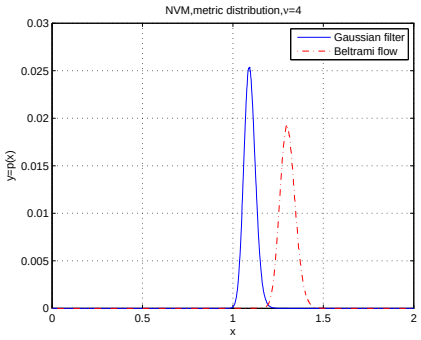


(c) MSE metric distribution, $\nu = 2$

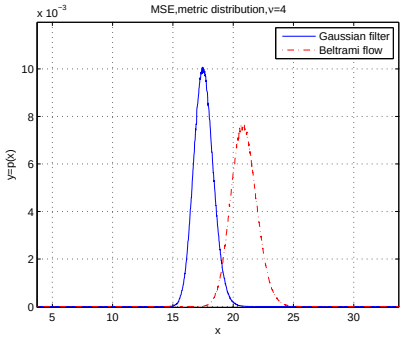
Figure 6.4: *Metrics distributions for Gaussian filter and Beltrami flow when applied to a K-distributed images set with* $\nu = 2$ *and* $L = 1$.



(a) ENL metric distribution, $\nu = 4$



(b) NVM metric distribution, $\nu = 4$

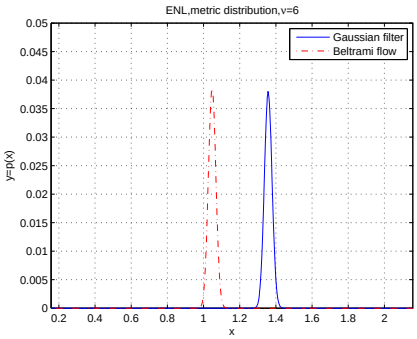


(c) MSE metric distribution, $\nu = 4$

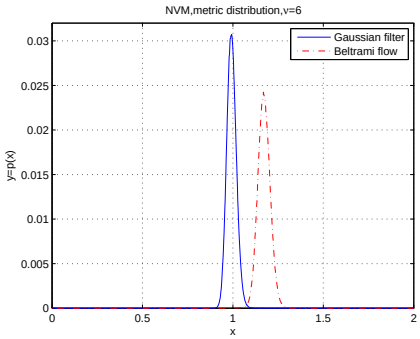
Figure 6.5: *Metrics distributions for Gaussian filter and Beltrami flow when applied to a K-distributed images set with* $\nu = 4$ *and* $L = 1$.

6.6 Summary

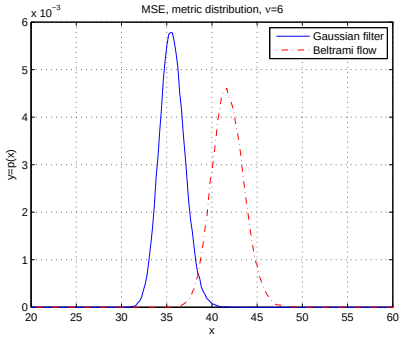
The main topic of Chapter 6 has been the analysis of the most important image quality metrics for ATR/SAR systems. An introduction of the most popular metrics adopted by researchers has been given and as a consequence metrics suitability for ATR/SAR systems have been analysed. In order to perform such an analysis some issues related to SAR images simulation have been resolved. Moreover a set of algorithms for image denoising have been described (i.e. the median filter, The Gaussian filter and Beltrami flow).



(a) ENL metric distribution, $\nu = 6$

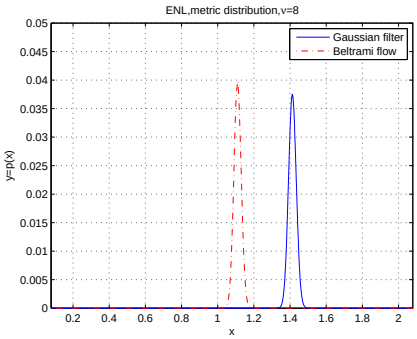


(b) NVM metric distribution, $\nu = 6$

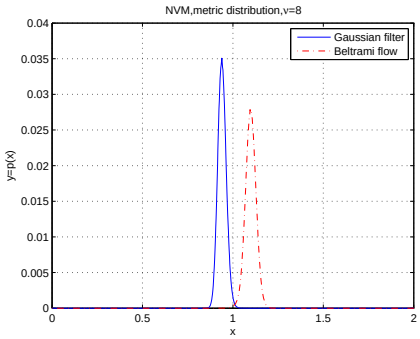


(c) MSE metric distribution, $\nu = 6$

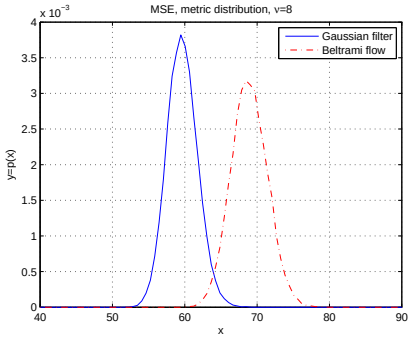
Figure 6.6: *Metrics distributions for Gaussian filter and Beltrami flow when applied to a K-distributed images set with* $\nu = 6$ *and* $L = 1$.



(a) ENL metric distribution, $\nu = 8$

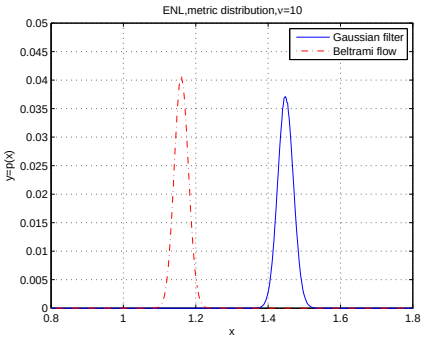


(b) NVM metric distribution, $\nu = 8$

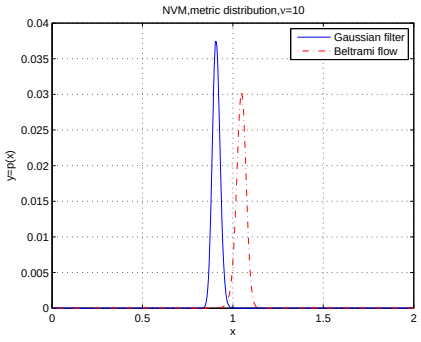


(c) MSE metric distribution, $\nu = 8$

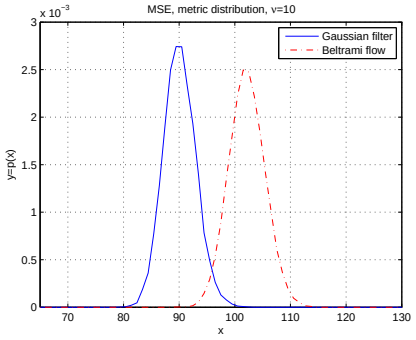
Figure 6.7: *Metrics distributions for Gaussian filter and Beltrami flow when applied to a K-distributed images set with* $\nu = 8$ *and* $L = 1$.



(a) ENL metric distribution, $\nu = 10$



(b) NVM metric distribution, $\nu = 10$



(c) MSE metric distribution, $\nu = 10$

Figure 6.8: *Metrics distributions for Gaussian filter and Beltrami flow when applied to a K-distributed images set with* $\nu = 10$ *and* $L = 1$.

Finally a new image quality metric (i.e. NVM) has been proposed because of its advantage in estimating the presence of artefacts in filtered images/RCS reconstruction filter output.

Chapter 7

Gaussian filter, Median filter and Beltrami flow: Analysis and comparison in ATR systems

7.1 Introduction

One of the most important tasks in a SAR/ATR system is to remove correctly the speckle from the images (i.e. removing the speckle and preserving target features as well as avoid introducing artifacts). There are many algorithms which have been adopted in the last decades as reported by Oliver in [9]. Recently a new set of algorithms has been developed by Spira *et al.* which tend to satisfy most of the requirements of digital processing (i.e. denoising/smoothing trade-off). These new algorithms have originated from studies

of high energy physics and use an important mathematical instrument: The Beltrami operator. It is proved by Spira *et al.* that Beltrami flow, described in 6.4.1, is a generalization of many algorithms based on partial differential equations (such as linear scale space, generalized Peona-Malik flows, mean curvature flow etc. as described in [119]) and it seems to be able to perform a good RCS reconstruction as well as preserves the edges. In this chapter the Beltrami flow algorithm and Median filter are analysed, and their performances are compared with the Gaussian filter which is the most popular noise filter in image processing.

7.2 Problem description

In Automatic Target Recognition (**ATR**) feature preserving is a crucial requirement in order to perform a correct classification. Hence despeckling algorithms should be able to reduce the noise level, perform a correct RCS reconstruction, preserve the edges and last, but not the least, not introduce any kind of artifacts.

The metrics used to compare algorithm performances are grouped in two sets: NVM as well as ENL , as defined in equation (6.2) and (6.8) as for denoising performance/RCS reconstruction and presence of artifacts, and edge preserving metrics.

As for the edge preserving metrics, they consists of several criteria to estimate how much the edges are preserved or distorted. In the literature, [115], the most important methods to evaluate edge preservation are:

1. Edge position: how much the position of the edge is shifted by the

despeckling algorithm;

2. Edge height: estimate the edge amplitude distortion introduced by the algorithm;

3. Slope angle of the edge;

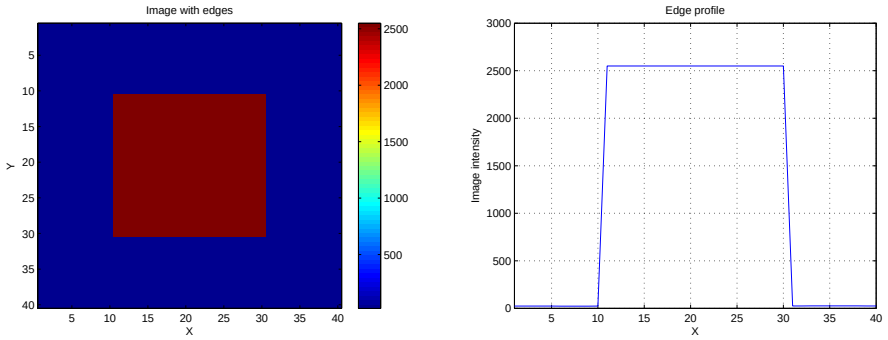
4. Spatial orientation on the edge introduced by the denoising algorithm.

7.3 Simulator description

By using the simulations performed in Chapter 6, a set of 1000, 100 by 100 pixels, K-distributed images was simulated.
However for testing of edge preserving metrics, a set of 1000, 100 by 100 pixels, K-distributed images with parameters $L = 1$ (number of look) and $\nu = 8$ (Gamma distributed RCS order parameter) was simulated and a square of 20 by 20 pixels was placed at the centre of the images (as reported in Figure 7.1(a) and 7.1(b)). In order to obtain a clear SAR image, the SCR of this images was set to $5dB$, as indicated by Skolnik in [2].

7.4 Experiment description and result analysis

Two sets of experiments have been performed in order to compare the performance of despeckling algorithms: First the algorithms have been analysed in terms of RCS filter reconstruction, then their edge-preserving properties



(a) Image with edges, the red part represent a 20×20 prevailing pixels

(b) Edge profile of simulated images with edges

Figure 7.1: *Examples of edge profile images: example of a simulated image with edge 7.1(a); Corresponding edge profile 7.1(b).*

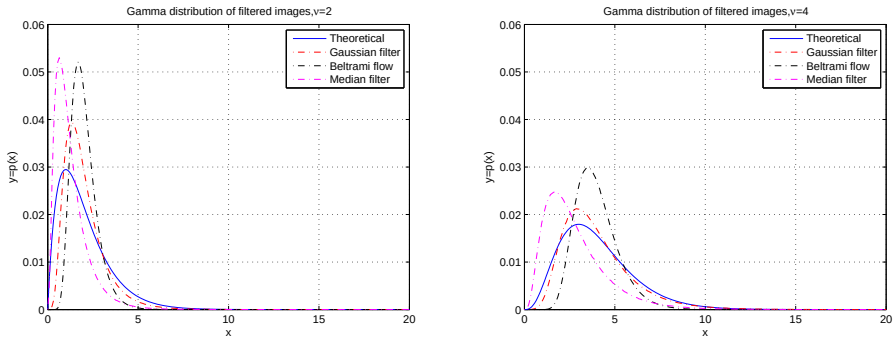
have been evaluated.

7.4.1 RCS reconstruction filter performances

As for RCS filters reconstruction, the algorithms have been compared in terms of ENL and NVM. Besides in order to check which metric performs the most realistic result, the PDF have been computed from the filter output and they have been compared in terms of MSE with the expected theoretical one (i.e. the underlying Gamma-distributed RCS).

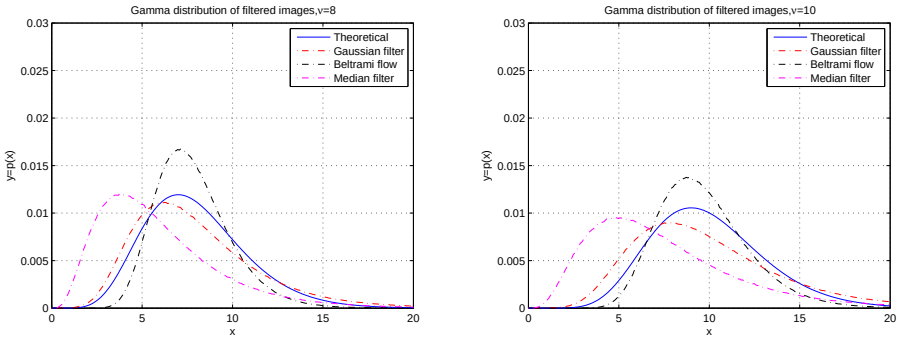
As for ENL and NVM values they are the mode of PDFs computed in chapter 6.

In Figure 7.2 and Figure 7.3 the outcomes of filters are depicted in terms of PDFs (Gamma-distributed RCS and exponential distributed removed speckle respectively). As for RCS reconstruction, it is clear that no filter is able to estimate correctly the RCS, it is clear that no filter is able to estimate



(a) RCS distribution filtered images, $\nu = 2$

(b) RCS distribution filtered images, $\nu = 4$



(c) RCS distribution filtered images, $\nu = 8$

(d) RCS distribution filtered images, $\nu = 10$

Figure 7.2: *Gamma distribution of filtered images, for* $\nu = 2, 4, 8, 10$

Metric	Beltrami flow	Gaussian filter	Median filter
NVM	1.666	1.367	0.9359
ENL	0.77	1.046	0.4080

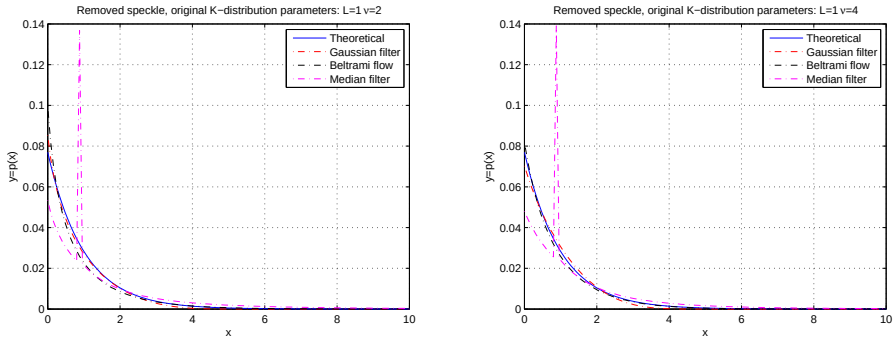
Table 7.1: *Metrics values of RCS reconstruction for filtered images,* $\nu = 2$

correctly the RCS PDF especially for low order parameter of the underlying RCS. Moreover for $\nu = 2$ the indirect measures tend to fail.

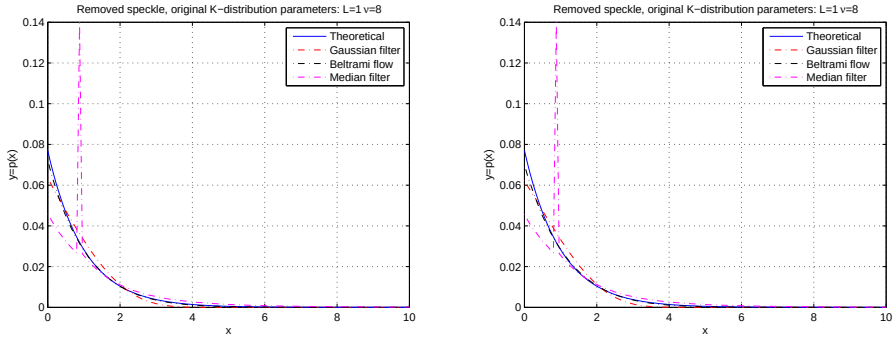
By comparing Figure 7.2(a) with values of Table 7.1 and 7.5, the minimum of MSE is given by the Gaussian filter output despite the NVM values seems to give the best result for the median filter. However ENL seems to be unaffected by this kind of problem.

As for the other outputs NVM seems to be confirmed by MSE values of Table 7.5. Indeed if the first lines of Tables 7.2-7.4 are considered and they are compared with the last three line of Table 7.5, the minimum of NVM corresponds to the minimum of MSE.

However if the second lines of Tables 7.1-7.4 is compared with Table 7.6 and Figure 7.3 it is clear that Median filter, which presents a spike near to 1, remove less correctly the speckle as confirmed by the ENL values. Moreover the higher ENL values of the Gaussian filter with respect to Beltrami flow confirms the property of 'blurring' of the Gaussian filter. In conclusion we can confirm that generally both ENL and NVM reflect properties of PDFs simulations (better than MSE).



(a) Removed speckle PDF, $L = 1$ $\nu = 2$ (b) Removed speckle PDF, $L = 1$ $\nu = 4$



(c) Removed speckle PDF, $L = 1$ $\nu = 8$ (d) Removed speckle PDF, $L = 1$ $\nu = 10$

Figure 7.3: *Speckle distribution of filtered images, from original K-distribution images with parameters equal to* $L = 1$ $\nu = 2, 4, 8, 10$

Metric	Beltrami flow	Gaussian filter	Median filter
NVM	1.2951	1.0931	0.6125
ENL	0.9581	1.262	0.7821

Table 7.2: *Metrics values of RCS reconstruction for filtered images,* $\nu = 4$

Metric	Beltrami flow	Gaussian filter	Median filter
NVM	1.0991	0.9391	0.675
ENL	1.1121	1.4121	0.7545

Table 7.3: *Metrics values of RCS reconstruction for filtered images,* $\nu = 8$

Metric	Beltrami flow	Gaussian filter	Median filter
NVM	1.0471	0.9051	0.6752
ENL	1.1601	1.4481	0.6551

Table 7.4: *Metrics values of RCS reconstruction for filtered images,* $\nu = 10$

ν of simulated p_Γ	Beltrami flow	Gaussian filter	Median filter
$\nu = 2$	$1.2641e-5$	$3.1411e-6$	$6.5360e-6$
$\nu = 4$	$4.1016e-6$	$2.7629e-7$	$4.6143e-6$
$\nu = 8$	$8.3332e-7$	$2.3925e-7$	$3.9653e-6$
$\nu = 10$	$4.1451e-7$	$4.3035e-7$	$3.8160e-6$

Table 7.5: *MSE between Gamma distributed filtered images and theoretical one.*

ν of simulated p_Γ	Beltrami flow	Gaussian filter	Median filter
$\nu = 2$	$1.6342e-6$	$1.1386e-6$	$1.6148e-5$
$\nu = 4$	$1.3052e-6$	$1.8797e-6$	$1.7146e-5$
$\nu = 8$	$1.5633e-6$	$2.5771e-6$	$1.7648e-5$
$\nu = 10$	$1.6923e-6$	$2.7518e-6$	$1.7754e-5$

Table 7.6: *MSE of PDF of removed speckle with theoretical one.*

7.4.2 Filter parameters analysis

Another important issue is to understand which parameters of algorithms affect their performances and how. In this section the Gaussian filter, Beltrami flow and Median filter are considered separately.

Gaussian filter parameters

As for Gaussian, two sets of investigation have been performed in order to understand which parameter between filter variance and window size affects more the algorithm performance. First PDFs for several filter with different window sizes are applied ($3-by-3$, $5-by-5$, $7-by-7$, $9-by-9$, $15-by-15$), but fixing the variance s^2 of (6.12) to 1 were computed and then the same analysis was performed with different window sizes and variances (i.e. s^2 equal to 1, 2, 3, 4, 8).

As depicted in Figure 7.4 and Figure 7.5, fixing the filter variance s^2 to

1 and changing the filter window size, the filtered images PDFs tend to be estimated better with a large filter window size, whereas the removed speckle PDFs are not estimated correctly. However if Figure 7.6 and Figure 7.7 are considered, then the filtered images and removed speckle PDFs are very well estimated for $s = 1, 2$, whereas for higher variances Gaussian filter does not reconstruct RCS correctly. As for removed speckle (as shown in Figure 7.7), increasing the order parameter ν (i.e. $\nu = 8, 10$) of the underlying Gamma-distributed RCS seems to be independent of filter variance $s \geq 2$.
In conclusion one can state that the Gaussian filter behaviour depends more on the filter variance than its window size.

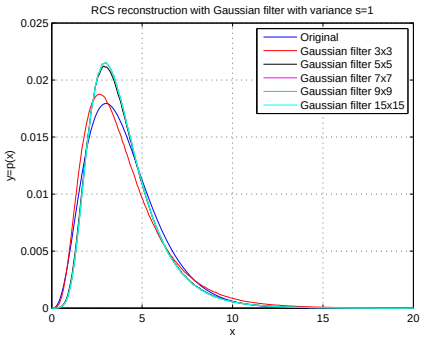
Beltrami flow parameters

As for Beltrami flow, two parameters have been analysed:

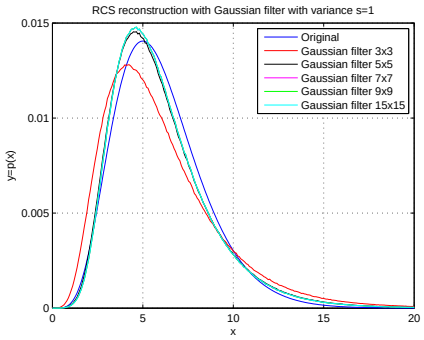
1. The number of iterations, because the kernel has been applied iteratively;

2. Kernel size.

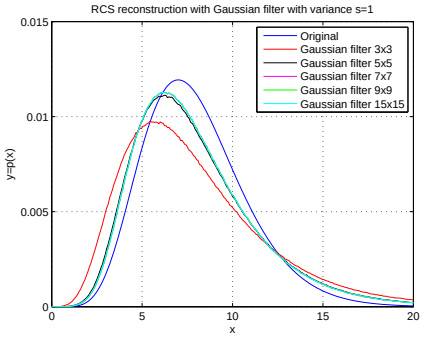
Figures 7.8(a), 7.9(a) and 7.10(a) represent the outcome PDFs of Beltrami flow. Increasing the number of iterations produce more distortions in output images. In practice, doubling the number of iterations (i.e. 2) increases the value and the position (i.e. shifting to the right on x-axis) of PDF mode and as a consequence reduce the tails of filtered RCS images (see MSE values in Table 7.7).
As for the removed speckle (Figure 7.8(b), 7.9(b) and 7.10(b)), increasing the number of iterations however produces better results as demonstrated in



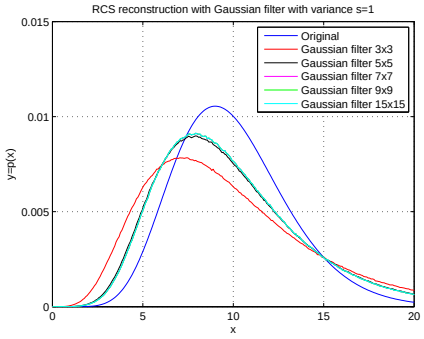
(a) RCS distribution filtered images, $\nu = 4$



(b) RCS distribution filtered images, $\nu = 6$

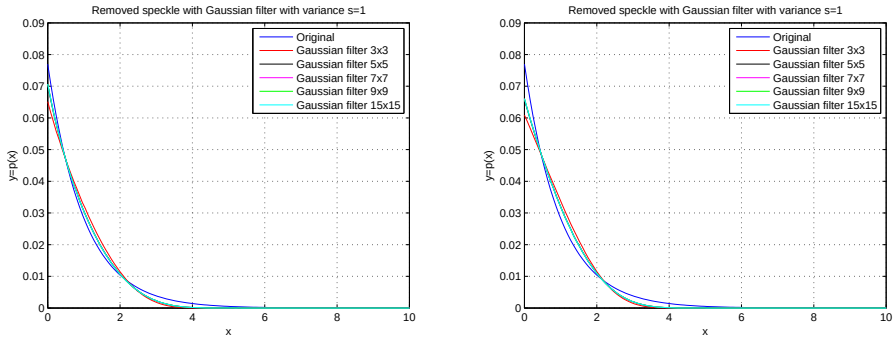


(c) RCS distribution filtered images, $\nu = 8$

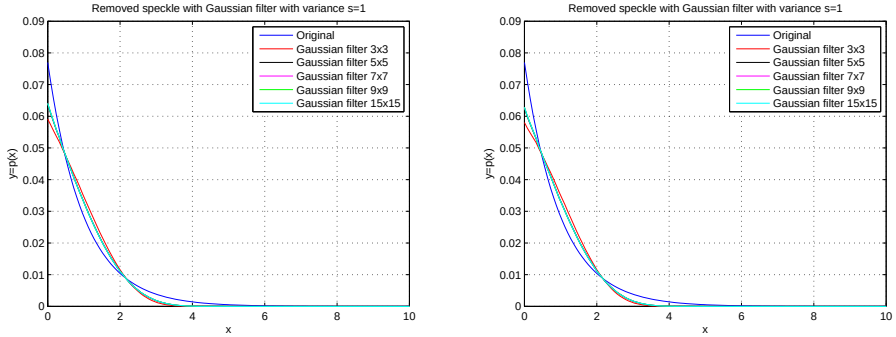


(d) RCS distribution filtered images, $\nu = 10$

Figure 7.4: *Gamma distribution of filtered images estimated by Gaussian filter with fixed variance* $s^2 = 1$ *and different window size* ($3-by-3$, $5-by-5$, $7-by-7$, $9-by-9$, $15-by-15$)*, for* $\nu = 4, 6, 8, 10$

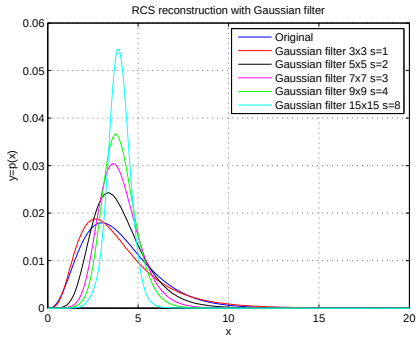


(a) Removed speckle PDF, $L = 1$ $\nu = 4$ (b) Removed speckle PDF, $L = 1$ $\nu = 6$

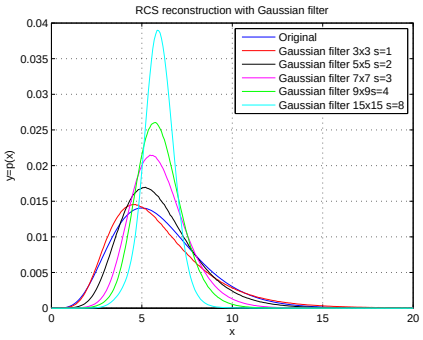


(c) Removed speckle PDF, $L = 1$ $\nu = 8$ (d) Removed speckle PDF, $L = 1$ $\nu = 10$

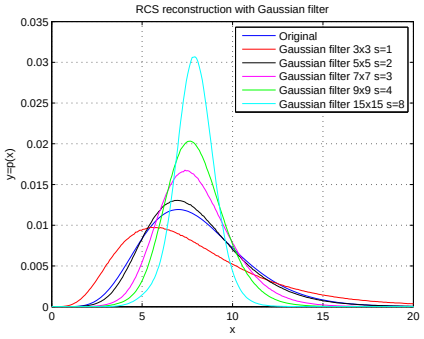
Figure 7.5: *Speckle distribution of filtered images removed by Gaussian filter with fixed variance* $s^2 = 1$ *and different window size* $(3-by-3,\ 5-by-5,\ 7-by-7,\ 9-by-9,\ 15-by-15)$*, from original K-distribution images with parameters equal to* $L = 1\nu = 4, 6, 8, 10$



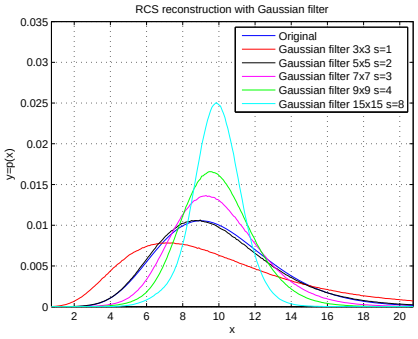
(a) RCS distribution filtered images, $\nu = 4$



(b) RCS distribution filtered images, $\nu = 6$

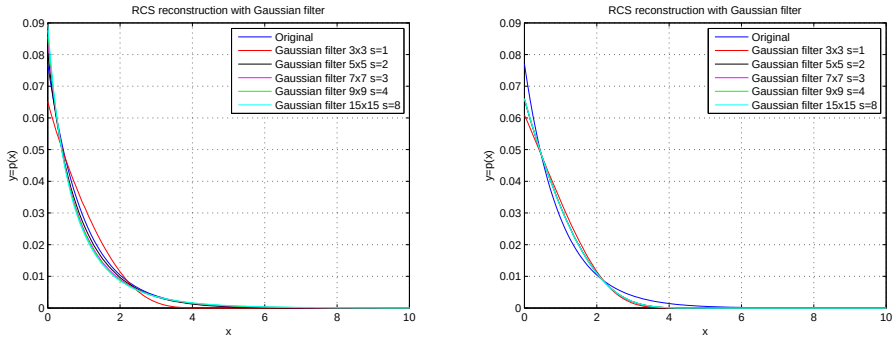


(c) RCS distribution filtered images, $\nu = 8$

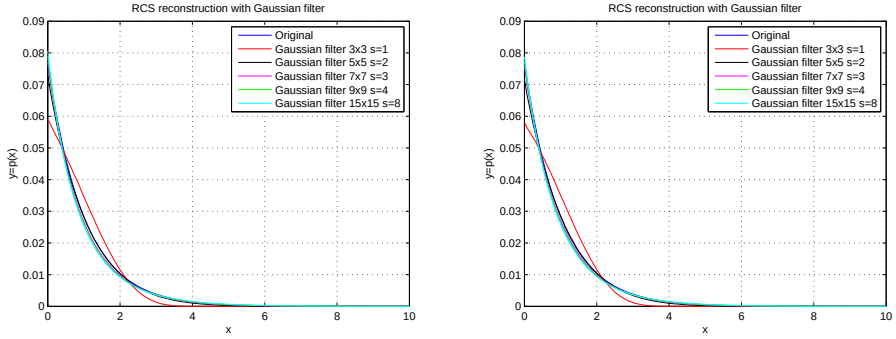


(d) RCS distribution filtered images, $\nu = 10$

Figure 7.6: *Gamma distribution of filtered images estimated by Gaussian filter with different values for variance* s^2 *and different window size* ($3-by-3$, $5-by-5$, $7-by-7$, $9-by-9$, $15-by-15$)*, for* $\nu = 4, 6, 8, 10$



(a) Removed speckle PDF, $L = 1$ $\nu = 4$ (b) Removed speckle PDF, $L = 1$ $\nu = 6$

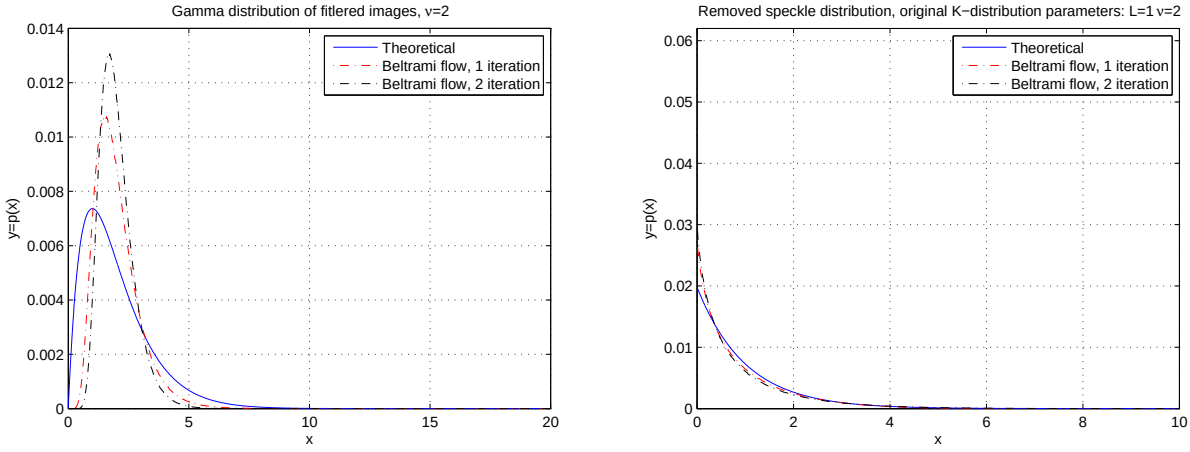


(c) Removed speckle PDF, $L = 1$ $\nu = 8$ (d) Removed speckle PDF, $L = 1$ $\nu = 10$

Figure 7.7: *Speckle distribution of filtered images removed by Gaussian filter with different values for variance* s^2 *and different window size* $(3-by-3,\ 5-by-5,\ 7-by-7,\ 9-by-9,\ 15-by-15)$*, from original K-distribution images with parameters equal to* $L = 1$ $\nu = 4, 6, 8, 10$

Table 7.8.

The second set of experiments consists of processing a set of K-distributed images with Beltrami flow (number of iteration equal to 1) and the kernel size equal to $7-by-7$, $9-by-9$, $11-by-11$ pixels. The outcomes are depicted in Figure 7.11 and 7.12 both RCS PDFs and speckle one are not estimated correctly therefore algorithm performances are affected by both parameters.



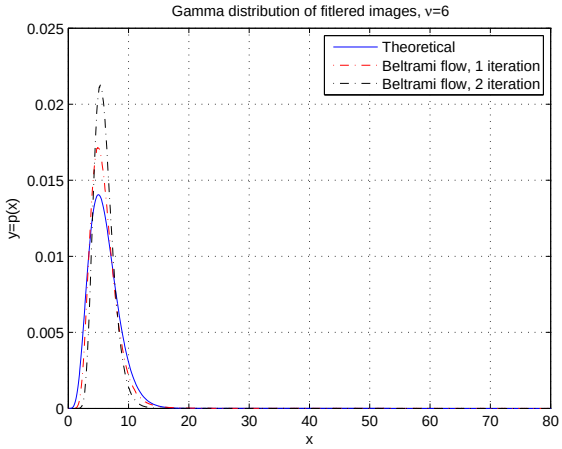
(a) RCS reconstruction PDF with Beltrami flow

(b) Removed speckle PDF by Beltrami flow

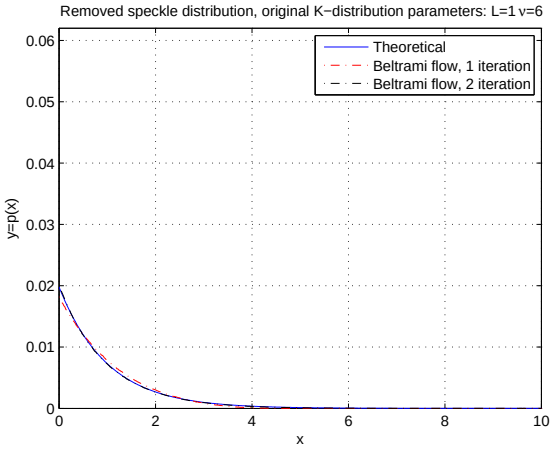
Figure 7.8: *Beltrami flow outputs: comparison single iteration algorithm and 2 iterations algorithm. Original K-distributed images parameters:* $L = 1$ *and* $\nu = 2$.

ν	Beltrami flow 1 iteration	Beltrami flow 2 iterations
$\nu = 2$	$1.5344e-006$	$3.1671e-006$
$\nu = 6$	$3.1725e-007$	$2.0522e-006$
$\nu = 10$	$2.3411e-007$	$4.1451e-007$

Table 7.7: *Beltrami flow: MSE between Gamma distributed filtered images and theoretical one (*ν *is the order parameter of Gamma-distribution).*

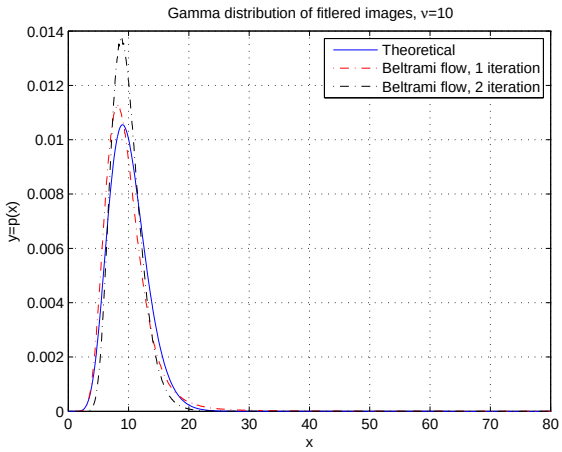


(a) RCS reconstruction PDF with Beltrami flow

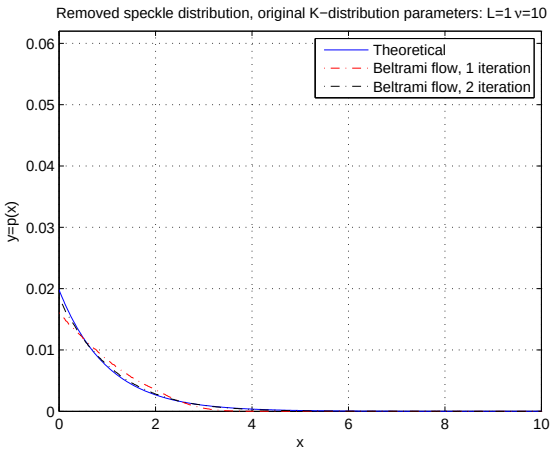


(b) Removed speckle PDF by Beltrami flow

Figure 7.9: *Beltrami flow outputs: comparison single iteration algorithm and* 2 *iterations algorithm. Original K-distributed images parameters:* $L = 1$ *and* $\nu = 6$.



(a) RCS reconstruction PDF with Beltrami flow

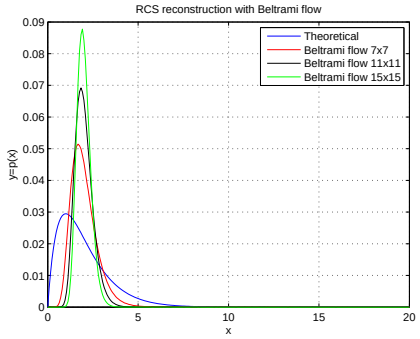


(b) Removed speckle PDF by Beltrami flow

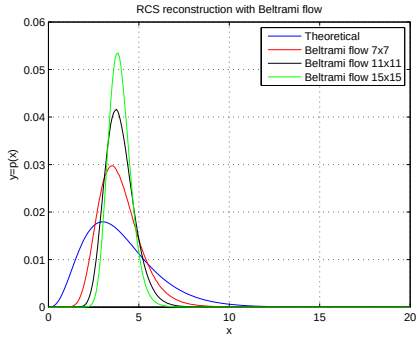
Figure 7.10: *Beltrami flow outputs: comparison single iteration algorithm and* 2 *iterations algorithm. Original K-distributed images parameters:* $L = 1$ *and* $\nu = 10$.

ν	Beltrami flow 1 iteration	Beltrami flow 2 iterations
$\nu = 2$	$2.2603e-007$	$4.3903e-007$
$\nu = 6$	$4.0975e-008$	$3.8088e-008$
$\nu = 10$	$2.0068e-007$	$1.9546e-008$

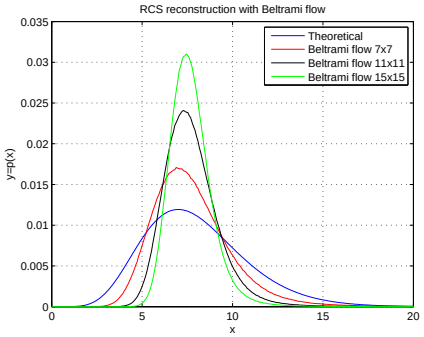
Table 7.8: *Beltrami flow: MSE of PDF of removed speckle with theoretical one (ν is the order parameter of Gamma-distribution).*



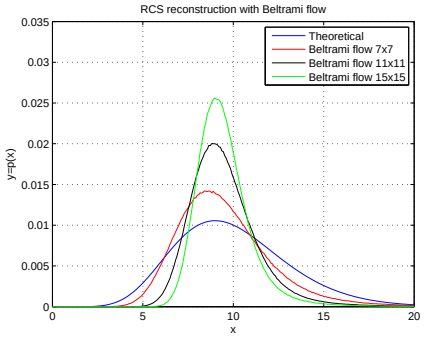
(a) RCS distribution filtered images, $\nu = 2$



(b) RCS distribution filtered images, $\nu = 4$

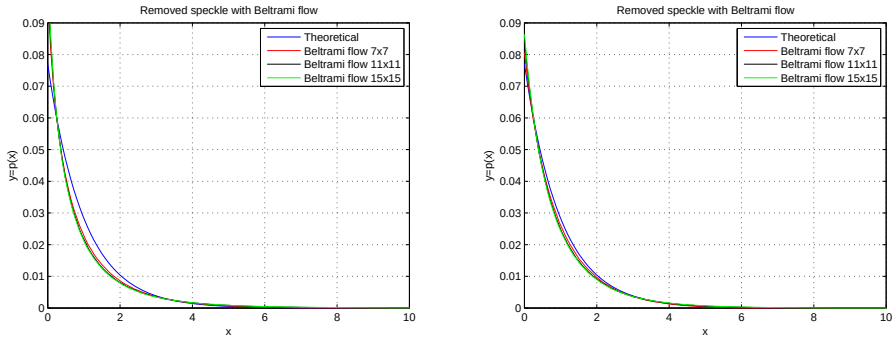


(c) RCS distribution filtered images, $\nu = 8$

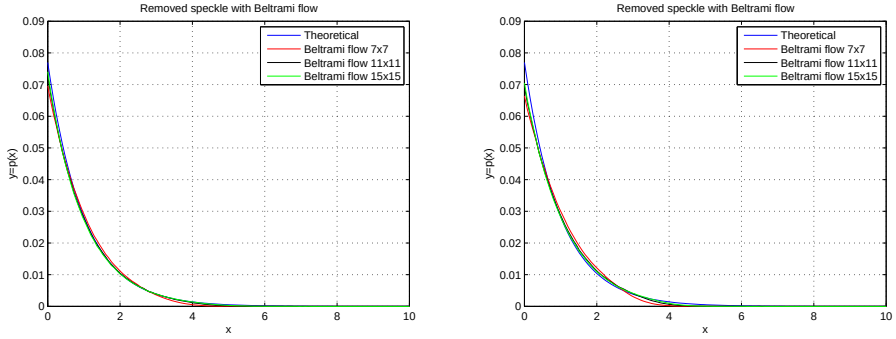


(d) RCS distribution filtered images, $\nu = 10$

Figure 7.11: *Gamma distribution of filtered images estimated by Beltrami flow single iteration and different window size ($7-by-7$, $9-by-9$, $11-by-11$), for $\nu = 2, 4, 8, 10$*



(a) Removed speckle PDF, $L = 1$ $\nu = 2$ (b) Removed speckle PDF, $L = 1$ $\nu = 4$



(c) Removed speckle PDF, $L = 1$ $\nu = 8$ (d) Removed speckle PDF, $L = 1$ $\nu = 10$

Figure 7.12: *Speckle distribution of filtered images removed by Beltrami flow single iteration and different window size* ($7-by-7$, $9-by-9$, $11-by-11$)*, from original K-distribution images with parameters equal to* $L = 1$ $\nu = 2, 4, 8, 10$

Median filter parameters

The only parameter which can affect Median filter performance is its window size, therefore a set of experiments has been performed by varying the window size of the median filter (i.e. $3-by-3$, $5-by-5$, $7-by-7$ pixels). As depicted in Figure 7.13 and 7.14, by increasing the window size the performances of the denoising algorithm/RCS reconstruction filter improve slightly (i.e. RCS PDF estimation is related by window size and order parameter ν, whereas the removed speckle PDFs show a better estimation proportional to the window size - spike amplitude reductions).

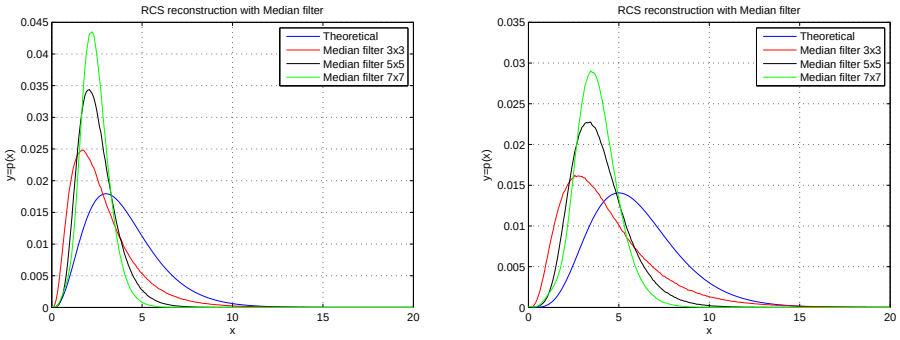
7.5 NVM vs ENL

As shown in the previous sections both ENL and NVM reflect properly RCS filter reconstruction and denoising property accurately and they can be considered dual. In order to manage them correctly in a more efficient way, we have been interested in studying the behaviour of their values as function of order parameter of underlying Gamma-distributed RCS.
In order to manage the collected values, as shown in Figure 7.15, two quantities were introduced, defined as following:

$$a = |L - ENL| \tag{7.1}$$

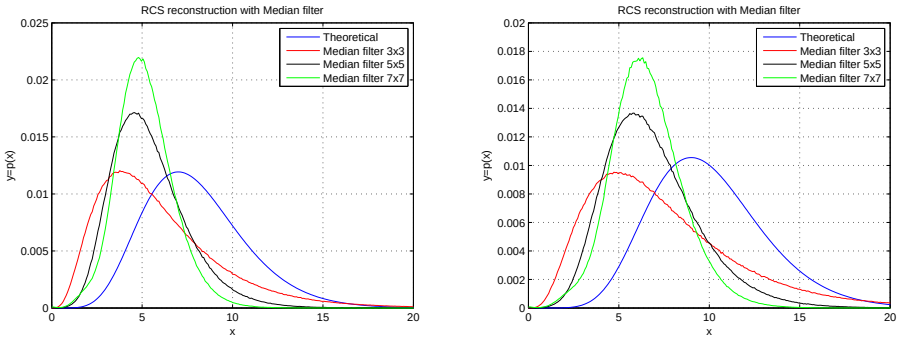
$$b = |L - NVM| \tag{7.2}$$

where L is the actual number of look, as a consequence of the definition of (7.2) and (7.2), the best performances are given by the algorithm which



(a) RCS distribution filtered images, $\nu = 4$

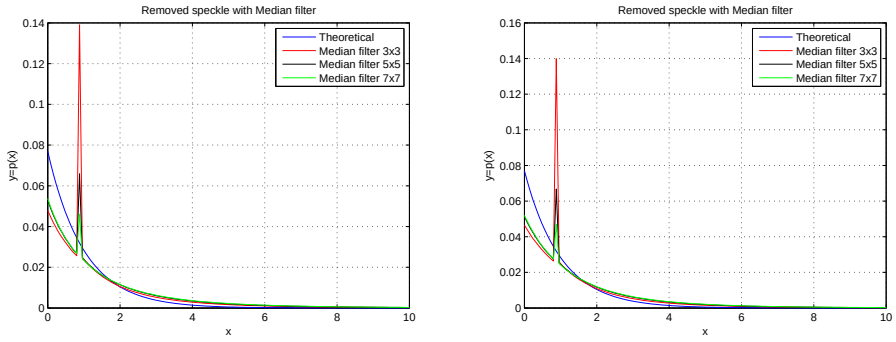
(b) RCS distribution filtered images, $\nu = 6$



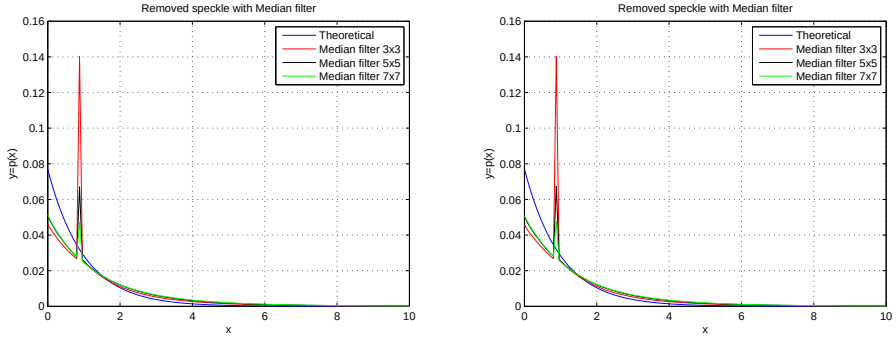
(c) RCS distribution filtered images, $\nu = 8$

(d) RCS distribution filtered images, $\nu = 10$

Figure 7.13: *Gamma distribution of filtered images estimated by Median filter with different window size ($3-by-3$, $5-by-5$, $7-by-7$), for $\nu = 4, 6, 8, 10$*



(a) Removed speckle PDF, $L = 1$ $\nu = 4$ (b) Removed speckle PDF, $L = 1$ $\nu = 6$



(c) Removed speckle PDF, $L = 1$ $\nu = 8$ (d) Removed speckle PDF, $L = 1$ $\nu = 10$

Figure 7.14: *Speckle distribution of filtered images removed by Median filter with different window size ($3-by-3$, $5-by-5$, $7-by-7$), for $\nu = 4, 6, 8, 10$ from original K-distribution images with parameters equal to $L = 1$ $\nu = 4, 6, 8, 10$*

has coordinates (a, b) near to $(0, 0)$. By excluding values for underlying

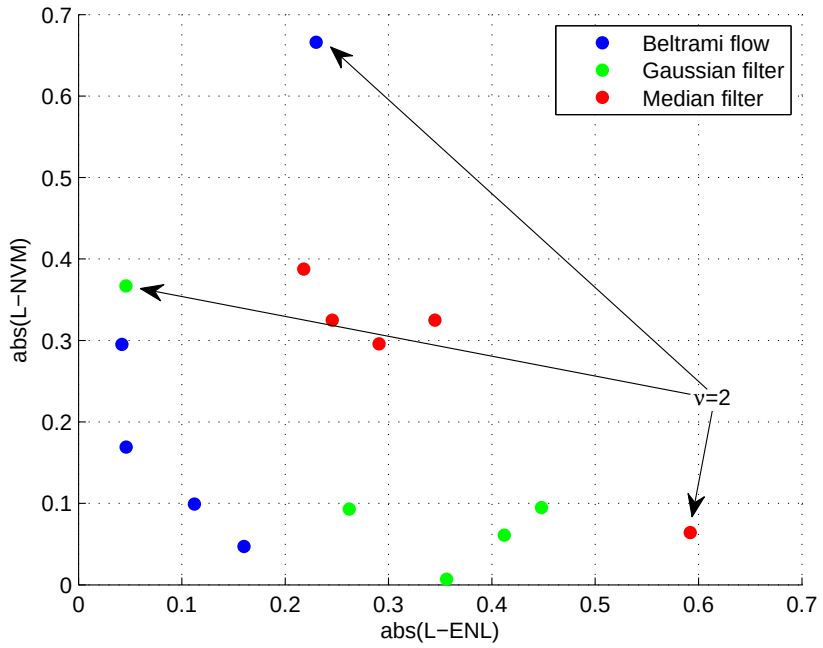


Figure 7.15: *NVM vs ENL: each point of graph has coordinate* (a, b) *defined by (7.2) and (7.2).*

Gamma-distributed RCS order parameter $\nu = 2$ (which as proved in [9] represents a woodland scenario, therefore not important for our investigations, i.e. SAR/ATR systems in GHz bandwidth), one can consider the opportunity to combine both metrics values and compare the algorithms in term of this quantities in order to estimate how correctly the speckle have been removed and how the RCS has been reconstructed (i.e. estimation of the presence of artifacts) well.

7.6 Edge preserving properties

Another important requirment for despeckling algorithms is to preserve the geometrical features of potential targets. Some very important features are indeed related to the area of bright pixels of potential targets, therefore it is important to understand how denoising algorithms are able to preserve edges of images.

For all of the algorithms hence (Beltrami flow, Gaussian filter, and Median filter) applied on a set of images as described in section 7.3, the preserving edges properties have been analysed in terms of the following metrics [115]:

- Change of edge position;
- Height distortion;
- Slope angles;
- Spatial orientation of the edge.

A change in position occurs when there is a translation between edges of original and filtered images. Note that there is no change in edge position if the algorithm produce a symmetric shifting (e.g. Gaussian filter in Figure 7.17). The distortion in edge height introduced by an algorithm is computed by considering the following formula:

$$10 \log E \left\{ \left| (\bar{M}_{or} - \bar{M}_{fltrd}) \right| \right\} \tag{7.3}$$

Where $E\{\cdot\}$ is the mean value, $|\cdot|$ is the absolute value as well as $\bar{M}_{or}$ and $\bar{M}_{fltrd}$ are the mean values of the profile of the edges.

Slope angle however is computed only for the left size of edges by considering the formula:

$$\hat{m} = \arctan \frac{y_2 - y_1}{x_2 - x_1} \tag{7.4}$$

where y_2 is maximum value of the edge y_1 represent the position of the first pixel of the edge. As a consequence, x_2 and x_1 are the abscissa corresponding to y_2 and y_1 respectively.

Finally a spatial orientation occurs if there is an asymmetry in slope angle and/or edge position of the two side of edge.

As for the Gaussian filter, two sets of experiments have been performed: First the metrics have been computed by changing the window size and by fixing the filter variance to 1 and then by changing the variance and as a consequence the filter window size. This is important indeed to understand which parameter has the biggest impact on the images edges.

In Figure 7.16 the results of first experiments are depicted. As for distortions introduced by the filters note that no change in edge position and no spatial orientation occurred. As for height and slope angle, as reported in Table 7.9, the slope angle tends to be independent of changes in filter window size, where are height edge distortion are less than $30dB$ intensity units.

As for the outcomes of Gaussian filter with different variance values, the

Window size	3×3	5×5	7×7	9×9	15×15
MAE (dB)	25.9273	28.4235	28.7022	28.7140	28.7142
S.A. (Deg)	89.3069	88.8371	1.5424	88.3730	88.3730

Table 7.9: *Edge distortion and slope angle introduced by Gaussian filter, fixed variance $s^2 = 1$ and different window size (MAE is the Mean of Absolute Error, whereas S.A. is the Slope Angle).*

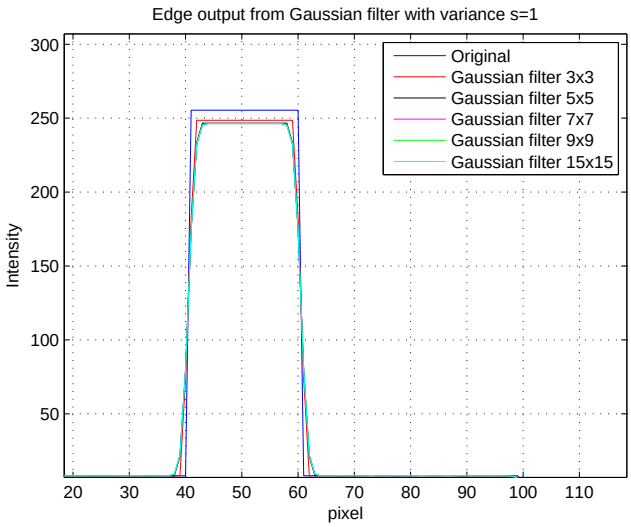


Figure 7.16: *Edge distortion introduced by Gaussian filtering. By increasing the value of the filter window size,* $s^2 = 1$.

results are depicted in Figure 7.17 and in Table 7.10. Also in this case no spatial orientation and edge shift occur, nevertheless the smoothing is bigger than in the previous case, therefore the distortion of edge height and slope angle are bigger than the previous case. As a consequence one can state that the dependencies of the Gaussian filter output depends more on filter variance than window size. As for Beltrami flow, two scenarios were analysed: First,

Window size	3×3 $s^2 = 1$	5×5 $s^2 = 2$	7×7 $s^2 = 3$	9×9 $s^2 = 4$	15×15 $s^2 = 8$
MAE (dB)	25.9273	32.3931	35.9446	38.4950	43.6093
S.A. (Deg)	89.3069	88.4532	88.2526	87.9604	85.9551

Table 7.10: *Edge distortion introduced by Gaussian filter (MAE is the Mean of Absolute Error, whereas S.A. is the Slope Angle).*

checking the edge preserving properties in the case of the iterative algorithm and then how the size of the window filter can affect the edge properties. The results are depicted in Figure 7.18 and Figure 7.19. In both case one

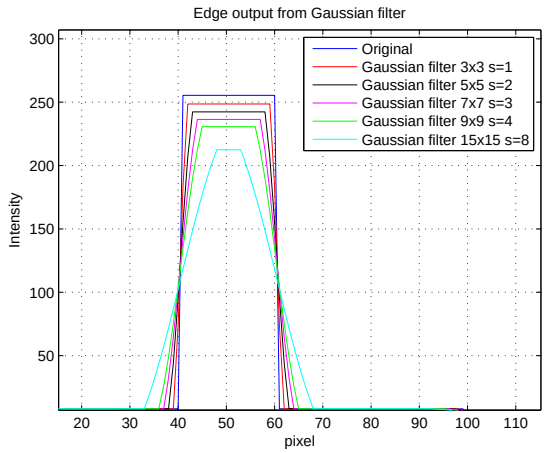


Figure 7.17: *Edge distortion introduced by Gaussian filtering. By increasing the value of the variance the distortions of edges in terms of their height and slope angle increase.*

can state that the number of iterations and filter window size do not affect in any case the edge preserving properties (i.e. no distortions, position change, slope angle and orientation changing), in case $SCR \geq 5dB$.

Figure 7.18 and Figure 7.19 prove that Beltrami flow does not introduce any distortion, any change of edge position and there is no spatial orientation on the edge.

As for the median filter, the results are depicted in Figure 7.20. The window size of the filter affect the results less than Gaussian filter, but much more than Beltrami flow. As shown in Table 7.11, the Median filter introduced a smaller distortion than the Gaussian filter, but slope angle values are very similar to the Gaussian filter results.

Window size	3×3	5×5	7×7
MAE (dB)	3.7503	8.4548	10.6811
S.A. (Deg)	89.3126	88.8542	88.3959

Table 7.11: *Edge distortion introduced by Median filter.*

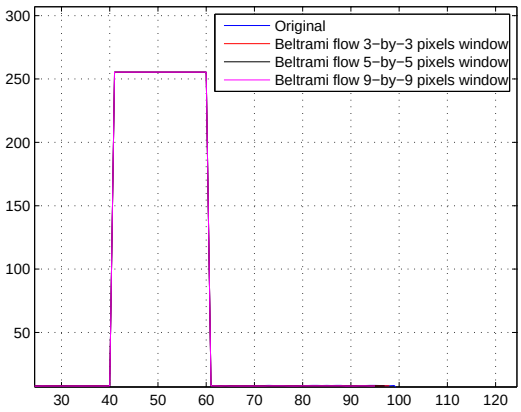


Figure 7.18: *Beltrami flow: edge preserving properties in case of increasing filter window sizes, 2 iterations.*

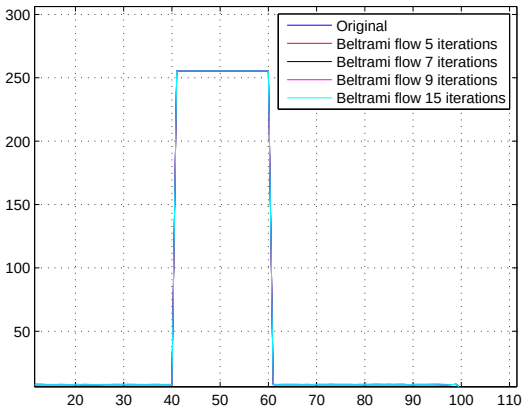


Figure 7.19: *Beltrami flow: edge preserving properties in case of increasing number of iterations, 3-by-3 pixels window filter.*

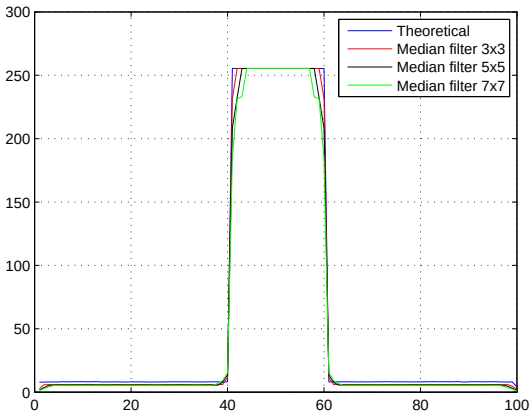


Figure 7.20: *Median filter edge processing:Edge smoothing is slightly better than Gaussian filter in terms of distortion and slope angle*

7.7 Summary

In Chapter 7 the Gaussian filter, the median filter and Beltrami flow have been analysed as despeckling algorithms. The analysis consisted in: firstly assessing the distortions introduced by the algorithms as RCS reconstruction filter; secondly comparing the filtered results with the ENL and NVM predicted values. Finally the most interesting edge preserving metrics have been described and as a consequence the edge preserving properties of the analysed algorithms have been reported.

It is worth pointing out that Beltrami flow preserves edges better than Gaussian and Median filter, nevertheless Barbaresco's work [148] argues that, from a geometrical point of view, Beltrami flow tends to minimize the mean curvature [130] of the edges, instead of protecting the principal curvature direction properly ([131], [132] and [130]). Experiments confirm Barbaresco's conclusions.

Chapter 8

Detection analysis

8.1 Introduction

A SAR/ATR system usually consists of three main actions: detection, discrimination and classification [2], [9] and [114]. First, the entire SAR image is scanned for the target detection stage which requires at least knowledge of the background clutter model. It yields a large number of false alarms in addition to identifying potential targets, therefore it is very important to perform a very effective and efficient detection process.

The outputs of the detection step are then passed to a discrimination stage, which should be able to reject further false targets based on simple properties of potential targets, including both geometrical and electromagnetic effects. Once the detection and discrimination stages have rejected as much clutter as possible, the final stage of an ATR scheme consists of target classification using all the information in the data.

The presence of speckle noise in SAR images affects the discrimination of

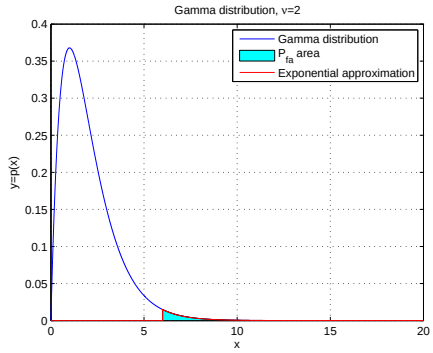


Figure 8.1: *Exponential approximation of a Gamma-distribution (order parameter $\nu = 2$). The idea is to approximate the P_{fa} area with a function which is easier to manage in the integral in (8.1).*

potential target features, therefore denoising algorithms are usually applied. Moreover other operations, such as incoherent averaging, are performed in order to improve the detection performance. All of the described operations change the background clutter model (i.e. clutter probability density function, **PDF**), which is crucial for estimating detection process parameters.

Detection can be described in terms of **probability of false alarm** (P_{fa}, also known as **error type I**) which represents the probability that the clutter is considered erroneously a potential target by the detection subsystem, defined as:

$$P_{fa} = \int_t^{\infty} p(x|B)dx \tag{8.1}$$

where $p(x|B)$ represents the probability that the pixel x is clutter given a clutter model B. The performance of the detector is also described in terms of **probability of detection** (P_d) which is defined as:

$$P_d = \int_t^{\infty} p(x|T)dx \tag{8.2}$$

where $p(x|T)$ is the likelihood function, i.e. the probability of a data value x when the target is present.

Unfortunately these two quantities are conflicting, therefore an optimization criterion has to be adopted to maximize the P_d with the constraint $P_{fa} \leq \alpha$ (with $0 \leq \alpha \leq 1$).

In radar systems the Neyman-Pearson test is usually considered as the best criterion to overcome the optimization problem and to determine which hypothesis is true (i.e. pixel x is a target or clutter respectively). It states that the target is detected if:

$$\frac{p(x|T)}{p(x|B)} > \tau_{NP}(SNR) \tag{8.3}$$

where SNR is the Signal-to-Noise Ratio.

The threshold τ_{NP} is usually selected to give a previously fixed value of $P_{fa} \leq \alpha$ [2] and it can be estimated from the available samples, therefore the knowledge of the clutter model is crucial. Unfortunately in most cases a closed form for the filtered clutter model is not available thereby suboptimal solutions are adopted (e.g. Exponential or Gamma-distributed clutter model [2] [9]).

Fortunately for Detection problems a global clutter model is not necessary, but an approximation of the filtered clutter tails is sufficient because the P_{fa} represents numerically the underlying area of the clutter model tail.

In this Chapter a novel mathematical approach is introduced to approximate the data output from the denoising process. The idea, as depicted in Figure 8.1, can be summarized as follows: the filled area underlying the global

clutter model (i.e. Gamma density function with order parameter $\nu = 2$ and scale parameter 1) has to be equal to the underlying area of the approximating function and the initial approximation point has to be equal for both the models. Note that the approximating function should be a function which allows us to compute the Neyman-Pearson threshold through (8.1) easily.
As for our experiments a two-parameters Constant False alarm Rate (*CFAR*, which can be considered as a quantization process due to the thresholding procedure, i.e. it is an asymptotic non-information preserving transformation as reported in Figure 1.10) has been considered [9] and it has been applied to image intensity. The target detection occurs when:

$$\frac{\bar{I}_T/\bar{I}_B - 1}{\sqrt{V_B}} > \tau_{NP} \tag{8.4}$$

where $\bar{I}_T$ is the average target intensity, estimated over the ROI of m pixels, whereas $\bar{I}_B$ and $\sqrt{V_B}$ are the average intensity and normalized variance of background respectively, estimated over the ROI of M pixels (as depicted in Figure 8.2).

Several techniques have been developed in order to increase the performances of CFAR subsystems and one of the most important is termed *incoherent averaging* [9] (under the assumption that the target spreads over as many pixels in the ROI, i.e. targets detected in high resolution radar systems) which consists of averaging the targets pixels, in order to reduce the speckle effects. As a consequence the single pixel target and multiples pixels target scenarios have been investigated in order to understand which parameters are important in CFAR subsystems and how despeckling algorithms

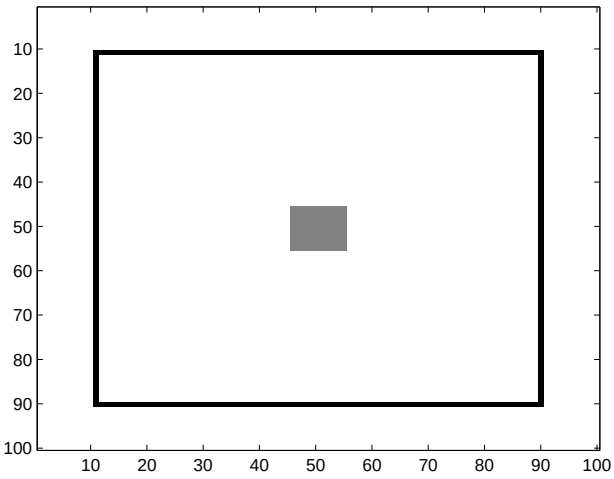


Figure 8.2: *ROI example, the target is placed at the centre of the image, whereas the contour represent the pixels over which I_B and V_B are estimated.*

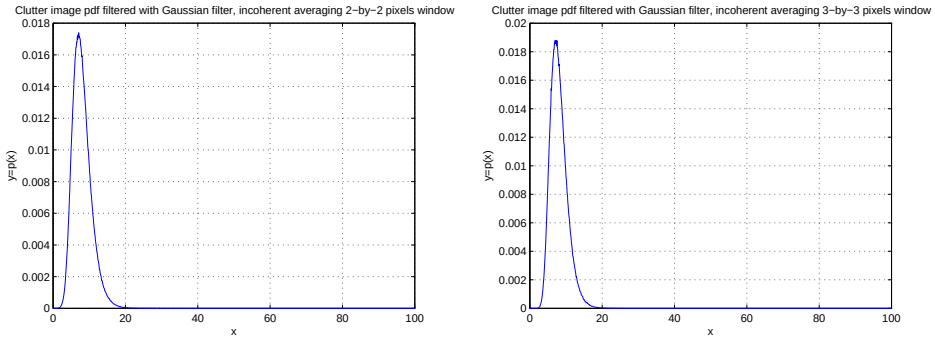
affect them.

8.2 Incoherent averaging

If a target subtends many pixels, CFAR conditions can be estimated over more than a single pixel, as a consequence the performances of CFAR subsystems improve because the speckle effects are reduced.

Unfortunately no investigation have been produced in order to understand how the clutter distribution changes when it is averaged over more than a single pixel of the denoised SAR image (in Figure 8.3 it is reported a simulated set of K-distributed SAR images was filtered with a Gaussian filter and then the outcomes were averaged by considering $2 - by - 2$ and $3 - by - 3$ adjacent pixels sub-matrices respectively).

As depicted in Figure 8.1, incoherent averaging tends to make the distri-



(a) Gaussian filter output distribution after incoherent averaging, $2-by-2$ pixels window

(b) Gaussian filter output distribution after incoherent averaging, $3-by-3$ pixels window

Figure 8.3: *PDFs of output of Gaussian filter (applied to a K-distributed images set with* $\nu = 8$ *and* $L = 1$*). The filtered image is averaged by considering a subset of* $2-by-2$ *and* $3-by-3$ *adjacent pixels submatrices respectively.*

butions of clutter narrower and at the same time it shifts the mode of the distribution to the right since random variables are correlated. As a consequence the output of a denoising filter is no more Gamma-distributed and therefore a closed form for the clutter model in the Neyman-Paerson criterion (8.3) is impossible .

From a mathematical point of view, P_{fa}, as defined in (8.1), represents

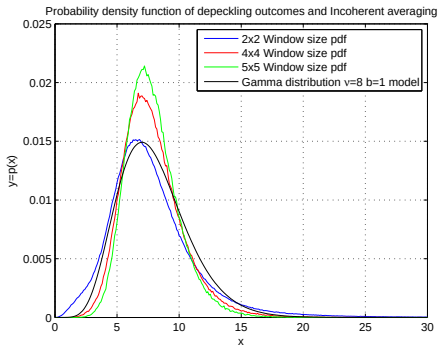


Figure 8.4: *Probability density functions of outcomes from despeckling and incoherent averaging filters. The simulated data are compared with a global clutter model: Gamma distribution* $\nu = 8$ *and scale parameter* $b = 1$.

the area under the clutter-RCS tail, as depicted in Figure 8.1. From (6.9), which represents the function of a Gamma-distribution, it is clear that for high values of variable x, the exponential slope becomes prevailing, therefore the tails can be approximated by an exponential function, which is easier to manage in order to determine the value of the CFAR threshold.
How to determine such an exponential function? An easy approach is given by considering a two parameter exponential function, defined as:

$$f(x; c_1, c_2) = c_1 \exp(-c_2(x - x_0)) \tag{8.5}$$

where c_1 and c_2 are two parameters which have to be estimated by fixing a value for variable x (set $x_0 = 2\nu + 2$, where ν is the order parameter of Gamma-distribution which is to estimate) and by considering the following constrains:

- c_1 is chosen by satisfying the condition: $f(x = x_0; c_1, c_2) = p_x(x = x_0)$ in (6.9).
- c_2 is chosen by satisfying the condition: $\int_{x_0}^{\infty} f(x; c_1, c_2)dx = P(x >= x_0)$ which states that the two underlying areas have to be equal.

As reported in Table 8.1, the proposed method is able to estimate the exponential approximation and assure that the computed P_{fa} is smaller than the expected one.
In conclusion, one is not interested in finding a function which is able to fit accurately the whole PDF under analysis, but just its tail.

ν	$T_{s,10^{-6}}$	$T_{s,10^{-9}}$	$\hat{P}_{fa}\|_{T_{s,10^{-6}}}$	$\hat{P}_{fa}\|_{T_{s,10^{-9}}}$	$\text{MSE}_{Appr.}$
$\nu = 2$	17.3889	25.4483	$5.1601\ 10^{-7}$	$2.35\ 10^{-10}$	$4.7078\ 10^{-11}$
$\nu = 4$	22.6274	32.0640	$3.29\ 10^{-7}$	$7.1764\ 10^{-11}$	$3.8744\ 10^{-11}$
$\nu = 6$	26.7941	37.0489	$3.23\ 10^{-7}$	$5.4394\ 10^{-11}$	$1.1873\ 10^{-11}$
$\nu = 8$	30.4669	41.2720	$3.6404\ 10^{-7}$	$5.7712\ 10^{-11}$	$3.0319\ 10^{-12}$
$\nu = 10$	33.8663	45.0696	$4.2527\ 10^{-7}$	$7.0087\ 10^{-11}$	$7.3809\ 10^{-13}$

Table 8.1: *Exponential approximation values. Note that the thresholds (columns* 2 *and* 3*) are computed by using the approximating function, whereas the values of columns number* 4 *and* 5 *represent the values of* P_{fa} *of approximated CDF corresponding to the thresholds reported in columns* 2 *and* 3 *(*ν *is the order parameter).*

8.3 Considerations on clutter models

In this Chapter the considered clutter models have been K and *Weibull* distributions, but unfortunately their approximations with an exponential distribution have failed. As for Weibull distribution indeed the cumulative density function (**CDF**), because it is easier to handle than the corresponding PDF) is defined as:

$$P_w(x) = 1 - e^{-(\frac{x}{\lambda})^k} \tag{8.6}$$

where $\lambda > 0$ is a real value, termed scale parameter, whereas $k > 0$ is the shape parameter (real). If the Mclaurin series expansion [109] for the exponential term of (8.6) and the exponential term of CDF of an exponential distribution (i.e. $P_{exp}(x) = 1 - e^{\lambda x}$) are computed, it is easy to show that the difference of two expansions is:

$$e^{d_0 x} - e^{d_1 x^k} = \sum_{n=0}^{\infty} \left[\frac{(d_0)^n}{n!} - \frac{(d_1)^n}{n!} x^{(k-1)n} \right] x^n \tag{8.7}$$

the term between brackets in (8.7) is always non-null (for non-trivial cases $k \neq 1$) therefore an exponential function is not able to approximate the Weibull distribution. Similar considerations are also valid for the K-distribution. A modified Bessel function of second kind $K_s(x)$ indeed has asymptotes at [109]:

$$K_s(x) \approx \sqrt{\frac{\pi}{2x}} e^{-x}, \quad x \to \infty \tag{8.8}$$

therefore an exponential approximation is not possible.

Another interesting interesting property of the clutter models considered is reported in the following theorem:

Theorem 1 (Exponential upper bound) *Let x be an Exponential distributed random variable, if its mean value $\langle x \rangle \geq 1$ then the Exponential distribution is an asymptotic upper bound for Weibull (for $k > 1$) and K clutter models, i.e. it is valid the following inequality:*

$$\textit{Exponential model} \geq \textit{K-model} \geq \textit{Weibull's model} \tag{8.9}$$

Proof. Let us consider the equation (8.8) and the following ones:

$$p_1(x) = k \frac{x^{k-1}}{\lambda_1^k} e^{-\frac{x^k}{\lambda_1^k}}, \; k > 1 \text{ and } \lambda_1 > 0 \tag{8.10}$$

$$p_2(x) = \lambda_2 e^{-\lambda_2 x}, \; \lambda_2 > 0 \tag{8.11}$$

To prove the theorem, compute the following limits:

$$\lim_{x\to\infty}\frac{p_1(x)}{p_2(x)} = \lim_{x\to\infty}\frac{k\frac{x^{k-1}}{\lambda_1^k}e^{-\frac{x^k}{\lambda_1^k}}}{\lambda_2 e^{-\lambda_2 x}} \tag{8.12}$$

$$\lim_{x\to\infty}\frac{\sqrt{\frac{\pi}{2x}}e^{-x}}{p_2(x)} = \lim_{x\to\infty}\frac{\sqrt{\frac{\pi}{2x}}e^{-x}}{\lambda_2 e^{-\lambda_2 x}} \tag{8.13}$$

(8.12) is always equal to zero, whereas the limit in equation (8.13) is equal to zero if and only if $\lambda_2 \leq 1$, i.e. the expected value of the Exponential distribution (8.11) is ≥ 1. As for the inequality (8.9), it is sufficient to compute the following limit:

$$\lim_{x\to\infty}\frac{k\frac{x^{k-1}}{\lambda_1^k}e^{-\frac{x^k}{\lambda_1^k}}}{\sqrt{\frac{\pi}{2x}}e^{-x}} = 0 \text{ always for } k > 1 \tag{8.14}$$

■

Corollary 2 *If a Weibull distribution has shape parameter $k < 1$ and assumption of Theorem 1 are preserved, then inequality (8.9) becomes:*

$$\textit{Weibull's model} \geq \textit{Exponential model} \geq \textit{K-model} \tag{8.15}$$

Proof. The proof is a consequence of (8.14) and (8.13). ■

8.4 Method description

Unfortunately SAR signal processing tends to change the statistics of the background clutter model (as depicted in Figure 8.4) and in most cases the

Windows size	λ	k	MSE	T_3	T_4	ϵ_3	ϵ_4
2×2	4.2748	0.9370	10^{-10}	33.6260	45.7100	10^{-4}	10^{-7}
4×4	5.3647	1.4642	10^{-11}	20.0806	24.4402	10^{-4}	10^{-7}
5×5	5.2757	1.6121	10^{-11}	17.4959	20.9140	10^{-4}	10^{-5}

Table 8.2: *Weibull parameters approximation:* ϵ_j *(*$j = 3, 4$*) ensures that at the threshold* T_j *the approximating CDF acts as an lower bound for the approximated CDF*

outcomes are not computationally feasible in a closed mathematical form. Skolnik in [2] introduces the classical Swerling model II, whereas Oliver in [9] suggests to use K-distribution clutter model and a Gamma approximation for large number of looks L (i.e. the number of radar antenna sub-apertures); Roy in [134] uses a K-distributed form of non-Gaussian clutter. Levanon in [135], Anatassopoulos in [136] however use a global Weibull background clutter model.

A local approach of approximating filtered clutter can be more efficient than the classical approach of approximating the clutter distribution (i.e. assuming that outputs are Exponential or Gamma-distributed for instance). A local approach can be made even easier if CDFs are considered. CDFs indeed can be mathematically more manageable than PDFs . Hence, our problem can be summarized as follows: *'Finding a function which approximates the filtered outputs CDF so that the approximating CDF value corresponding to the threshold is a lower bound for the value of the approximated CDF '* [173]. Let (x_0, y_0) and (x_1, y_1) be two points of the CDF which is to be approximated and consider, for example, the Weibull CDF:

$$P_{fa} = 1 - \Pr(\mathtt{x} \leq x) = 1 - (1 - e^{-(\frac{x}{\lambda})^k}) = e^{-(\frac{x}{\lambda})^k} \tag{8.16}$$

where $\Pr(\mathtt{x} \leq x)$ is the CDF, k is the scale parameter and λ is the shape parameter of the Weibull clutter model.

The approximating CDF can be computed by solving:

$$\begin{cases} \left(\frac{x_0}{\lambda}\right)^k = -\ln(1-y_0) \\ \left(\frac{x_1}{\lambda}\right)^k = -\ln(1-y_1) \end{cases} \tag{8.17}$$

which determines the values of the Weibull parameters $k > 0$ (scale) and $\lambda > 0$ (shape).

As for the choice of the parameters of (8.17), they will be discussed in the next sections.

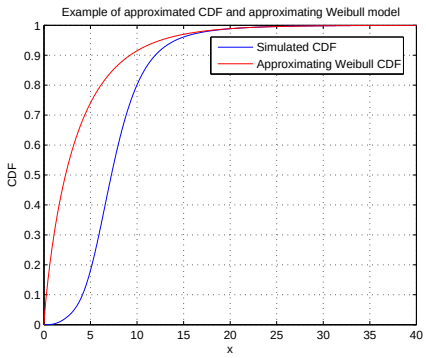


Figure 8.5: *Example of global approximation (MSE= 0.0057): The Weibull model (obtained by a filtered images set which has been incoherently averaged with non-overlapped 2 by 2 pixels window) with parameters* $\lambda = 3.5370$ $k = 0.8677$ *(initial point* $x_0 = 32$*) respectively and the processed images CDF.*

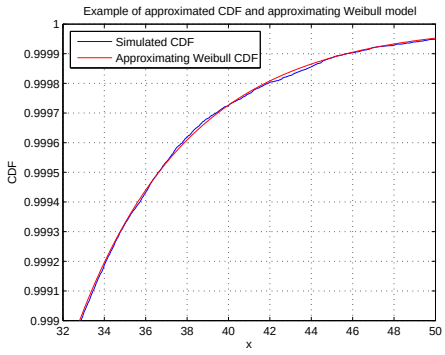


Figure 8.6: *Example of local approximation (MSE*$\approx 10^{-12}$*): The Weibull model (obtained by a filtered images set which has been incoherently averaged with non-overlapped* 2 *by* 2 *pixels window) with parameters* $\lambda = 3.5370$ $k = 0.8677$ *(initial point* $x_0 = 32$*) respectively and the processed images CDF.*

8.5 Results

A set of 1000, 100 *by* 100 pixels, SAR images have been simulated with a clutter model defined by a K-distribution as follows [9]:

$$P(I) = \frac{2}{\Gamma(L)\Gamma(\nu)} \left(\frac{L\nu}{\langle I \rangle}\right)^{(L+\nu)/2} \times I^{(L+\nu-2)/2} K_{\nu-L} \left[2\left(\frac{\nu L I}{\langle I \rangle}\right)^{1/2}\right] \tag{8.18}$$

where $L = 1$ is the number of images averaged (number of looks) $\nu = 8$ is the order parameter, $\langle I \rangle = 8$ image intensity mean value, $\Gamma(\cdot)$ is the Gamma Function, K is the modified Bessel Function of second kind. As for the despeckling algorithm, Beltrami flow [124] (single iteration and window size 5 *by* 5) has been adopted. The despeckled images have been then averaged over non-overlapped 2 *by* 2, 4 *by* 4 and 5 *by* 5 pixels windows. Finally the CDFs have been computed.

The computed CDF have been approximated by using (8.17) with following

Windows size	T_3	T_3 (est.)
2×2	32.8833	33.6260
4×4	19.6820	20.0806
5×5	17.1617	17.4959

Table 8.3: *Comparison thresholds: The actual threshold is compared with the estimated one for $P_{fa} = 10^{-3}$ by using the local model (8.17)*

parameters: $x_0 = 15$ ($y_0 = CDF(x_0)$), whereas x_1 is the first value of the approximated CDF such that $|CDF(x_1) - 1| \leq 10^{-4}$ ($y_1 = CDF(x_1)$).

An example is depicted in Figure 8.5 and Figure 8.6, which represent the same solution seen globally and locally respectively.

The simulations are performed in order to compute the threshold for P_{fa} equal to 10^{-3} and 10^{-4}.

The parameters of approximating CDF are reported in Table 8.2: λ and

Windows size	T_4	T_4 (est.)
2×2	45.7246	45.7100
4×4	24.4524	24.4402
5×5	20.8721	20.9140

Table 8.4: *Comparison thresholds: The actual threshold is compared with the estimated one for $P_{fa} = 10^{-4}$ by using the local model (8.17)*

k are the parameters estimated through (8.17), whereas the Mean Squared Error (**MSE**) between the approximating CDF and the approximated one is computed from the initial point of approximation. T_3 and T_4 however are the thresholds computed (for P_{fa} equal to 10^{-3} and 10^{-4} respectively) by using equation (8.16):

$$T_s = \lambda \left[-\ln P_{fa} \right]^{\frac{1}{k}} \tag{8.19}$$

As reported in Table 8.3 and Table 8.4 the thresholds are better estimated for small values of the P_{fa}.

The values of thresholds have been also tested by considering the frequency (i.e. the percentage of pixels) of filtered and averaged images pixels, as previously described, which exceed them (No. of samples: $2.5 \cdot 10^6$ (2×2), $6.25 \cdot 10^5$ (4×4) and $4 \cdot 10^5$ (5×5) respectively). As reported in Table 8.5, the value of the thresholds produces values of error type I smaller than the original P_{fa} (P_{T_3} and P_{T_4} represent the probability of clutter pixels which exceed the thresholds T_3 and T_4 respectively).

The local approximated clutter model has been compared by two global

Windows size	P_{T_3}	P_{T_4}
2×2	$\approx 10^{-4}$	$\approx 10^{-5}$
4×4	$\approx 10^{-4}$	$\approx 10^{-5}$
5×5	$\approx 10^{-4}$	$\approx 10^{-5}$

Table 8.5: *Estimated P_{fa} by using a local approximated clutter model approach*

clutter models: Exponential and Gamma clutter model ($\nu = 8$ and scale parameter 1) respectively.

As for the Exponential clutter model [9], the threshold is computed by considering:

$$T_s = -\sigma_c \ln P_{fa} \tag{8.20}$$

where σ_c is the mean power of the clutter. The estimated thresholds are reported in Table 8.6.

As for the Gamma-distribution clutter model [137], the thresholds are com-

Windows size	T_3	T_4
2×2	64.9240	86.5654
4×4	34.7892	46.3856
5×5	26.5142	35.3522

Table 8.6: *Estimated thresholds by using a global Exponential clutter model*

Windows size	P_{T_3}	P_{T_4}
2×2	$\approx 10^{-6}$	$< 10^{-6}$
4×4	$< 10^{-6}$	$< 10^{-6}$
5×5	$< 10^{-6}$	$< 10^{-6}$

Table 8.7: *Estimated P_{fa} by using a global Exponential clutter model*

puted by inverting numerically the following formula:

$$P_{fa} = \sum_{i=0}^{\nu-1} \frac{(\frac{x}{\theta})^i}{i!} e^{(-\frac{x}{\theta})} \tag{8.21}$$

where $\nu = 8$ and $\theta = 1$ are the order parameter and scale parameter of the Gamma-distribution respectively. As a consequence the threshold assume values: $T_3 = 19.2104$, $T_4 = 20.1830$ respectively.

By comparing Table 8.5, Table 8.7 and Table 8.8 it is clear that the proposed method is more efficient than other clutter models.

Windows size	P_{T_3}	P_{T_4}
2×2	$\approx 10^{-2}$	$\approx 10^{-3}$
4×4	$\approx 10^{-4}$	$\approx 10^{-4}$
5×5	$\approx 10^{-4}$	$\approx 10^{-5}$

Table 8.8: *Estimated P_{fa} by using a global Gamma ($\nu = 8$ and scale parameter 1) clutter model*

8.6 Analysis parameters

The parameters which have been chosen accurately are the points (x_0, y_0) and (x_1, y_1). The method has been developed by considering x_1 the first value of the approximated CDF such that $|CDF(x_1) - 1| \leq 10^{-4}$ (i.e. $y_1 = CDF(x_1)$). The value of thresholds is insensitive to the value of the initial point which has been fixed to $x_0 = 2\mu$, i.e. μ is the mean value of the data. Moreover we suggest to introduce two margins $0 < \eta_j < 10^{-8}$, $j = 0, 1$ (subtracted to the actual values y_j) in order to obtain positive errors $\epsilon_i = CDF_{i,ted} - CDF_{i,ing}$ (see last two columns of Table 8.2, for $i = 3, 4$). Under this assumption a solution is always found and the estimated thresholds shows that the corresponding estimated P_{fa} is always smaller than the expected one.

8.7 Clutter attenuation

Another set of experiments in order to define which parameters affect detection step has been performed [171]. In this section how the Signal-to-Clutter Ratio (SCR) affects performances of detection algorithm is considered. Since incoherent averaging presents the problem of losing an huge amount of information (e.g. optimal threshold cannot be estimated, therefore some potential target features can be accidentally removed and classification, as a consequence, fails), other approaches more information preserving have been investigated. One of them has been inherited and modified by mammogra-

phy image processing. As reported in [138], by applying a sigmoid function and fixing a threshold value it is possible to increase the contrast of a blurred and darkness image. In our case, however we are interested in the dual problem: reduce the value of the clutter before detection, therefore we modify the equation as:

$$O = \begin{cases} i & \text{if } i > t \\ i - (i\frac{c}{(1+e^{-i})}) & \text{if } i \leq t \end{cases} \tag{8.22}$$

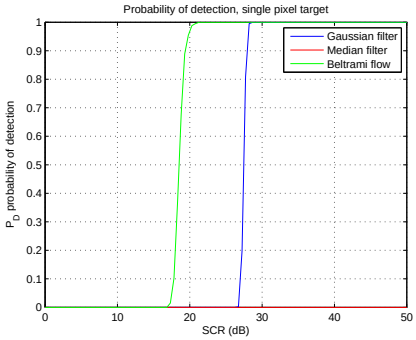
where c is the percentage of intensity reduction ($c = 0.9$ in our case), i is the intensity of the analysed pixel and t is a threshold (in out case $T_c = \mu_c + 1\sigma_c$ (i.e. μ_c is the intensity clutter mean value, whereas σ_c is the intensity clutter standard deviation) of the clutter computed on an homogeneous clutter area).

As for the experiments, two possible scenarios were considered:

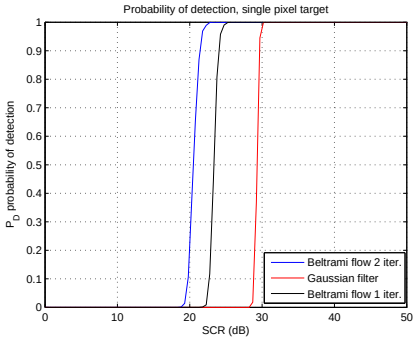
1. Exponential distributed single pixel target;
2. Exponential distributed extended pixel target:
    - central pixel detection;
    - corner pixel detection, as reported in Figure 8.7.

Moreover the detection after despeckling has been performed for the cases of Gaussian and median filters as well as Beltrami flow. As for the detection threshold for a set of K-distributed images ($\nu = 8$, $L = 1$), the values are reported in Table 8.1. In the case of an extended target, the previous set of images was modified by introducing an exponential distributed squared object (i.e. $10 - by - 10$ pixels) at the centre of the scene.

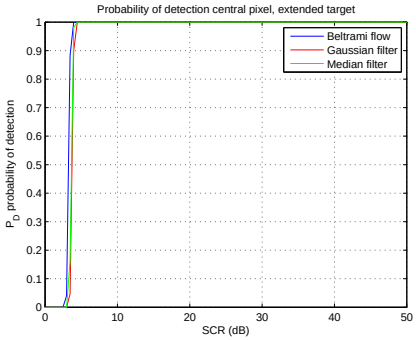
Figure 8.7(a) reports the probability of detection for an Exponential dis-



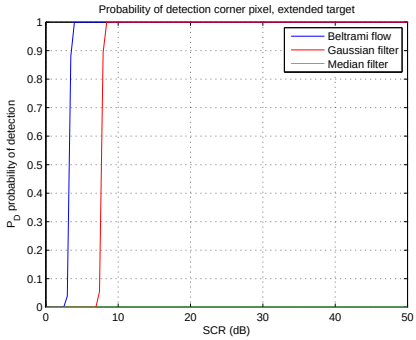
(a) P_D single pixel target comparison, $P_{fa} = 10^{-6}$



(b) P_D single pixel target for iterative Beltrami flow, $P_{fa} = 10^{-9}$



(c) P_D extended pixel target, central pixel $P_{fa} = 10^{-6}$



(d) P_D extended pixel target, corner pixel $P_{fa} = 10^{-6}$

Figure 8.7: *Probability of detection for an exponential distributed single and extended targets*

tributed single target. The results show thet the median filter tends to remove any target independently of the SCR value, whereas Beltrami flow performs better in terms of information preserving (i.e. it 'removes' less single targets than the Gaussian filter).

In Figure 8.7(b) the effects of the numbers of iterations for Beltrami flow is investigated and compared with the Gaussian filter. As reported Beltrami performs better results than the Gaussian filter when the number of iterations increase.

Figure 8.7(c) reports the results of the detection of the central pixel in an Exponential distributed extended target when the clutter suppression is performed. With respect the previous two images, clutter suppression improves the results. Moreover the presence of a neighbor near the central pixel increases the detection performance as well (i.e. in this case also the Median filter is able to preserve information of the central pixel).

However in Figure 8.7(d) the situation become worse in the case of the detection of the corner pixel. Beltrami flow produces the best results, despite the required SCR with respect to Figure 8.7(c) is bigger (i.e. $P_{fa} \leq 10^{-6}$), and the results tend to confirm the edge preserving properties of the algorithm, the Gaussian filter tends to smooth the edges more than Beltrami flow. As for the Median filter, it tends to remove the information regarding the object corners.

8.8 Hough Transform correction

As reported in the previous section, some algorithms tend to introduce a degree of distortion into SAR images which corrupted the geometrical and statistical properties of the potential targets. In particular, the Gaussian filter and Median filter tend to remove important information especially near to the potential targets edges. In order to restore this kind of information, some image transformation can be adopted. This section therefore explains how a backward procedure can be used in order to preserve the information of the samples.

Hough transform is a normal parametrization for lines [139], adopted as a feature extraction techniques in image analysis, computer vision and digital image processing. The classical Hough transform has been concerned with the identification of lines in the image but it has been extended to identifying objects of arbitrary shapes. As illustrated in Figure 8.8, this parametrization specifies a straight line by the angle θ of its normal and its algebraic distance ρ from the origin. The equation of the line corresponding to this geometry is:

$$\rho = x\cos\theta + y\sin\theta \tag{8.23}$$

If θ is restricted to the interval $[\,0, \pi\,)$, then the normal parameters for a line are unique. It is therefore possible to associate uniquely each line of the image a pair (ρ, θ). The (ρ, θ) plane is sometimes referred to as *Hough space* for the set of straight lines in two dimensions. Equation (8.23) corresponds hence to a sinusoidal curve in the (ρ, θ) plane, which is unique to a point of the line. If the curves corresponding to two points are superimposed, the

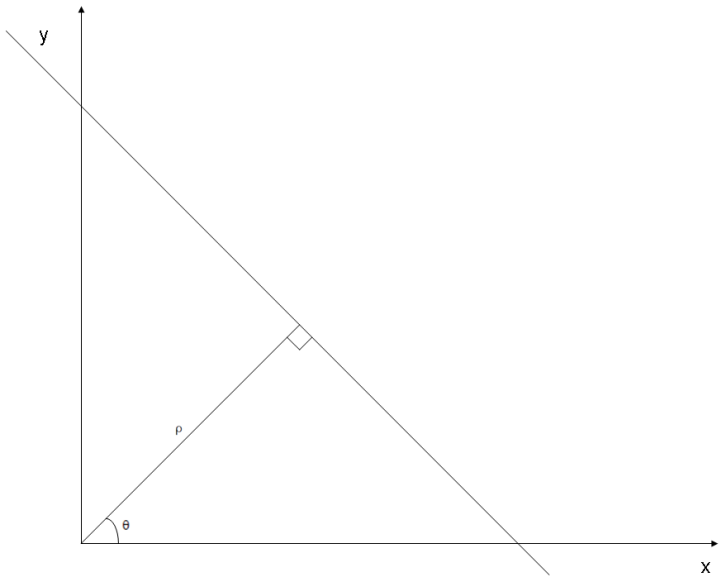


Figure 8.8: *Hough transform parameters*

location in the Hough space where they cross corresponds to a line in the original image space that passes through both points. Generally hence, a set of points placed on the same line in the original image will produce sinusoids which all cross parameters for that line. As a consequence the problem of line detection can be converted in a problem of finding concurrent curves.

As for the implementation of the Hough transform, it consists in quantizing the Hough parameter space (ρ, θ) into finite intervals stored in array, called an accumulator. As the algorithm runs, each image point (x_i, y_i) is transformed into a discretized (ρ, θ) curve and the accumulator cells which lie along this curve are incremented. The resulting peaks in the accumulator array represent strong evidence that a corresponding straight line exists in the image, as reported in Figure 8.9 and the corresponding Hough transform in Figure 8.10.

As for our purpose a simple line, as reported in Figure 8.11(a), and its cor-

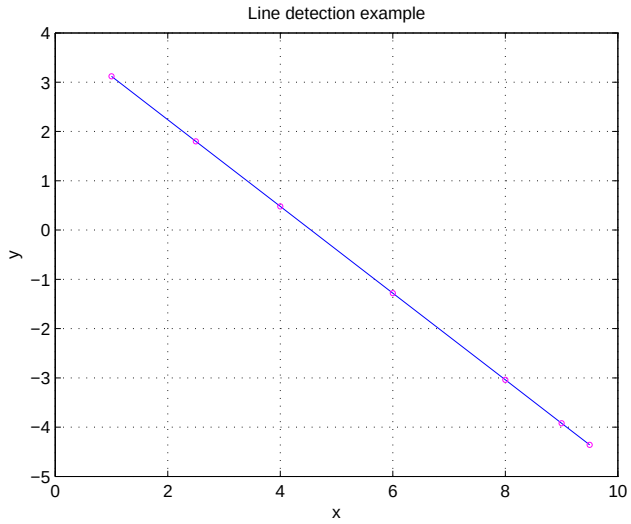


Figure 8.9: *Example of a line*

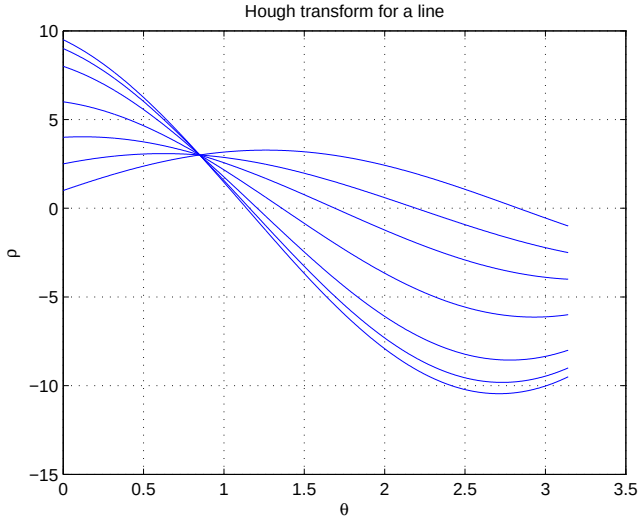
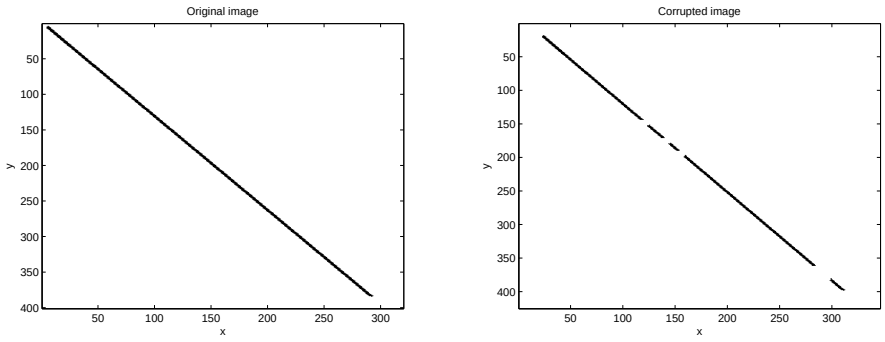
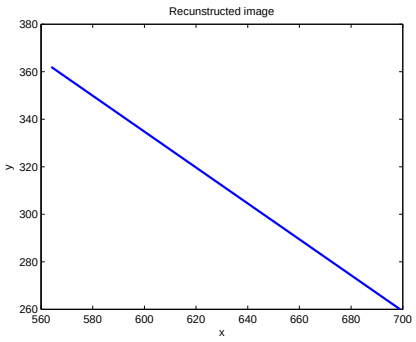


Figure 8.10: *Hough transform for a line*

rupted version, as reported in Figure 8.11(b), were considered. The Hough transform was considered and then from the Hough space the image was restored, as reported in Figure 8.11(c). As shown, the Hough transform could help in correcting the distortion introduced by a despeckling algorithm before detection processing. As for the correction of distortions introduced by



(a) Original line (b) Corrupted line



(c) Hough transform output

Figure 8.11: *Hough Transform correction for a line*

despckling algorithms (e.g. median filter, which remove pixels at the corner of the potential targets), the idea can be summarized as follows: firstly the Hough Transform is computed at corner edges (i.e. one checks from Hough space if a line is present for the first two rows and columns of the edge corner

'*A*' in Figure 8.12) if the lines are present, then the corner is reconstructed as an average of adjacent pixels.

In conclusion the proposed algorithm consists of four steps:

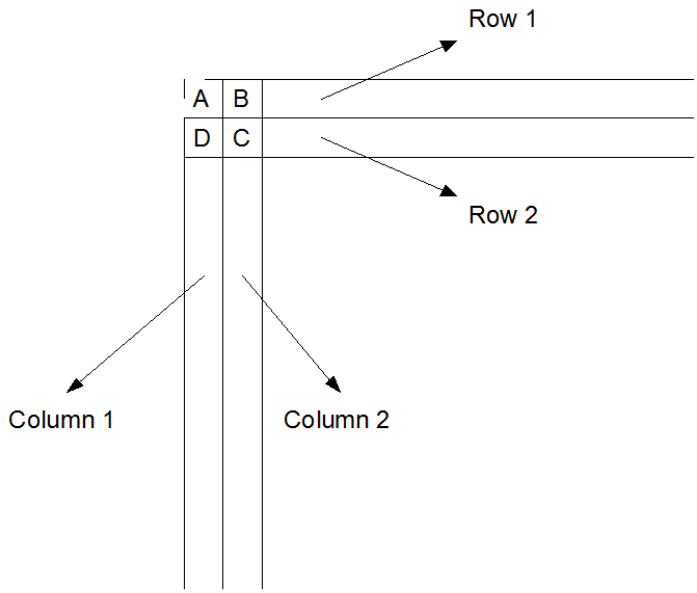


Figure 8.12: *Hough Transform correction: Hough transform is used to evaluate the presence of Row* 1 *and* 2 *as well as Column* 1 *and* 2. *If the lines are detected, then the pixel corner A (dashed) is computed as an average of pixels B, C and D.*

1. Hough transform is applied to the original images in order to determine the presence of corners;

2. The Median filter is applied to the SAR images;

3. Hough transform correction is applied to the images (i.e. the intensity of replaced pixels are an average of the neighbors);

4. CFAR algorithm is performed;

In terms of detection, a CFAR algorithm and CFAR plus Hough transform correction algorithm in terms of P_d (pixel at the corner) for a set of K-distributed images ($100 - by - 100$ pixels) plus an Exponential distributed squared object ($20 - by - 20$ pixels) placed at the centre of the scene were compared. The results are reported in Figure 8.13. The comparison is between the median filter (corrected), Beltrami flow and the Gaussian filter (uncorrected).

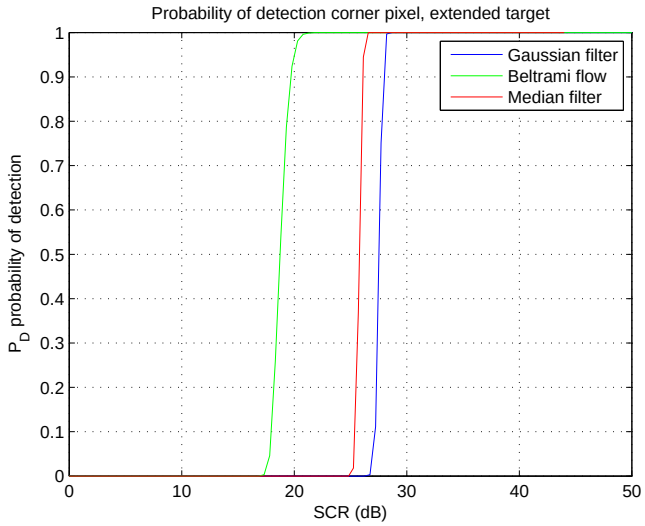


Figure 8.13: *Hough Transform correction: Probability of detection of the corner for a set of SAR images filtered with a Median filter. Result is compared with Beltrami flow and Gaussian filter outcomes (uncorrected).*

8.9 Summary

Chapter 8 analyses the most important issues of the detection step. Firstly an introduction of detection processing has been given. Secondly, methods for increasing the detection ability of an ATR/SAR system (i.e. Incoherent

averaging and clutter attenuation) have been described. The analysis allowed us to stress the most important parameters of detection step (i.e. SCR and the clutter modelling). Moreover the analysis allowed us to understand the importance of the models clutter tail and its effects on the detection of potential targets features. Finally an example of backward function analysis have been developed in order to solve an unwanted distortion introduced by the median filter (i.e. Hough transform correction).

Chapter 9

Conclusions

9.1 Overview of chapters

In this Chapter the conclusion of the thesis are reported. Before writing them, an overview of previous Chapters will be given and a Brief conclusion will be reported.

9.1.1 Chapter 2

In Chapter 2 an introduction of the research on ATR systems has been reported. Firstly a general introduction on image classification and the most important techniques for improving classification performance have been introduced, then the SAR/ATR literature survey was reported: Comparison of ATR methods in SAR image classification (i.e. MSE, HMM, SVM and non-linear classifiers); Survey on most important parameters which affect the systems performances (i.e. investigations on different polarimetric techniques applied to ATR problems, importance of class-skew, etc.); Introduction of

several features (i.e. RCS) used in SAR/ATR systems and the most important investigation on how the features affect performance of ATR systems as well as their best representation (i.e. scattering centre); Methods for understanding how to measure the accuracy of the metrics for ATR performances and the most important techniques used to assess them. Finally Computational performances for SAR/ATR systems as well as SAR/ATR datasets and Multi-sensor data fusion have been briefly reported.

9.1.2 Chapter 3

In Chapter 3 an overview of the most important theoretical backgrounds of the thesis has been reported. Firstly, an introduction to Bayes theory and a Bayes classifiers has been reported, then the most important issues about the information flow model have been introduced and the most important problem about class-separability has been analysed. Secondly some methods for the analysis of information preservation have been described. In particular, the chapter was concerned with the dependency of the information flow on the sensor structure, describing and comparing some techniques which allow assessing the measure of the variation of information content through the processing chain (i.e. how the sample set changes through the processing chain) as well as how to measure the variation content in terms of class-separability (i.e. how to assess the degree of class-separability).

In this Chapter a model for information flow through the ATR/SAR sensor has been also introduced. It has been useful in order to determine the limits of subsystems and to estimate the variation of information content flow.

Moreover the most popular metrics for assessing class-separability have been investigated. Unfortunately most metrics failed when a constraint on the maximum tolerated type II error is considered. As reported in Section 3.5 the Hausdorff dimension and statistical fractal model theoretically are suitable for estimating the degree of class-separability. Since indeed Hausdorff dimension is related to the Lebesgue's measure theory, it is a important mathematical tool for assessing if the condition of class-separability is satisfied (i.e. the violation of value of Hausdorff dimension is the condition of non separability of sample classes). Despite this approach is mathematically suitable, it fails when samples data are considered. Indeed the class skew (i.e. varying, unequal occurrences of individual classes) for sampled data does not produce a meaningful results (i.e. Housdorff dimension for set of point is null). However the Hausdorff dimension problem allowed us to understand better the class-separability problem, i.e. reducing error area increases class separability. Hence a class separability metric requirements is to be able to measure error area.

In this chapter an introduction of the most powerful techniques used to assess the information content variation in ATR systems have been described. Particularly PCA, PDF mapping and Differential Geometry approach have been compared.

PCA has been considered not suitable because it is a linear method, therefore it works properly for Gaussian distributed samples and for linear transformation which are variance property preserving (i.e. they do not change the distribution properties). In case of non linear transformation however Kernel Tricks are necessary (i.e. Kernel PCA). Moreover PCA method is based on

a belief that large variances have important dynamics, i.e. large variances have a high Signal-to-Noise Ratio, i.e. a high information content, despite this requirement is often not satisfied by sample data (i.e. the sample set is not able to represent the whole set property properly).

Another proposed criterion consists in considering the PDF of class-samples and by transforming them through the information flow model compute the resulting PDF. As for SAR systems, some processing steps (e.g. detection) are not linear therefore computing non-linear computed PDF can be not feasible.

A more reliable method is considering a set of samples points in a non-Euclidean space and then transform them through the information flow model by using properties of differential geometry and Lie groups. An approximation of this method can be performed by unscented theory which approximately a generic transformed random variable distribution as a Gaussian distribution. This method can be considered as an upper bound for our problem analysis, i.e. the performance in the worst case design can be determined.

9.1.3 Chapter 4

The fourth chapter was arranged in order to report and compare the performances of the most common methods used to estimate the degree of class-separability. Moreover new methods have been considered as metrics for assessing the degree of class-separability.

Covariance based methods show the disadvantage either being dependent on

a priory probability of classes (i.e. LDA) or it is suitable for distributions whose 'shape' is completely described by the covariance matrix (i.e. Gaussian), besides it does not give any information on the error area, i.e. no guarantee that requirements on false alarm rate are satisfied.

As for thresholding criteria, they are not able to assess any information related to error area. Moreover some measures are not suitable for class-separability estimation (i.e. their values are not able to determine if two classes are separate and at what degree). Chernoff and Bhattaccharrya distances indeed give an upper bound for error area, but they are not suitable for sampled data. Moreover they can be computed in a closed form only for Gaussian distributed samples. Mahalanobis distance is suitable only for Gaussian distributions as well, because it depends on the covariance matrix, but it is not able to give any information about neither the class-separability or the degree of class-separability. Moreover Mahalanobis distance in unidimensional case for normally distributed samples is equal to the Kolmogorov-Smirnov Threshold (T_{KS}), which is the most reliable threshold among the analysed ones.

The KAPS metric however is suitable especially for sampled data, it is independent of sample distribution and it is related to maximum tolerated Type II error. Compared with ROC analysis (which is difficult to interpret) and AUC (which is not defined in a unique mode) indeed the KAPS metric is able to emphasize the problem, often encountered in pattern recognition, reported in Figure 4.10 and Figure 4.12, i.e. when the maximum tolerated Type II error performance requirements are reached as provided ω_2-class is classified as ω_1.

9.1.4 Chapter 5

In this chapter the effects of sample size have been investigated. Since T_{MHD} is dependent on the number of samples, it is difficult to determine its variation with respect to the number of samples considered, therefore its order statistics analysis have been investigated. However as for the T_{KS} value a Monte Carlo simulation has been preformed, because its order statistic analysis was too complicated to be performed. As demonstrated by results the KAPS value, for Exponential and Gaussian example described by equations (4.17) and (4.19) is independent of number of samples, nevertheless it is suggested to use a number of samples of the order 10^3.

9.1.5 Chapter 6

In this Chapter which is the best metric for SAR/ATR systems denoising step were investigated. Two common metrics adopted by researchers in image processing were analysed in detail as well as the afore-mentioned metrics with NVM as described in section 6.3 were compared.

MSE should be rejected because it is not able to give any useful information on speckle reduction as it is dependent on the features of the image not degrading and therefore raising the MSE value, whereas ENL and NVM can be considered as dual. They indeed give us information about reconstructed RCS and removed speckle, which are often in contrast, therefore it is suggested to use both of them contemporaneously in order to establish that no artifacts were introduced or too much structure from the observed scene was removed: ENL indeed ensures that the ratio between the filtered image and

original one is Gamma-distributed, but unfortunately it does not ensure that the RCS has been reconstructed in the correct way. Vice-versa NVM confirms that the Gamma-distributed RCS has been estimated correctly if its value is equal to the actual number of looks, but it does not certify that the speckle has been removed correctly.
Ideally an algorithm should give the same value for both metrics, i.e. the actual number of looks, to establish the correct separation of noise (i.e. speckle) from actual RCS.

9.1.6 Chapter 7

This Chapter was addressed in order to compare the performances of three algorithms (i.e. Gaussian filter, Median filter and Beltrami flow) used for despeckling. They were analysed in terms of RCS reconstruction and as edge preserving algorithms, because the features of SAR/ATR images are crucial to perform a correct target classification.
First one can affirm that the ENL metric is more robust than the NVM with respect to changing of RCS, despite the fact that it gives no information about distortion in the output images. However the latter has shown that a distant value from the actual number of look means that the tail of output images PDF tend to zero faster than an algorithm which has a NVM value near to L. This could be appreciable because in CFAR step clutter tail plays an important role.
As for RCS filter reconstruction performances, no algorithm is able to estimate correctly the RCS PDF.

The Gaussian filter is a good despeckling algorithm especially for low order parameter ν, whereas for higher order parameters the RCS reconstruction depends on the filter variance and as a consequence on filter window size. Beltrami flow however makes more evident the dual problem ENL/NVM because by increasing the number of iteration better denoising performances are achieved, but at the same time the RCS distortion in output images increases as well. The median filter tends to estimate better the RCS and removed speckle PDFs by increasing the window size filter.

As for edge preserving properties, Beltrami performs the best results (in terms of number of iterations and window size), followed by the median filter which introduces less distortion in edge-height than the Gaussian filter. In terms of slope angle it seems to be as good as the Gaussian filter.

The Gaussian filter produces the worst performance as preserving geometrical features of potential targets because it tends to smooth the images too much.

In conclusion the Beltrami flow seems to be the best candidate as a SAR/ATR despeckling algorithm, because it performs the best trade-offs between a despeckling action and edge preserving. Beltrami was adopted by Barbaresco in [148] as CFAR subsystem for Doppler and polarimetric data. In their articles the authors argued that Beltrami cannot be considered an edge preserving algorithm because, from a geometrical point of view, it tends to minimize the mean curvature [130] of the edges, instead of computing the principal curvature direction properly ([131], [132] and [130]). The performed experiments confirmed the Barbaresco results, indeed edge preserving properties are not numerically satisfied (i.e. Slope Angle of filtered edges is little less

than 90°, as expected). Moreover Beltrami flow performances depend on the SCR value (i.e. for SCR$\leq$ 1 dB the edges are considered as clutter, therefore cancelled, whereas for SCR$\geq$ 5 dB Beltrami flow can be considered an edge preserving algorithm. As for 1 <SCR< 5 dB the algorithm tends to preserve more and more the edges).

9.1.7 Chapter 8

This Chapter focused on the efficiency of an approximated local clutter model. Three models have been investigated: a local approximation, Exponential and Gamma clutter model. The results confirm that a local approach can be considered more suitable than a global model in terms of the detection threshold estimation as well as model fitting of the clutter tail. As a consequence the information content of detection input can be preserved/emphasized better (e.g. estimation of SAR/ATR parameters for the discrimination of potential targets such as Mass, Diameter, Rotational inertia, Percent bright CFAR, Standard deviation etc. [9] can be evaluated better) if an approximating local clutter model is adopted. In this Chapter a Weibull model has been adopted, but also Gaussian and Log-normal CDFs can be adopted as approximating the CDF if necessary.

As for the parameter selection at the detection stage, an important role in object information preserving is played by SCR, nevertheless detection performances are strictly related to the CFAR threshold (i.e. clutter model) and to the P_{fa} requirements as well. Moreover the detection step has to consider the distortion introduced by the despeckling algorithm and try to

minimize them in order to preserve information/remove unwanted information (i.e. Hough correction). Finally as for the clutter models, it has been proved that Exponential distribution can be considered as an upper bound for land clutter distributions (i.e. the Exponentially distributed clutter tail is longer than Weibull and K-distributed ones), whereas Weibull represents an upper bound in case of sea clutter.

9.2 Thesis Conclusion

In this thesis several aspects of the limits in SAR/ATR systems have been investigated. Firstly the condition under which the class-separability is ensured have been investigated and how to assess the degree of class-separability. A necessary condition for class-separability is that the error area defined in Chapter 3 ideally has to be equal to zero. The investigation on the class-separability produced a new metric for assessing the class-separability, defined as KAPS , which is related to the error area minimization problem when a constraint on the Type II error is considered. KAPS , hence, is suitable to compare the performances of different signal processing chains for SAR/ATR systems.

In Chapter 3 another problem has been analysed, indeed a new information flow model has been introduced in order to assess how the information content changes through the processing chain. The most interesting approach is given by the differential geometry and Lie groups, because a set of samples can be mapped in the feature space so that their variation can be assessed mathematically. As for dimensionality reduction, it can be considered as a

non *bi-Lipschitz* mapping of the samples through the feature space, therefore the geometrical method helps in understanding where the bottlenecks of information flow are located. Unfortunately no investigated methods are able to perform properly a backward procedure for the selection of the optimal parameters of each subsystem of a SAR/ equipment, therefore it is desirable to compare qualitatively the outcomes of the subsystems with respect to the ideal response in order to find the optimal subsystem parameters.

The advantages of the information flow model has been used in the case studies, where a cascade of two subsystems has been considered. Indeed denoising processing and detection are strictly related, therefore the selection of denoising parameters (i.e. the despeckling algorithm) can affect the performances of the detection algorithm, in terms of the clutter model as reported in Chapter 8. As reported in Chapter 7, however, the analysis of the outcomes of filtered images, gave us important information regarding the information content of SAR images (i.e. how the despeckling algorithm distorted the input in terms of clutter modelling and edge preserving information, which are very important parameters for detection and discrimination). Moreover the comparison of the expected ideal output of the despeckling allowed the researchers understanding which are the optimal criteria for the selection of the denoising algorithm.

In conclusion, the thesis has been able to define the definition of class-separability and of the degree of separability and as a consequence to understand which are the conditions under which the separability is ensured. Moreover by defining the forward and backward procedure analysis it is possible to define the parameters which play an important role in each subsystem of an ATR processing

chain as well as to set the processing chain in order to preserve/maximize the information content.

Appendix A

Publications

G. Marino, E. J. Hughes, *Information Content Variation In CFAR Processing*, Cranfield University Research Student Symposium, May 2011.

G. Marino, E. J. Hughes, *Automatic Target Recognition in Synthetic Aperture Radars*, 9th Electro-Optics & Infrared Conference, Shrivenham, 4th/6th July 2011.

G. Marino, E. J. Hughes, *A novel mathematical approach for the problem of CFAR clutter model approximation*, 3rd Microwaves, Radar and Remote Sensing Symposium, Kiev August 2011.

Bibliography

[1] Smith, G.E. *Radar target micro-Doppler signature classification* Doctoral thesis, University of London 2008.

[2] Merril Skolnik *Introduction to radar systems 3rd edition* McGraw-Hill 2001.

[3] Burl, M. C., G. J. Owirka and L. M. Novak *Texture Discrimination in Synthetic aperture Radar Imagery* 23rd Asilomar Conf. Signals, Systems and computer, Pacific Grove, CA, 1989, pp.399-404.

[4] L. M. Novak, Gregory J. Owirka *The Automatic Target-Recognition System in SIAP* The Lincoln Laboratory Journal, Vol.10, No. 2, 1997.

[5] Nilubol C. Mersereau R. M., Smith M.J.T. *A SAR Target Classifier Using Radon Transforms and Hidden Markov Models* DSP, Vol. 12, Issues 2-3, 2002.

[6] Paul J.L., Lupo J.C. *From tanks to tumors* Engineering in Medicine and Biology Magazine, IEEE, Vol. 21, Issue 6, November 2002.

[7] P. Jorion *Value at Risk: The New Benchmark for Managing Financial Risk (3rd ed.)* McGraw-Hill 2006.

[8] H.M. Markowitz *Portfolio Selection* The Journal of Finance, March 1952.

[9] Chris Oliver *Understanding Synthetic Aperture Radar Images* Artech House 1998.

[10] Tough, R. J. A. and K. D. Ward *The generation of correlated K-distributed noise* DRA technical report DRA/CIS/CBC3/WP94001/2.0, 1994, DRA Malvern, St. Andrews Road, Malvern, Worcs.

[11] Zelnio E., Garber F. *A characterization of ATR performance evaluation* Proc. of Signal Processing, Sensor Fusion and Target, SPIE, Vol. 2755, 1996.

[12] Lu D., Weng Q. *A survey of image classification methods and techniques for improving classification performance* International Journal of Remote Sensing Vol. 28, No. 5, March 2007.

[13] E. G. Zelnio *Advanced decision-making systems in future avionics : automatic target recognition example* Aerospace conference Proc. IEEE, Vol. 1, 1998.

[14] J. A. Ratches, C. P. Walters, R. G. Buser, and B. D. Guenther *Aided and automatic target recognition based upon sensory inputs from image forming systems* IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 19, No. 9, 1997.

[15] L. M. Novak *State-of-art of SAR automatic target recognition* IEEE International radar conference, Alexandria, USA, 2000.

[16] L. M. Novak, G. J. Owirka, and W. S. Brower *An efficient multi-target SAR ATR algorithm* Asilomar 32, Vol. 1, p3-13, 1998.

[17] L. M. Novak, G. J. Owirka, and A. L. Weaver *Automatic target recognition using enhanced resolution SAR data* IEEE Trans AES, Vol. 35, No. 1, 1999.

[18] G. J. Owirka, S. M. Verbout and L. M. Novak *Template-based SAR ATR performance using different image enhancement techniques* Algorithms for SAR VI, SPIE Proc., Vol. 3721, p302-319, 1999.

[19] L. M. Novak, G. J. Orwirka, and C. M. Netishen *Performance of a high-resolution polarimetric SAR automatic target recognition system* The Lincoln Laboratory Journal, Vol. 6, pp11-24, 1993.

[20] L. M. Novak, M. C. Burl, W. W. Irving *Optimal polarimetric processing for enhanced target detection* IEEE Trans AES, Vol. 29, pp234-244, 1993.

[21] R. D. Chaney, M. C. Burl, L. M. Novak *On the performance of polarimetric target detection algorithms* IEEE International radar conference, pp 520-525, 1990.

[22] L. M. Novak, G. J. Owirka, and W. S. Brower *Performance of 10 and 20 target MSE classifier* IEEE Trans AES, Vol. 36, No. 4, 2000.

[23] R. Kohavi, F. Poster *Editorial for special issue on Application of Machine Learning and the knowledge Discovery Process* Machine Learning, Vol. 30, No. 2-3, 1998.

[24] M. Cetin, W. C. Karl, and D. A. Castanon *Feature enhancement and ATR performance using non-quadratic optimization-based SAR imaging* IEEE Trans AES, Vol. 39, No. 4, 2003.

[25] J. S. Ahn, B. Bhanu *Model-based recognition of articulated objects* Pattern Recognition Letters 23, pp 1019-1029, 2002.

[26] B. Bhanu, Y. Lin *Stochastic models for recognition of occluded targets* Pattern Recognition 36, pp 2855-2873, 2003.

[27] G. J. Ettinger, G. A. Klanderman, W. M. Wells, W. E. L. Grimson *A probabilistic optimization approach to SAR feature matching* Algorithms for Synthetic Aperture Radar III, Proc. SPIE 2757, pp 318-329, 1996.

[28] E. R. Keydel, S. W. Lee *Signature prediction for model-based Automatic Target Recognition* Algorithms for Synthetic Aperture Radar III, Proc SPIE 2757, pp 306-317, 1996.

[29] H. J. Wolfson, I. Rigoutsos *Geometric hashing : an overview* IEEE Computational science and engineering, Vol. 4, pp 10-21, 1997.

[30] H. Chiang, R. L. Moses, L. C. Potter *Model-based classification of radar images* IEEE Trans Information Theory, Vol. 46, No 5, 2000.

[31] H. Chiang, R. L. Moses, L. C. Potter *Model-based Bayesian feature matching with application to synthetic aperture radar target recognition* Pattern Recognition 34, pp 1539-1553, 2001.

[32] Keinosuke Fukunaga *Introduction to statistical pattern recognition 2nd edition* Academic Press Limited, 1990.

[33] K. Fukunaga, R. R. Hayes *Estimation of classifier performance* IEEE Trans Pattern Analysis and Machine Intelligence, Vol. 11, No. 10, 1989.

[34] K. Fukunaga, D. L. Kessell *Nonparametric Bayes error estimation using unclassified samples* IEEE Trans. Information Theory, IT-19, No. 4, 1973.

[35] C. Daniell, A. Mahalanobis, R. Goodman *Object recognition in sub-band transform-compressed images by use of correlation filters* Applied Optics, Vol. 32, No. 32, 2003.

[36] S. Suvorova, J. Schroeder *Automated target recognition using the Karhunen-Loeve transform with invariance* DSP, Vol. 12, pp 295-306, 2002.

[37] Q. Zhao, J.C. Principe *Support Vector Machines for SAR Automatic Target Recognition* IEEE Trans. on Aerospace And Electronic Systems, April 2001.

[38] A. Jain, D. Zongker *Feature Selection: Evaluation, Application and small Sample Performance* IEEE Trans. on Pattern Analysis and machine intelligence, Vol. 19. No. 2, February 1997.

[39] P. Pudil, J. Novovicova, J. Kittler *Floating Search Methods in Feature Selection* Pattern Recognition Letters, vol. 15, pp. 1,119-1,125, November 1994.

[40] J. O'Sullivan, M.D. DeVore *SAR ATR Performance Using a Conditionally Gaussian Model* IEEE Tran. on Aerospace and Electronic Systems, Vol. 37, No.1, January 2001.

[41] M. D. DeVore, J. A. OSullivan *Performance complexity study of several approaches to automatic target recognition from SAR images* IEEE Trans. AES, Vol. 38, No. 2, pp 632-648, 2000.

[42] A. Mahalanobis, V.A. Forman *Multi-class SAR ATR using shift-invariant correlation filters* Pattern recognition, Vol. 27, Issue 4, April 1994.

[43] B. Bhanu, Y. Lin *Adaptive Target Recognition* Machine Vision and Applications, 2000.

[44] D. S. Rosario *Managing within-class target variability in SAR imagery with a target decomposition model* IEEE Int. Conf. Image Processing, Vol. 3, pp 935-938, 1996.

[45] E. Krogager *New decomposition of the radar scattering matrix* Electronics Letters, Vol. 26, pp 1525-1527, 1990.

[46] D. Turner *RESPECT: Rapid electromagnetic scattering predictor for extremely complex targets* IEE Proc., Vol. 137, part F, No. 4, August 1990.

[47] M. J. Gerry, L. C. Potter, I. J. Gupta *A parametric model for synthetic aperture radar measurements* IEEE Trans. Antennas and Propagation, Vol. 47, No. 7, July 1999.

[48] M. J. Coulombe, T. Horgan, J. Waldman, J. Neilson, S. Carter, W. Nixon *A 160 GHz polarimetric compact range for scale model RCS measurement* Proc. Antenna measurements and techniques association, pg 239, October 1996.

[49] R. H. Giles, H. J. Nielson, D. M. Healy Jr., T. Grayson, R. Williams, W. Nixon *Acquisition and analysis of X-band moving target signature data using a 160 GHz compact range* Proc. SPIE, Vol. 4379, ATR XI, pp 289-299, April 2001.

[50] R. H Giles, W. T. Kersey, A. J. Gatesman, M. J. Coulombe, M. S. McFarlin, R. Finley, W. Nixon *A study of the X-band radar signature characteristics for main battle tanks in operational environments* Proc. SPIE, Vol. 4718, pp 336-343, April 2002.

[51] D. Blacknell *Statistical target behaviour in SAR images* IEE Proc. Radar, Sonar, and Navigation, Vol. 147, Issue 3, pp 143-148, 2000.

[52] L. C. Potter and R. L. Moses *Attributed scattering centers for SAR ATR* IEEE Trans. Image Processing, Vol. 6, No. 4, 1997.

[53] B. Bhanu, G. Jones III *Recognizing target variants and articulations in synthetic aperture radar images* Optical Engineering, Vol. 39, No. 3, 2000.

[54] G. Jones III, B. Bhanu *Recognizing occluded objects in SAR images* IEEE Trans. AES, Vol. 37, No. 1, 2001.

[55] G. Jones III, B. Bhanu *Quasi-invariants for recognition of articulated and non-standard objects in SAR images* IEEE workshop on computer vision beyond the visible spectrum, Fort Collins, CO, pp 88-97, 1999.

[56] G. Jones III, B. Bhanu *Recognizing articulated objects in SAR images* Pattern recognition, Vol. 34, pp 469-485, 2001.

[57] G. Jones III, B. Bhanu *Increasing the discrimination of synthetic aperture radar recognition models*, Optical Engineering, Vol. 41, No. 12, 2002.

[58] S. V. Stehman *Selecting and interpreting measures of thematic classification accuracy* Remote Sensing and Environment, Vol. 62, pp 77-89, 1997.

[59] R. G. Congalton *A review of assessing the accuracy of classifications of remotely sensed data* Remote Sensing and Environment, Vol. 37, pp 35-46, 1991.

[60] D. E. Dudgeon *ATR performance modeling and estimation*, DSP, Vol. 10, pp 269-285, 2000.

[61] A. M. Horne *An information theory for prediction of SAR Target classification performance* Algorithms for synthetic Aperture Radar Imagery VII, Edmund G. Zelnio Editor, Proc. of SPIE Vol. 4382, 2001.

[62] S. D. Briles *Information-theoretic performance bounding of Bayesian identifiers* Automatic Object Recognition III, Proc. SPIE, Vol. 1960, 1993.

[63] F. Garber, E. Zelnio *On some simple estimates of ATR performance and initial comparisons for a small data set*, SPIE Algorithms for Synthetic Aperture Radar IV, Proc. SPIE 3070, pp 150-161, 1997.

[64] K. Takkola *Feature extraction by Non-parametric Mutual Information Maximization*, Journal of Machine learning research, Vol. 3, 2003.

[65] M. L. Cooper, M. I. Miller *Information measures for object recognition accommodating signature variability*, IEEE Trans. Information Theory, Vol. 46, No. 5, 2000.

[66] F. Kanaya, K. Nakagawa *On the Practical Implication of Mutual Information for statistical Decisioning*, IEEE Trans. Information Theory, Vol. 37, No.4, July 1991.

[67] M. I. Miller, P. Moulin, K. Ramchandran *Information-theoretic bounds on target recognition performance based on degraded image data* IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No.9, September 2002.

[68] R. O. Duda *Pattern recognition and scene analysis* John Wiley & Sons, 1973.

[69] M. Boshra, B. Bhanu *Predicting an upper bound on SAR ATR performance* IEEE Trans. AES, Vol. 37, No.3, July 2001.

[70] M. Boshra, B. Bhanu *Predicting the performance of object recognition*, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, No. 9, 2000.

[71] M. I. Miller, U. Grenander, J. A. OSullivan, D. L. Snyder *Automatic target recognition organised via jump-diffusion algorithms* IEEE Trans. Image Processing, Vol. 6, No. 1, 1997.

[72] U. Grenander, M.I. Miller, A. Srivastava *Hilbert-Schmidt lower bounds for estimators on matrix Lie groups for ATR* IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 8, September 1998.

[73] W. Irving, R.B. Washburn, W. Grimson, L. Eric *Bounding performance of peak-based target detectors* Algorithms for Synthetic Aperture Radar Imagery IV, Edmund G. Zelnio Editor, Proceedings of SPIE Vol. 3070, 1997.

[74] B. Bhanu, J. H. Yi *Target indexing in SAR images using scattering centers ad the Hausdorff distance*, Pattern Recogntion Letters, Vol. 17, 1996.

[75] Z. Knowles, D. Parker *A Monte Carlo simulation based approach to a priori performance prediction for target detection and recognition in cluttered synthetic aperture radar imagery*, Target Tracking 2004: Algorithms and Applications, IEE.

[76] G.E. Smith, M. Vespe, K. Woodbridge, C.J. Baker *Radar classification evaluation*, IEEE Radar Conference 2008.

[77] J. A. Richards, W. J. Bow, B. K. Bray *An informative confidence metric for ATR*, Algorithms for SAR X, Proc. SPIE, Vol. 5095, pp 336-348, 2003.

[78] T. D. Ross *Confidence intervals for ATR performance metrics* Algorithms for Synthetic Aperture Radar Imagery VIII, Proc. SPIE, Vol. 4382, pp 318-329, 2001.

[79] U. Grenander, A. Srivastava, M. I. Miller *Asymptotic performance analysis of Bayesian target recognition*, IEEE Trans. Information Theory, Vol. 46, No. 4, 2000.

[80] L. L. Horowitz, G. F. Brendel *Fundamental SAR ATR performance predictions for design tradeoffs*, SPIE Algorithms for Synthetic Aperture Radar IV, Vol. 3070, 1997.

[81] M. D. Devore, J. A. OSullivan, R. D. Chamberlain, M. A. Franklin *Relantionships between computational system performance and recognition system performance*, Automatic Target Recognition XI, Proc. SPIE 4379, pp 355-363, 2001.

[82] M. D. DeVore, R. Chamberlain, G. Engel, J. A. OSullivan, M. A. Franklin *Trade-offs between quality of results and resource consumption in a recognition system*, Proc. IEEE Conference on application specific systems, architectures and processors, pp 391-402, 2002.

[83] D. Andre, D. Blacknell, J. Hare *Generation and validation of a simulated radar ground target database* SET080, NATO conference, 2004.

[84] E. Keydel, W. Williams, R. Sieron, V. Rajlich, and S. Stanhope *Reasoning support and uncertainty prediction in model-based vision SAR ATR*, Algorithms for Synthetic Aperture Radar VI, Proc. SPIE 3721, pp 620-631, 1999.

[85] B Bhanu and G Jones III *Object recognition results using MSTAR synthetic aperture radar data*, Proc. IEEE workshop on computer vision beyond the visible spectrum, Hilton Head, SC, pp 55-62, 2000.

[86] T. D. Ross, J. J. Bradley, L. J. Hudson, M. P. OConnor *SAR ATR-So Whats The Problem? An MSTAR Perspective*, SPIE Algorithms for SAR VI, Vol. 3721, pp 662-672, 1999.

[87] V. Velten, T. Ross, J. Mossing, S. Worrell, M. Bryant *Standard SAR ATR evaluation experiments using the MSTAR public release data set* Algorithms for Synthetic Aperture Radar V, Proc. SPIE 3370, pp 566-573, 1998.

[88] T. D. Ross and J. C. Mossing *The MSTAR evaluation methodology* Algorithms for Synthetic Aperture Radar VI, Vol. 3721, pp 705-713, 1999.

[89] C. Alexandrov, A. Draganov, N. Kolev*An application of Automatic Target recognition in Marine Navigation*, IEEE International Radar Conference, 1995.

[90] F. Qiang, Y. Wenxian *Automatic Target Recognition Based on Incoherent Radar Returns* Proc. of the IEEE 1995 National Aerospace and Electronics Conference, Vol.1, pp 123 - 128, 22-26 May 1995.

[91] A. K. Mishra, B. Mulgrew *Ground target classification for airborne bistatic Radar* 1^{st} EMRS DTC Technical Conference, Edinburgh 2004.

[92] R. Williams, J. Westerkamp,*Automatic target recognition of time critical moving targets using 1D High Range Resolution (HRR) Radar*, IEEE AES Magazine, Vol. 15, Issue 4, pp 37-43, April 2000.

[93] M. Vespe, C.J. Baker, H.D. Griffiths *Automatic target recognition using multi-diversity radar* IET Radar Sonar Navigation, Vol. 1, No. 6, pp 470-478, 2007.

[94] B.B. Mandelbrot *Self-affine fractal and fractal dimension* Physica Scripta, Vol. 32, 1985.

[95] K.J. Falconer *The geometry of fractal sets* Cambridge University Press, 1985.

[96] K.J. Falconer *Fractal geometry* John Wiley & Sons, 1990.

[97] A. Cichocki, S. Amari *Adaptive Blind Signal and Image Processing* John Wiley & Sons, 2003.

[98] D.C. Lay *Linear Algebra and its Application, 2nd ed.* Addison Wesley, 1996.

[99] P. Comon *Independent Component Analysis: a new concept?* Signal Processing, Elsevier, Vol. 36, No.3 pp 287-314, 1994.

[100] A. Hyvärinen, E. Oja *Independent Component Analysis: Algorithms and Application* Neural Networks, Vol. 13, No. 4-5, pp. 411-430, 2000.

[101] N. Metropolis *The beginning of the Monte Carlo Method* Los Alamos Science, Special Issue, 1987.

[102] S. Amari, H. Nagaoka *Methods of information geometry* Oxford University Press, 2000.

[103] R. E. Blahut *Principle and practice of information theory* Addison Wesley Series in Electrical and computer engineering, 1987.

[104] M. Fecko *Differential geometry and Lie groups for physicists* Cambridge University Press, 2006.

[105] W. Lederman, A. J. Weir *Introduction to Group Theory,* 2^{nd} *Ed.* Addison Wesley Longman, 1996.

[106] J. A. De Azcrraga, J.M. Izquierdo *Lie groups, Lie Algebras cohomology and some application in physics* Cambridge University Press, 1995.

[107] B. C. Hall *Lie groups, Lie Algebras and representations. An Elementary introduction* Springer-Verlag Inc., 2004.

[108] S.J. Julier, J.K. Uhlmann *Unscented filtering and nonlinear estimation*, Proc. of the IEEE, March 2004.

[109] M. Abramowitz, I. A. Stegun *Handbook of mathematical Functions* Dover Publication INC., New York 1964.

[110] G. Fasano, A. Franceschini *A multidimensional version of the Kolmogorov Smirnov test* Monthly Notices of the Royal Astronomical Society, Vol. 225 1987.

[111] C.J.C. Burge *A Tutorial on Support Vector Machines for Pattern Recognition* Data Mining and Knowledge Discovery 1998.

[112] T. Fawcett *An introduction to ROC analysis* Pattern Recognition Letters, Vol. 27, Issue 8, pp 861-874, 2006.

[113] H. A. David, H. N. Nagaraja *Order Statistics 3rd Edition* John Wiley & Sons, 2003.

[114] A. Hein *Processing of SAR data* Springer-Verlag Inc., 2004.

[115] A. Bolvik, J. Gibson *Handbook of Image and Video Processing (Communications, Networking and Multimedia)* Academic Press, 2000.

[116] T. Crimmins *The geometric filter for speckle reduction* Applied Optics, Vol. 24, No. 10, May 1985.

[117] N. J. Redding *Estimating the Parameters of the K Distribution in the Intensity Domain* DSTO-TR-0839, July 1999.

[118] R. Kimmel, R. Malladi, N. Sochen *Image processing via Beltrami operator* presented at the 3rd Asian Conf. Computer Vision, Hong Kong, January 1998.

[119] N. Sochen, R. Kimmel, R. Malladi *A general framework for low level vision* IEEE Trans. Image Processing, Vol. 7, No. 3, pp 310-318, March 1998.

[120] N. Sochen, R. Kimmel, R. Malladi *From high energy physics to low level vision* LBNL Rep. 39243, Univ. California, Berkeley, Aug. 1996.

[121] A. M. Polyakov *Quantum geometry of bosonic strings* Physics Letters B, 103(B), pp 207-210, 1981.

[122] R.J. Dekker *Speckle filtering in satellite SAR change detection imagery* International Journal of Remote sensing, 19:6, pp 1133-1146.

[123] E. Kreyszing*Differential Geometry* Dover Publications, Inc., New York 1991.

[124] A. Spira, R. Kimmel*A short-time Beltrami kernel for smoothing images and manifolds* IEEE Trans. Image Processing, Vol. 16, No. 6, pp. 1628-1635, June 2007.

[125] N. Sochen, R. Kimmel, R. Malladi *On the Geometry of Texture* Proceedings of the 4th International conference on Mathematical Methods for Curves and Surfaces, 1996.

[126] G. Xu *Discrete Laplace-Beltrami operators and their convergence*, Computer Aided Geometric Design, Vol. 21 , Issue 8, pp. 767 - 784, October 2004.

[127] S. Osher *Image Processing Based on Partial Differential Equations* Proceedings of the International Conference on PDE-Based Image Processing and Related Inverse Problems, CMA, Oslo, August 812, 2005.

[128] W. K. Pratt *Digital image processing: PIKS inside, Third Edition* John Willey & Sons, 2001.

[129] G. Aubert, P. Kornprobst *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations Second Edition* Springer 2006.

[130] W. Kühnel *Differential Geometry: curves - surfaces - manifolds*, American Mathematical Society, 2002.

[131] J. A. Thorpe *Elementary topics in differential geometry* Springer-Verlag Inc., 1979.

[132] R.W. Sharpe *Differential Geometry* Springer-Verlag New York Inc., 1996.

[133] G.W. Corder, D.I. Foreman *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach* John Willey & Sons, 2009.

[134] L.P. Roy, R.V.R Kumar *Accurate K-distributed clutter model for scanning radar application* IET Radar Sonar Navigation, Vol. 4, Iss. 2, pp. 158-167, 2010.

[135] N. Levanon, M. Shor *Order statistics CFAR for Weibull background* IEE Proc., Vol. 137. Pt. F, No. 3, June 1990.

[136] V. Anastassopoulos, G. A. Lampropoulos *Optimal CFAR detection in Weibull clutter*, IEEE Trans. AES, Vol. 31, No. 1, January 1995.

[137] A. Papoulis *Probability, Random Variables, and Stochastic Processes* McGraw-Hill Kogakusha, Tokyo, 9th edition 1965.

[138] N.Y. Hassan, N. Aakamatsu *Contrast enhancement technique of dark blurred image* IJCSNS, Vol. 6, No. 2A, 2006.

[139] R. Duda, P. E. Hart *Use of the Hough transformation to detect lines and curves in pictures*, Comm. ACM, Vol. 15, No. 1, pp. 11-15, 1972.

[140] R. Gunn Steve *Support Vector Machines for Classification and Regression* Technical report, Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science, May 1998.

[141] E. E. Osuna, R. Freund, F. Girossi *Support Vector Machines: Training and Applications* Massachussetts Institute of Technology - Artificial Intelligence Laboratory, C.B.C.L Paper No. 144, March 1997

[142] C. Cortes, V. Vapnik *Support-Vector Networks* Journal of Machine Learning, Vol. 20, 1995.

[143] L. M. Novak, S.D. Halversen, G. Owirka, M. Hiett *Effects of polarization and resolution on SAR ATR* IEEE Trans. AES, Vol. 33, pp 102-116, 1997.

[144] R. Hummel *Model-based ATR using synthetic aperture radar*, IEEE Radar Conference, pp 856-886, 2000.

[145] J. Wissinger, R. Washburn, D. Morgan, C. Chong, N. Friedland, A. Nowicki, R. Fung *Search algorithms for model-based SAR ATR*, Algorithms for Synthetic Aperture Radar III, Proc. SPIE 2757, pp279-293, 1996.

[146] B. Ulug, S. C. Ahalt, and R. A. Mitchell *Efficient ATR using compression* IEEE Trans AES, Vol. 33, No. 4, 1997.

[147] M. Z. Brown *Analysis of multiple-view Bayesian classification for SAR ATR* Algorithms for Synthetic Aperture Radar X, Proc. SPIE 5095, pp 265-274, 2003.

[148] F. Barbaresco, N. Rivereau *Diffusive CFAR & its extension for Doppler and Polarimetric data* IET International Conference on Radar Systems, 2007.

[149] J. P. Egan *Signal Detection Theory And ROC Analysis* Academic Press, London 1975.

[150] T. Fawcett *ROC Graphs: Notes and Practical Considerations for Data Mining Researchers* Technical Report HPL-2003-4, HP LAboratories, Intelligent Enterprise Technologies Laboratory HP Laboratories, Palo Alto.

[151] J. C. Mossing, T. D. Ross *An evaluation of SAR ATR algorithm performance sensitivity to MSTAR extended operating conditions* Algorithms for Synthetic Aperture Radar V, Proc. SPIE 3370, pp 554-565, 1998.

[152] J. Illingworth, J.V. Kittler *A Survey of the Hough Transform* Computer vision, graphics and image processing, Vol. 44, pp 87-116, 1998.

[153] J. Illingworth, J.V. Kittler *The Adaptive Hough Transform*, IEEE Transaction on Pattern Analysis and Machine Intelligence, September 1987.

[154] F. C. Morabito, G. Simone, A. Farina *Automated lineament detection in SAR images based on the joint use of wavelet and hough transforms* Proc. of 5th International Conference on Radar Systems, Brest, France 1999.

[155] W. Haihui, W. Yanli, Z. Tongzhou *Automated Detection in SAR Images by Using Wavelet Filtering and Hough Transform* Second International Workshop on Education Technology and Computer Science, 2010.

[156] J. Cheng-li, J. Ke-feng, J. Yong-mei, K. Gang-yao *Road Extraction from High-Resolution SAR Imagery. Using Hough Transform* Geoscience and Remote Sensing Symposium, 2005.

[157] N. Kiryati, Y. Eldar a A. M. Bruckstein *A probabilistic Hough transform* Pattern Recognition, Volume 24 Issue 4, 1991.

[158] D. Shaked , O. Yaron , N. Kiryati *Deriving Stopping Rules for the Probabilistic Hough Transform by Sequential Analysis* Computer Vision and Image Understanding 1996.

[159] A.D. Lanterman, J. A. O'Sullivan, M.I. Miller *Kullback-Leiber distances for quantifying clutter and model* Optical Engineering, Vol. 38, No. 2, pp 2671-2674, December 1999.

[160] T.M. Cover, J. A. Thomas. *Elements of Information Theory* John Wiley & Sons New York, 1991.

[161] F. Sadjadi *Improved Terget Classification Using Optimum Polarimetric SAR Signature*, IEEE Trans. AES, Vol. 39, No. 1, January 2002.

[162] K. Arwini, C.T.J. Dodson *Information Geometry: Near Randomness and Near Independence (Lecture Notes in Mathematics)* Springler-Verlang Inc., 2008.

[163] M. B. Sechtin, L. M. Novak, M.C. Burl *Algorithms for optimal processing of polarimetric radar data* Project report TT-73, MIT-Lincoln Lab., 1989.

[164] A. W. Rihaczek, S. J. Hershkowitz *Radar Resolution and Complex-Image Analysis* Norwood, MA, Artech House, 1996.

[165] B. Borden *Radar Imaging of airborne targets* Philadelphia: Institute of Physics, 1999.

[166] I. Antipov *Analysis of sea clutter data* DSTO Electronic and Surveillance Research Laboratory, DSTO-TR-0647, March 1998

[167] H.C. Chan *Radar sea clutter at low grazing angles* IEEE Proc., Vol.137, Pt. F, No.2, April 1990.

[168] E. Jakeman, P.N. Pusey *A model for non-Rayleigh sea echo*, IEEE Trans. On Antennas and Propagation, Vol. 24, No. 6, November 1970.

[169] T.P. Leonard, I. Antipov, K.D. Ward *Comparison of radar sea clutter models*, IEE, No.490, Radar 2002, Edinburgh.

[170] C.J. Oliver *Optimum texture estimators for SAR clutter* Journal of Physics: Applied Physics, Vol. 26, pp 1824-1835 , 1993.

[171] G. Marino, E. J. Hughes *Information Content Variation In CFAR Processing*, Cranfield University Research Student Symposium, May 2011.

[172] G. Marino, E. J. Hughes *Automatic Target Recognition in Synthetic Aperture Radars*, 9th Electro-Optics & Infrared Conference, Shrivenham, 4th/6th July 2011.

[173] G. Marino, E. J. Hughes *A novel mathematical approach for the problem of CFAR clutter model approximation*, 3rd Microwaves, Radar and Remote Sensing Symposium, Kiev August 2011.