# Advanced Maritime Target Recognition from SAR Images Exploiting Target's Micro-Motions and AI

---

**Laura Parra García**

A thesis submitted for the degree of
Doctor of Philosophy

Sensor Signal Processing & Security Lab
Department of Electronic and Electrical Engineering
University of Strathclyde
Glasgow

2026

# Abstract

The work presented in this thesis is an exploration of advanced techniques for maritime target recognition using Synthetic Aperture Radar (SAR) data, focusing on innovative approaches to target classification, coregistration, and micro-motion analysis, in the field of maritime surveillance. SAR technology offers unique advantages in maritime environments, including the ability to operate under diverse weather and lighting conditions. This research aims to exploit these capabilities to address the challenges of detecting and classifying maritime targets, particularly in scenarios involving moving or vibrating ships due to their engine.

The literature review establishes the foundation by discussing SAR geometry, image formation processes, and acquisition modes, providing a comprehensive understanding of the principles that underpin SAR data and imaging. It also highlights state-of-the-art advancements in coregistration of multitemporal SAR images and micro-motion analysis, offering insights into their applications in maritime environments.

A novel cross-cross-correlation-based method is proposed for joint coregistration of rotated multitemporal SAR images, addressing challenges related to alignment and displacement field estimation. The effectiveness of the method is demonstrated through experimental evaluations, highlighting its potential for improving SAR image processing workflows for maritime surveillance.

This research also introduces a novel framework for maritime target classification using Single Look Complex (SLC) SAR data. Key contributions include the extraction of spectral profile features and the integration of invariant features, such as Krawtchouk moments, which enhance the classification performance of the overall system. A combination of machine learning and deep learning techniques, including Recurrent Neural Networks (RNNs) and confidence-based ensemble methods, is employed to improve accuracy and resilience against noise.

Furthermore, a second framework for detection, location, and classification of targets based on autocorrelation is proposed. By exploiting sub-aperture processing and physics-informed feature extraction, the framework enables robust identification of vibrating targets and discrimination from static objects and clutter. The results demonstrate that micro-motion signatures embedded in the autocorrelation domain can be effectively characterised and leveraged for target classification, improving sensitivity to subtle object dynamics.

The findings presented in this thesis introduce novel methodologies for target recognition and categorisation in SAR-based maritime surveillance. By incorporating micro-motion–aware features, the proposed approaches enhance the capability of SAR systems to detect, characterise, and interpret maritime activities. These advancements have direct relevance to military surveillance, environmental monitoring, and the detection of illegal or anomalous maritime behaviour.

# Contents

# List of Symbols

| | |
|---|---|
| $(\cdot)^*$ | Complex conjugate operator |
| $\alpha_k$ | Complex scaling/rotation factor for image $k$ |
| $\boldsymbol{\epsilon}$ | Absolute residual error vector |
| $\boldsymbol{w}$ | Weighting vector |
| $\Delta\phi_{\text{max}}$ | Maximum phase deviation |
| $\Delta R$ | Spacing between consecutive dominant autocorrelation peaks |
| $\delta_k$ | Complex translation parameter for image $k$ |
| $\delta_r$ | Range resolution (m) |
| $\delta_{az}$ | Azimuth resolution (m) |
| $\delta_{nm}$ | Kronecker delta |
| $\ell$ | Discrete lag variable in autocorrelation |
| $\ell_{\text{dom}}[k]$ | Dominant autocorrelation lag in the $k$-th sub-aperture |
| $\ell_{\text{mode}}$ | Most frequent dominant lag across sub-apertures |
| $\eta$ | Slow-time variable (azimuth time) |
| $\gamma_k$ | Scaling (zoom) factor $k$ |
| $\Gamma_r$ | STFT Refinement score |

## List of Symbols

| | |
|---|---|
| $\hat{\phi}(\ell)$ | Least-squares linear approximation of phase |
| $\hat{f}_{\text{vib}}$ | Estimated dominant vibration frequency (Hz) |
| $\hat{y}$ | Predicted class label |
| $\lambda$ | Wavelength (m) |
| $\mathbf{I_{RD}}$ | Range-Doppler representation of SLC image |
| $\mathbf{S}$ | Focused SLC SAR image |
| $\mathbf{w}$ | Normal vector of the separating hyperplane |
| $\mathbf{x}_i$ | Feature vector of the $i$-th sample |
| $\mathcal{F}^{-1}$ | Inverse Fast Fourier Transform (IFFT) operator |
| $\mathcal{R}_{\text{STFT}}$ | Refined candidate set from STFT detection |
| $\mathcal{T}\{\cdot\}$ | Least-squares linear detrending operator |
| $\mu_X$ | Mean value of feature vector $X(y)$ |
| $\mu_{\Delta R}$ | Mean spacing between dominant autocorrelation peaks |
| $\omega_c$ | Carrier angular frequency (rad/s) |
| $\phi(\ell)$ | Unwrapped phase of the autocorrelation function |
| $\phi(t)$ | Instantaneous phase of the received signal |
| $\phi_r(k)$ | Unwrapped phase of the slow-time signal at range bin $r$ |
| $\rho(N, n, p)$ | Weight function of Krawtchouk polynomials |
| $\sigma^0$ | Sigma nought, normalised radar backscatter coefficient |
| $\sigma_w^2$ | Noise variance |
| $\sigma_X$ | Standard deviation of feature vector $X(y)$ |

| | |
|---|---|
| $\sigma_{\Delta R}$ | Standard deviation of peak spacing |
| $\sigma_{\text{static}}$ | Scattering coefficient of static target |
| $\sigma_{\text{vib}}$ | Scattering coefficient of vibrating target |
| $\tau$ | Fast-time variable (range time) |
| $\theta$ | Antenna aperture (m) |
| $\theta_k$ | Rotation angle |
| $\tilde{\phi}_r(k)$ | Unwrapped-detrended phase |
| $\tilde{f}(k)$ | Normalised feature vector |
| $\tilde{R}(\ell)$ | Normalised autocorrelation function at lag $\ell$ |
| $\tilde{X}(y)$ | Z-score normalised feature vector |
| $\tilde{X}_{log}(y)$ | Normalised logarithmic spectral feature |
| $\tilde{X}_{spec}(y)$ | Normalised magnitude spectral feature |
| $\zeta_l$ | Complex coordinate of the $l$-th tie-point in the secondary image |
| $A$ | Vibration amplitude (m) |
| $a_{\cdot,n}$ | Per-pulse amplitude fluctuation random variable |
| $A_{\text{speckle}}$ | Rayleigh-distributed speckle amplitude |
| $B$ | Transmitted signal bandwidth (Hz) |
| $b$ | Bias (offset) term of the hyperplane |
| $C_1, C_2, C_3$ | Class sets in the multiclass classification problem |
| $d_a$ | Antenna size azimuth direction (m) |
| $f(x, y)$ | Image intensity function at pixel $(x, y)$ |

List of Symbols

| | |
|---|---|
| $f_c$ | Radar carrier frequency (Hz) |
| $f_D$ | Doppler frequency shift (Hz) |
| $f_R$ | Received frequency at the sensor (Hz) |
| $f_s$ | Sampling frequency (Hz) |
| $f_{\text{vib}}$ | Vibration frequency (Hz) |
| $f_{D_n}(T)$ | Instantaneous Doppler frequency of target $n$ (Hz) |
| $f_{mD}$ | Micro-Doppler frequency (Hz) |
| $I_{RD}(r, k)$ | Range-Doppler representation of the SLC image |
| $I_{SLC}(x, y)$ | Focused Single Look Complex (SLC) image |
| $j(x)$ | Discrete weight function of the Krawtchouk polynomials |
| $K(\mathbf{x}_i, \mathbf{x}_j)$ | Kernel function |
| $k_n(x)$ | Krawtchouk polynomial of order $n$ |
| $K_r$ | Chirp rate of LFM waveform |
| $K_{nm}$ | Two-dimensional Krawtchouk moment of orders $n$ and $m$ |
| $L$ | Real antenna length (m) |
| $L_{sa}$ | Synthetic Aperture Length (m) |
| $mod_{idx}$ | Modulation index |
| $P$ | Phase Change parameter |
| $p$ | Krawtchouk polynomial parameter controlling spatial localisation |
| $P_r(f)$ | Time-collapsed frequency spectrum at range bin $r$ |
| $Q_n(T)$ | Position vector of target $n$ at slow time $T$ |

| | |
|---|---|
| $R$ | Slant range distance from the antenna to the target (m) |
| $R(\ell)$ | Autocorrelation function |
| $r(t, \eta_n)$ | Received raw echo signal |
| $R_0$ | Reference slant range to scene centre (m) |
| $S(y)$ | Azimuth spectral profile |
| $S_r(f, \tau)$ | STFT Energy Spectrogram |
| $S_{\text{az}}(k)$ | Collapsed Doppler profile |
| $SP_r(f)$ | STFT spectrogram integrated over time |
| $T_p$ | Pulse duration (s) |
| $v_p$ | Platform velocity along azimuth direction (m/s) |
| $v_{\text{drift}}$ | Linear drift velocity of static target (m/s) |
| $v_{r_n}(T)$ | Instantaneous radial velocity of target $n$ (m/s) |
| $w(t, \eta_n)$ | Complex white Gaussian noise |
| $x_n(T)$ | Azimuth coordinate of target $n$ (m) |
| $x_r(k)$ | Slow-time signal extracted at range bin $r$ |
| $X_{log}(y)$ | Logarithmic spectral feature |
| $X_{spec}(y)$ | Magnitude spectral feature |
| $Xcorr1$ | Single autocorrelation |
| $Xcorr2$ | Double autocorrelation |
| $y_i$ | Class label of the $i$-th training sample |
| $y_n(T)$ | Range coordinate of target $n$ (m) |

| | |
|---|---|
| $z_1(t)$ | Vertical displacement of vibrating target |
| $z_l$ | Complex coordinate of the $l$-th tie-point in the reference image |
| $f(k)$ | Generic feature vector |
| $c$ | Speed of light in a vacuum (m/s) |

# Acronyms

**Φ-OTDR** Phase-sensitive Optical Time-Domain Reflectometry.

**A-SVM** Adaptive Support Vector Machine.

**AI** Artificial Intelligence.

**AIS** Automatic Identification System.

**AR** Agreement Ratio.

**ATR** Automatic Target Recognition.

**BiLSTM** Bi-directional Long Short-Term Memory.

**CA-CFAR** Cell-Averaging CFAR.

**CFAR** Constant False Alarm Rate.

**CLS** Constrained Least Squares.

**CNN** Convolutional Neural Network.

**CPI** Coherent Processing Interval.

**CUT** Cell Under Test.

**DEM** Digital Elevation Map.

**DFRFT** Discrete Fractional Fourier Transform.

**DFT** Discrete Fourier Transform.

**DL** Deep Learning.

**ECOC** Error-Correcting Output Codes.

**FFT** Fast Fourier Transform.

**FT** Fourier Transform.

**GAN** Generative Adversarial Network.

**GPS** Global Positioning System.

**GRD** Ground Range Detected.

**HMM** Hidden Markov Model.

**HRR** Hankel Rank Reduction.

**IF** Instantaneous Frequency.

**IFFT** Inverse Fast Fourier Transform.

**InSAR** Interferometry SAR.

**JCLS** Joint CLS.

**KPCA** Kernel PCA.

**LDA** Linear Discriminant Analysis.

**LFM** Linear Frequency Modulated.

**LS** Least Squares.

**LSTM** Long Short-Term Memory.

**MAD** Median Absolute Deviation.

**mDSAR** micro-Doppler SAR.

**ML** Machine Learning.

**MLP** Multi-Layer Perceptron.

**MLR** Multinomial Logistic Regression.

**NGF** Naïve Geometric Features.

**NN** Neural Network.

**OTDR** Optical Time-Domain Reflectometry.

**OvO** One-vs-One.

**PCA** Principal Component Analysis.

**PolSAR** Polarimetric SAR.

**PRF** Pulse Repetition Frequency.

**RADAR** Radio Detection and Ranging.

**RBF** Radial Basis Function.

**RCM** Range Cell Migration.

**RD** Range-Doppler.

**RDA** Range Doppler Algorithm.

**RMSE** Root Mean Square Error.

**RNN** Recurrent Neural Network.

**SAR** Synthetic Aperture Radar.

**SCR** Signal-to-Clutter Ratio.

**SHM** Structural Health Monitoring.

**SIFT** Scale-Invariant Feature Transform.

**SLC** Single Look Complex.

**SNR** Signal-to-Noise Ratio.

**SoI** Signal of Interest.

**SPOT** Sub-Pixel Offset Tracking.

**SSP&S** Sensor Signal Processing & Security.

**STFT** Short-Time Fourier Transform.

**SURF** Speeded Up Robust Feature.

**SVM** Support Vector Machine.

**TDOA** Time Difference Of Arrival.

# List of Figures

# List of Tables

# Acknowledgements

First and foremost, I would like to extend my deepest gratitude to my sponsors, BAE Systems and the European Space Agency (ESA), for their generous support and for enabling this work to come to life.

I am incredibly grateful to my supervisors at the University of Strathclyde, Prof. Carmine Clemente, Dr. Christos Ilioudis, and Prof. Malcolm Macdonald. Your guidance, encouragement, and expertise have been invaluable throughout this journey.

A special mention goes to Dr. Gianluca Furano for believing in me, keeping my motivation levels high, and being my number one fan. Working with you at ESA was amazing.

To my parents, words cannot express the depth of my gratitude. Your constant support, encouragement, and sacrifices have been the foundation of all my achievements. This accomplishment would not have been possible without you, even if the topic remains somewhat mysterious.

To all my PQPses and especially to Martuki Fiesta, for providing me with a sunny place to work from, papas arrugás, and true crimes. Puri APS.

To Villa Rolliza, for keeping me motivated by asking (on repeat) when I will be done. For the late nights in Madrid, the trips, and the Horteralias.

To Sergio, for being an endless source of calmness, laughs, support, and salmon salad. Truly carried this thesis journey by doing everything except writing it.

Finally, to the friends I made along the way: Lara, Pete, il Parruco... and especially Greta, a ray of sunshine in the not-so-sunny-Glasgow, forza amiga! L'equipo sempre.

This thesis is a culmination of the contributions, love, and support from all of you.

Thank you for being part of this journey. The least you can do now is read it.

The project was funded by The University of Strathclyde, BAE Systems, and ESA. ESA Contract No. 4000130412/20/NL/MH/ac.

# Chapter 1

# Introduction

## 1.1 Context of the Research and Motivation

The motivation for the work presented in this study is the detection and identification of dark vessels. Ships that operate without sending information or identifying themselves have become a critical challenge in maritime surveillance, as they are often linked to illegal activities such as smuggling, piracy, and illegal fishing, posing significant threats to national security, environmental conservation, and global trade. Synthetic Aperture Radar (SAR) technologies have been fast developing over the last years as they offer important advantages over other types of sensors. The ability to monitor the Earth's surface in any weather, through clouds and smoke, and in any light condition, day or night, is a crucial feature of spaceborne and airborne SAR sensors. These characteristics make SAR suitable for a wide range of applications, such as agriculture monitoring, deforestation tracking, fire damage assessment, glaciers and permafrost hazards, and the monitoring of activities in the open ocean, among others. They can also be used for mapping other planets and moons that are not accessible by optical sensors due to thick atmospheres.

SAR functions by illuminating the targeted area or object with electromagnetic pulses and measuring the phase and magnitude of the scattered echoes sequentially to determine the position and characteristics of the targets under study. The amplitude and phase depend on the physical characteristics and electrical properties, which will be reflected in the final image. Through signal processing, a two-dimensional image of the mapped area can be formed.

In the context of maritime surveillance, there are challenges to address, such as the movement of the ships due to their trajectory and the rocking of the sea. The relative movement between the sensor and the target introduces artefacts in the resulting images, such as blurring and positional shifts of moving objects. These effects arise from Doppler frequency variations and phase errors caused by residual motion during signal acquisition and processing. Moreover, the distinct radar reflectivity characteristics of the sea surface and maritime vessels provide a reliable basis for detecting and classifying maritime targets. This capability makes SAR particularly suitable for identifying anomalies such as dark vessels, distinguishing them from other maritime objects, and supporting real-time decision-making. Leveraging these strengths, this project aims to advance SAR-based methodologies for enhanced detection and characterisation of dark vessels, contributing to safer and more secure maritime operations globally.

## 1.2 Contributions

The present thesis delivers several original contributions to the field of SAR image processing and maritime target recognition and classification. The key contributions of this research are outlined below:

1. In Chapter 3, a framework for the joint coregistration of rotated multitemporal SAR images is presented and substantially extended from earlier preliminary work [1]. The preliminary contribution introduced the use of cross-cross-correlation for joint estimation of misregistration parameters across multiple SAR images. Building on this foundation, the main contributions of this chapter are as follows. First, a rigorous mathematical formulation of the joint coregistration problem is developed, based on a Constrained Least Squares (CLS) framework, hereafter referred to as the Joint CLS (JCLS) method, with an explicit analytical solution. Second, the cross-cross-correlation approach is extended into a complete and reproducible algorithmic pipeline, including structured model construction

and displacement estimation. Third, a robust displacement estimation strategy is introduced, incorporating mechanisms for improved reliability under noise and complex alignment conditions. Fourth, an additional robustness enhancement based on outlier rejection is integrated into the estimation process. Finally, the proposed framework is extensively validated through a comprehensive experimental analysis, demonstrating improved accuracy and consistency compared to conventional pairwise coregistration approaches [2].

Accurate multitemporal coregistration is particularly critical in surveillance applications when fusing SAR acquisitions obtained from different orbital passes, viewing geometries, and squint angles, where rotational and translational misalignments are prevalent. Even small offsets can lead to erroneous vessel localisation, false change detection, or incorrect association of detections across time.

2. In Chapter 4, an innovative feature extraction for maritime target classification is proposed through the use of spectral profile features derived from Single Look Complex (SLC) SAR data for binary and multiclass classification of maritime targets. In addition, invariant features are also developed to enhance the classification performance and robustness against noise and varying acquisition conditions. A comprehensive evaluation across multiple classifiers and feature configurations demonstrates consistent performance gains, confirming both the effectiveness of the proposed spectral features and the added discriminative power provided by the invariant attributes in some scenarios [3].

3. Additionally, Chapter 4 presents a deep learning–based maritime Automatic Target Recognition (ATR) algorithm where a Recurrent Neural Network (RNN) is introduced for ship classification using SLC-derived spectral profiles, alongside a classifier fusion framework based on confidence weighting and logistic regression, achieving improved accuracy and robustness over individual classifiers [3].

4. Advanced techniques for vibrating target characterisation: introducing an autocorrelation-based framework for extracting and analysing vibration signatures of

maritime targets. A pipeline that includes Constant False Alarm Rate (CFAR) detection, time-frequency analysis, autocorrelation characterisation, and Machine Learning (ML) classification is presented to isolate and characterise vibrational components, to facilitate the detection and distinction of targets with unique vibrational profiles. This work is developed in Chapter 5 and it is to be presented and published at the 23rd European Radar Conference (EuRAD), in October 2026.

These contributions collectively advance the state-of-the-art in SAR-based maritime surveillance and provide new methodologies for image coregistration, and target detection, characterisation, and classification, with potential applications in defence, environmental monitoring, and illegal activity prevention.

## 1.3 Thesis Organisation

This thesis is organised into different chapters, each addressing key aspects of the research conducted. Chapter 1 introduces context and motivation, and outlines the main objectives and contributions of the thesis. Chapter 2 presents a comprehensive review of the relevant background and state-of-the-art literature, alongside an overview of SAR technology, including imaging geometry, image formation principles, and acquisition modes. Techniques for the coregistration of multitemporal SAR images are also reviewed. A Cross-Cross-Correlation-Based Method for Joint Coregistration of Rotated Multitemporal SAR Images is presented in Chapter 3. This chapter formulates the coregistration problem, details the proposed methodology for cross-cross-correlation matrix construction, and evaluates the approach through experimental results. The advantages and limitations of the method are discussed in the context of multitemporal SAR analysis. Chapter 4, focuses on maritime target classification using spectral profiles extracted from SLC SAR data. A classification framework based on these profiles is introduced, together with the use of invariant features, such as Krawtchouk moments, to improve robustness. Both binary and multiclass classification scenarios are examined, and the impact of invariant features on classification performance is analysed. Chapter 5 presents an end-to-end detection and classification framework based on autocorrelation analysis. In this chapter, a complete processing chain combining CFAR-based target detection, Short-Time Fourier Transform (STFT) analysis, autocorrelation-based feature extraction, and Support Vector Machine (SVM) classification is developed. The proposed approach exploits autocorrelation structure to characterise micro-motion effects, enabling the discrimination of vibrating targets from static objects and clutter. Experimental results demonstrate the effectiveness of the framework in identifying vibration signatures and improving target classification reliability. Finally, Chapter 6 summarises the key contributions of the research, including its significance and implications for maritime surveillance using SAR. Moreover, Recommendations for future work are presented, emphasising potential directions for improving SAR-based target recognition and classification methodologies.

# Chapter 2

# Background and Literature Review

SAR is an active remote sensing technology that creates high-resolution images of the Earth's surface using Radio Detection and Ranging (RADAR) principles. Radars emit short bursts of electromagnetic waves and, upon encountering an object, these pulses return to the radar sensor, which utilizes the characteristics of the reflected echo signal to accurately determine the object's position, distance, and speed [4]. SAR technology has a clear advantage over optical imaging systems since it is not affected by light or weather and can penetrate a variety of atmospheric conditions, including rain and clouds. SAR sensors can be mounted on either airborne or space-borne platforms, enabling the sensor to move along a specific area of interest, capturing sequential images. These individual captures are then combined to form a *synthetic aperture*, thereby enhancing the resolution of the final image by synthesising a much larger antenna. Consequently, SAR can distinguish closely spaced targets with high accuracy in both range and azimuth dimensions [5].

When the scene under study remains stationary, SAR products exhibit high fidelity. However, blurring effects in the images may occur when there is relative motion between the sensor and objects in the scene, such as moving ships and land vehicles. When a target is stationary, the frequency of the transmitted and received signals remains identical (ignoring noise and system effects). As a result, the received signal exhibits no Doppler frequency shift, and its phase evolves linearly in time according to the nominal radar frequency. However, for moving targets, the returned and transmitted signals' frequency differ due to the Doppler effect [6]. The Doppler frequency shift $f_D$ between the received frequency $f_R$ and the radar carrier frequency $f_c$ is given by:

$$f_D = f_R - f_c \approx \frac{-2v}{\lambda} \tag{2.1}$$

where $v$ is the target's radial velocity, $\lambda$ denotes the transmitted wavelength, and the sign of $f_D$ depends on the target's direction relative to the radar: positive if moving toward the radar and negative if moving away.

Furthermore, if a target exhibits small-scale, repetitive motions, such as vibrations, these induce a micro-Doppler effect. This phenomenon, which varies over time, is described by the micro-Doppler frequency $f_{mD}$ :

$$f_{mD}[n] = \frac{4\pi f_c f_{vib} A}{c} \cos\left(2\pi f_{vib} n\right) \tag{2.2}$$

where $f_{vib}$ represents the vibration frequency, A is the amplitude of the displacement, and $c$ is the speed of light [7]. Analysing the target's micro-motions requires examining the signal in both time and frequency domain. The resultant time-frequency representation, known as the micro-Doppler signature, is useful for differentiating between various moving or vibrating objects, such as distinguishing ships in motion from static maritime features [8].

These foundational principles of SAR and the Doppler effect establish a framework for understanding SAR's capabilities in high-resolution imaging, particularly for dynamic environments such as maritime monitoring, where accurate target identification is critical.

## 2.1 SAR Geometry and Collection Modes

SAR geometry and collection modes are critical to understanding how SAR images are acquired, processed, and interpreted. SAR geometry encompasses the spatial orientation and positioning of the SAR sensor relative to the Earth's surface, while collection modes define the different operational configurations of SAR systems based on application requirements, platform constraints, and desired resolution. Fig. 2.1 illustrates the SAR imaging geometry, where the axis aligned with the sensor's flight path is referred to as the azimuth direction, while the range direction extends perpendicular to it along the ground. SAR emits electromagnetic pulses across an area, known as the "footprint", and records the echoes reflected from the surface as it moves along its path for a period of time, known as the acquisition time. The received echoed data is collected as a function of fast time and slow time, corresponding to the range and azimuth directions, respectively. As the sensor moves, the image matrix is formed. The focusing process of the SAR image is computationally intensive, therefore, the processing is not usually done on-board, the data is saved to memory and then downloaded to a ground station when the satellite comes into visibility.

The SAR imaging geometry, which is side-looking, is further characterised by slant range and ground range. Slant range is the direct line-of-sight distance from the radar to a target point, while ground range represents the horizontal distance from the nadir point on the Earth's surface to the target point. The SAR signal data is collected in slant range, but images are often projected to ground range for easier interpretation and analysis in many applications, such as land monitoring and topographic mapping [9]. The radar transmits pulses, generally Linear Frequency Modulated (LFM) signals or *chirps*, that propagate towards the surface with a cadence called the Pulse Repetition Frequency (PRF), and the swath width depends on the real dimensions of the antenna. Depending on the duration of the pulse, it allows for a certain resolution in range. The movement of the platform enables the creation of a synthesised, extended antenna, effectively spanning kilometres, as if a single element was responsible for the larger aperture. The antenna aperture, or beam-width, $\theta$ depends on the dimension of the

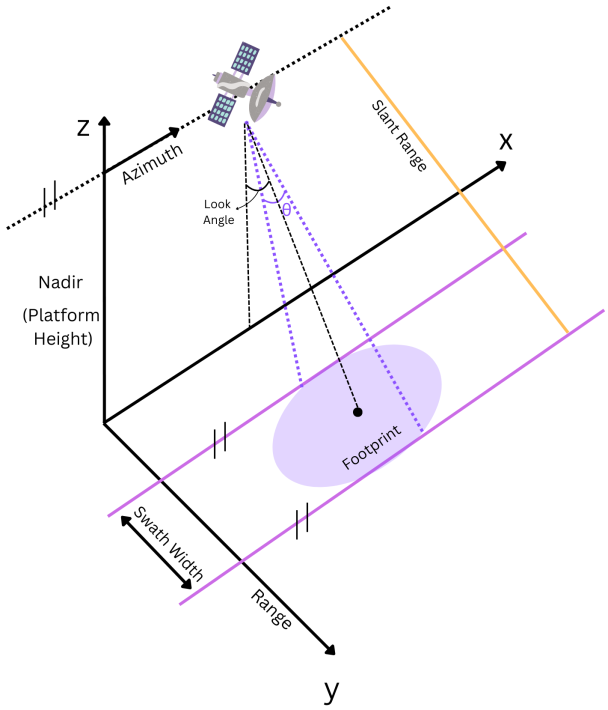


Figure 2.1: SAR imaging geometry, illustrating the azimuth and range directions.

real antenna in elevation/azimuth $L$ and the wavelength $\lambda$:

$$\theta \approx \frac{\lambda}{L} \tag{2.3}$$

The footprint depends on the beam-width and the distance from the antenna to a target point $R$,

$$Footprint = \theta \times R \tag{2.4}$$

as illustrated in Fig. 2.2, along with the distance to the centre of the footprint, the movement of the platform, and the beam-width. The length of the synthetic aperture produced by the platform's movement, which greatly increases the effective antenna length beyond its physical dimensions, is a critical factor in determining the resolution obtained in SAR imaging. Since the synthetic aperture allows for better spatial distinction between objects in the azimuthal plane without requiring a physically larger antenna, this idea is essential to SAR's capacity to produce fine azimuth resolution [10].

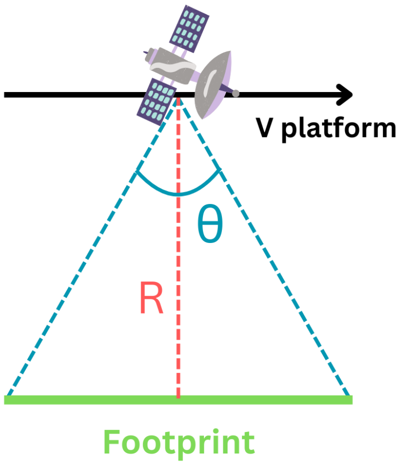


Figure 2.2: SAR footprint geometry

SAR images, in addition, depend on the frequency of the radar and the look angle (see Fig. 2.1. These two factors affect the power and the penetration of the wave, as well as the resolution. Objects with dimensions comparable to the wavelength used can be captured. Related to the quality of the SAR image, resolution is a fundamental design parameter, which depends on the requirements of the mission. The concept of resolution is a two-dimension one; the along-track dimension, or azimuth dimension, and the perpendicular to it, the cross-track direction, or range. The range resolution depends on the inverse of the bandwidth of the transmitted pulse $B$ or, equivalently, the duration of the pulse (T), as follows:

$$\delta_r = c/2B = c\tau/2 \tag{2.5}$$

Two targets can be distinguished if their relative distance is greater than the resolution value. The azimuth resolution, on the other hand, depends on the physical azimuth dimension of the antenna $d_a$, $\lambda$, and the distance $R_0$

$$\delta_{az} = \theta/R_0 = \frac{\lambda}{d_a}R_0 \tag{2.6}$$

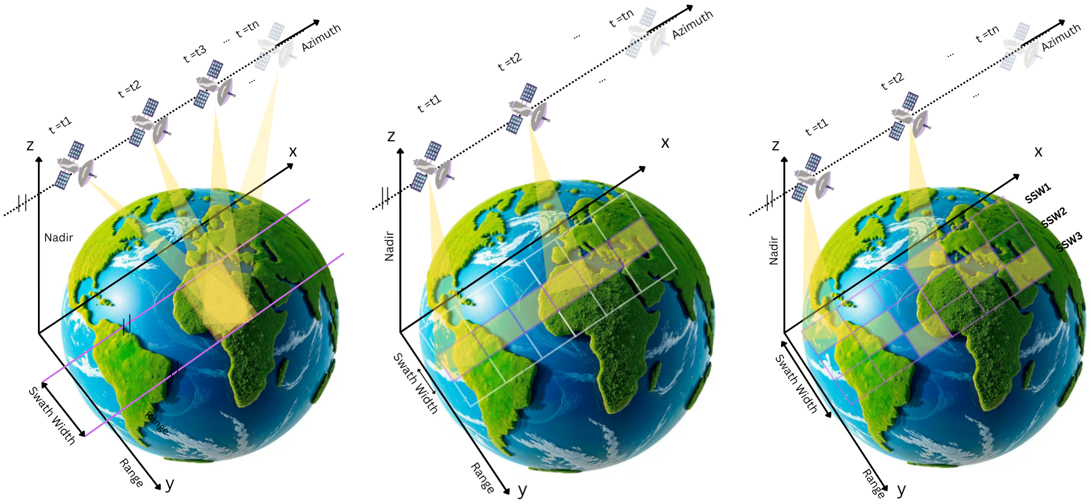
and two targets can be distinguished in azimuth if their angular distance is greater than the resolution in this dimension. Taking advantage of the synthetic aperture, which is equal to the along-track size of the instantaneous swath $L_{sa} = \frac{\lambda}{d_a} R_0$, the azimuth resolution detaches from the distance, allowing the system to maintain fine resolution in this direction as well and becomes:

$$\delta_{az} = \frac{\lambda}{L_{sa}} R_0 = \frac{\lambda}{2\frac{\lambda}{d_a R_0}} R_0 = d_a/2 \tag{2.7}$$

The factor 2 accounts for the round-trip travel time for the signal, so the relative phase shift is observed twice.

### 2.1.1 SAR Acquisition Modes

SAR systems operate in several distinct collection modes, each tailored to achieve specific imaging objectives, including variations in resolution, swath width, and scene coverage. The basic SAR collection modes are Stripmap, Spotlight, and ScanSAR, and they are shown in Fig. 2.3.



(a) Spotligth (b) Stripmap (c) ScanSAR

Figure 2.3: SAR acquisition modes [11].

- Spotlight Mode: The SAR system continuously adjusts the antenna beam to focus on a more specific, targeted area as the platform moves, prolonging its dwell time over that area. This method enhances azimuth resolution inside the smaller imaged region. Spotlight mode works best for applications that need high-resolution imagery, such as surveillance and reconnaissance, where precise detail in a limited area is more crucial than broad coverage [12]. A compromise between the spotlight and stripmap (see below) modes is the “Sliding” or “Sliding-Spotlight”, where the beam illuminates the target area for longer than stripmap, achieving better resolution, and keeping wider coverage than spotlight [13].

- Stripmap Mode: In this mode, the SAR antenna beam remains stationary at a specific angle to the path while continuously scanning a strip of ground parallel to the flight path. This mode’s continuous azimuth and range resolution over the imaged strip is useful for applications such as geological and environmental monitoring that need wide-area coverage [14].
- ScanSAR Mode: To achieve wider swath coverage, the radar beam alternates between adjacent swaths as the platform advances. With a trade-off in azimuth resolution, ScanSAR covers a far larger region than Stripmap mode by periodically changing the beam direction. When fine resolution is less important than having a wide coverage, in applications like sea ice monitoring, natural disaster assessment, and environmental monitoring, this mode is a fair choice [15].

The acquisition mode is selected based on the application and the ratio between resolution and coverage needed. With the right mode, SAR can fit into many different types of remote sensing scenarios, from high-detail target detection with Spotlight mode to high-density observation with ScanSAR mode. The versatility and operational effectiveness of SAR-based Earth observation have been improved by developments in multi-mode SAR systems, which now enable dynamic mode switching within a single mission [16].

Variations of these modes that improve the illuminated area while preserving a good resolution, are being used by companies such as Capella Space, ICEYE, and Umbra. For example, Capella Space offers the Sliding Spotlight mode, which lengthens the duration of high-resolution spotlight products by gradually adjusting the antenna’s orientation. Instead of keeping the radar beam fixed on a single spot, the focus point of the radar shifts throughout time along the radar's trajectory. This approach provides excellent resolution over a wider area than the standard spotlight mode [17]. Umbra provides a mode called "Mosaic" that allows high-quality data collection across several nearby places to produce a seamless, large-scale image with remarkable spatial detail. For a thorough perspective of broader areas, this method allows for enormous coverage while preserving detailed resolution [18]. Because of the modest size and mass of its satellites, ICEYE is able to use a long spotlight SAR technology in its Dwell mode. This mode achieves extremely fine azimuth resolution by focusing on the same target area for up to 25 seconds. While conventional image formation methods can generate highly detailed images, these can be complex for human analysts to interpret. However, the extended dwell time captures significantly more data than standard spotlight modes, enhancing analytical potential and supporting advanced processing methods [19].

The flexibility of SAR in remote sensing depends on its geometry and collection modes, which allow for a wide range of applications, from large-scale environmental evaluation to high-resolution target identification. These setups highlight SAR's importance as a crucial technology in Earth observation by allowing it to be tailored to certain mission objectives.

## 2.2 SAR image formation

After collection, the data is arranged and saved for later examination. A raw data matrix is created by organising the surface returns, and the information gathered is saved in the columns of the matrix each time a pulse is sent and the matching echo is returned. The azimuth-directional data is recorded in the rows of the matrix as the satellite moves along its trajectory, repeating this process [20], and the data is efficiently organised for further processing and image production. The main goal is to build a reflectivity map of the scene by focusing the received echoes, compensating for phase variations due to the motion of the platform, and aligning the signal in both azimuth and range directions. SAR systems generate fine-detailed images of the observed area thanks to the combination of high azimuth resolution, which is attained using the synthetic aperture, and high range resolution, accomplished using pulse compression. However, issues like azimuth ambiguity and range migration must be resolved. Range Cell Migration (RCM) is a phenomenon caused by the relative movement of the radar platform and the scene being imaged. When a target's echoed signal is captured on the data matrix, it moves across range cells as a result of variations in the radar platform's distance from the target. As seen in Fig. 2.4, this causes misalignments in the raw data matrix, which need to be fixed throughout the image generation process. The radar beam's curvature and the quadratic phase changes brought forth by changing target distances as the platform moves also affect RCM. Range migration can skew the representation of the scene and reduce the final image's spatial resolution if left uncompensated [21].

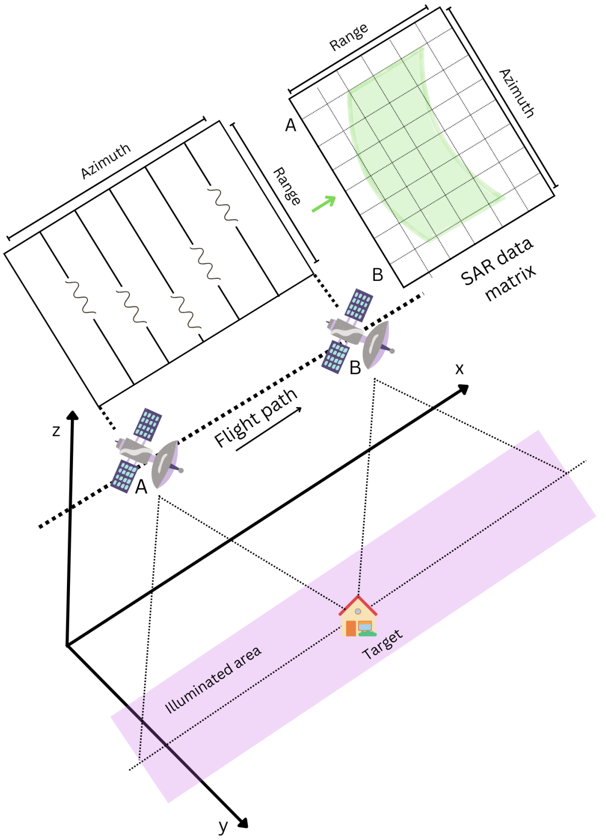


Figure 2.4: SAR data matrix creation.

One of the most commonly used imaging algorithms is the Range Doppler Algorithm (RDA), which consists of 3 main steps: Range compression, RCM correction, and azimuth compression. To focus the raw data in range, a Fast Fourier Transform (FFT) is performed in this direction, followed by a matched filter multiplication. Then, a range Inverse Fast Fourier Transform (IFFT) is performed to return to the fast time domain. Before RCM is implemented, the data is transformed into the range-Doppler domain, through an FFT in the azimuth direction, and after the data has been compensated in range, azimuth matched filtering is performed. The last step is to transform the data back to the time domain, by means of an IFFT in the azimuth direction, resulting in the final focused complex image to be analysed [22]. The process is illustrated in Fig. 2.5, where an example of a point target is displayed along with the imaging steps.

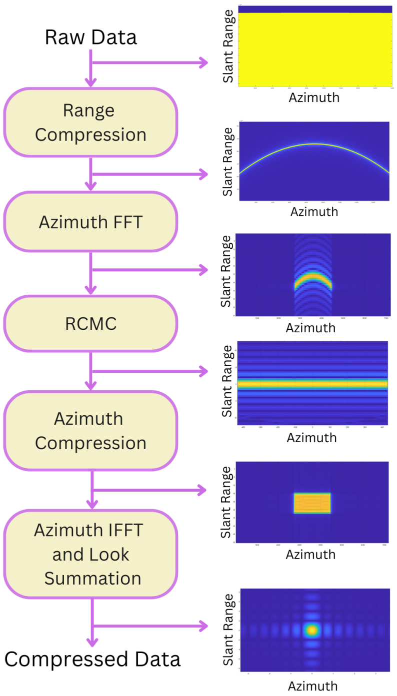


Figure 2.5: Functional block diagram of the RDA algorithm.

## 2.3 Coregistration of multitemporal SAR images

For applications like Interferometry SAR (InSAR), coherence analysis, displacement mapping, and coastal change detection, that demand high-precision alignment, coregistration of multitemporal SAR images is an essential tool [23, 24]. Variations in the trajectories followed by the platforms cause disparities between SAR images obtained over time or by numerous sensors. Even minor discrepancies can significantly impact the alignment of the final scene and the quality and accuracy of SAR-related products [25]. In maritime surveillance, accurate multitemporal coregistration can make it possible to track ship wakes, vessel motion patterns, and surface changes like oil spills over time, where even subpixel misalignments can result in false detections or misinterpretation of vessel displacements [26–28]. Coregistration techniques are typically applied after the image formation procedure to address these issues and ensure that matching pixels link to the same scatterers in the real scenario, typically done by aligning secondary images with a reference image. There are two primary techniques for SAR image registration in the literature [29]:

- The feature-based approaches use feature extraction and matching algorithms to find tie-points in both images to estimate the relative shift between them. The two most popular feature-based matching operators are the Scale-Invariant Feature Transform (SIFT) and the Speeded Up Robust Feature (SURF). In order to correlate images under a wide range of transformations, including translation, scaling, rotation, reflection, and shearing, as well as changes in 3-D viewpoint, noise levels, and illumination conditions, SIFT depends on the extraction of distinct invariant features from the images [30–32]. In maritime scenarios, feature-based coregistration relies on stable scatterers such as coastal infrastructures that provide persistent keypoints over time [33]. Through the use of the SURF technique, important points are identified in the image using the Hessian matrix, which makes it faster than the SIFT algorithm. SURF is efficient for real-time applications and is commonly used for tasks like image registration and object recognition due to its speed and reliability in matching features across images with varying conditions [34]. SIFT-based methods adapt the SIFT algorithm, accounting, for example, for the speckle noise present in SAR images, to make it more suitable for the unique characteristics of SAR data [31]. In [35], a combination of image segmentation with SIFT for feature extraction is presented. Segmentation improves the selection of meaningful regions, reducing false matches and enhancing accuracy in SAR coregistration. More recent work continues to adapt and extend these ideas to the SAR domain, for example by combining nonlinear diffusion filtering, Hessian-based salient structures and edge points to enhance keypoint stability in speckle [36], or by using deep networks to generate pseudo-optical images that facilitate more reliable feature detection and matching in SAR-optical registration [37]. Some studies propose multi-feature extraction algorithms, to improve the robustness and accuracy of the registration. A block-based technique in [32] illustrates two types of features, such as edges and corners and texture points, to increase matching speed. This method ensures reliable registration even in challenging scenarios by dividing the images into blocks and controlling the amount and uniform distribution of derived features. Speckle and geometric aberrations often cause outliers in feature matching in SAR images. Outlier removal methods like segmentation [38] and filtering [39] might help reduce this.

- Area-based methods focus on aligning images by comparing pixel intensities over large sections of the image. Usually, similarity measures like cross-correlation, mutual information, or sum of squared differences are used. Area-based techniques are well-suited for images where features are difficult to identify or where the images are less distinct because of noise or changes in imaging conditions [40–42]. Cross-correlation-based methods measure the similarity between two image patches by sliding one image over the other and they can be applied in both 2-D and 3-D data. This technique is used to estimate the shift between a reference and a secondary image. In multi-image corregistration, often, one image is chosen as the reference, and the remaining images are individually registered as secondary to the reference without considering their relative displacements. A joint coregistration method that simultaneously estimates parameters for every secondary in relation to the reference and takes into consideration displacements among the secondary themselves is provided in [1]. By adding cross-correlation data from several patches throughout the images to the CLS algorithm, this technique improves the robustness and accuracy of the computed parameters. The sub-pixel coregistration has been proposed in the literature, as a method that applies interpolation of the 2-D cross-correlation between reference and secondary images to refine the estimation of the peak value in the cross-correlation [43]. For maritime monitoring, intensity-based correlation can be affected by variations in sea state and wind conditions, modifying the sea clutter characteristics between acquisitions reducing the reliability of simple similarity measures, like cross-correlation.

In recent studies on sea state retrieval and maritime situational awareness, the dependence of SAR backscatter on wave and wind parameters is modelled [44]. Furthermore, Fourier domain correlation methods, like phase correlation, form a subset of area-based registration techniques. These techniques are ideal for remote sensing imagery because they are very resilient to noise that affects narrow frequency ranges as well as changes in light and reflectance. Due to the availability of FFT processing, these methods provide faster results than conventional cross-correlation techniques, which is especially useful for large remote sensing datasets. Additionally, phase correlation has been applied to detect translation, rotation, and scale transformations in optical remote sensing imagery, and to recover offsets in repeat-pass SAR imagery, and it can also address rotation and scale in SAR data.

New advancements demonstrate phase correlation for single-frequency repeat-pass imagery with temporal changes, as well as its extension to multi-frequency registration of X-band and C-band imagery [45]. To further increase robustness under large rotations and varying viewing geometries, correlation-based schemes have also been extended using more discriminative similarity measures, such as best-buddy similarity combined with SAR-specific operators for speckle suppression, producing improved matching performance for multi-view SAR data [46].

Feature-based approaches are recommended in situations involving significant geometric changes or when high precision is needed, whereas area-based approaches are effective for smaller displacements or datasets that are well aligned. These methods are frequently combined in recent research to improve accuracy and resilience in a variety of SAR applications [47].

In recent years, data-driven and learning-based strategies have been increasingly explored for SAR and SAR-optical image registration, complementing classical feature and area-based approaches, and improving robustness under complex radiometric and geometric conditions [48,49]. These works highlight the potential of deep networks and advanced similarity measures to handle strong heterogeneity, speckle and multi-sensor configurations. Recent approaches increasingly incorporate learning-based similarity metrics and region-adaptive keypoint selection to tackle SAR-optical heterogeneity and weak-feature regions, further improving registration robustness in challenging scenarios [50, 51]. There is an increasing trend towards hybrid pipelines that integrate physical models, classical optimisation and deep representations.

### Challenges in Pair-wise Multitemporal SAR Coregistration

Several challenges arise in the coregistration of multitemporal SAR images:

- Low Signal-to-Noise Ratio (SNR): Due to the coherent nature of radar backscatter, speckle noise, which is multiplicative, affects SAR images. It is challenging to extract and match related features across numerous images because of this noise, that drastically lowers the SNR. Weak scatterers are especially impacted by the low SNR, which might result in unclear or inaccurate tie-point selection during coregistration. Speckle noise reduces precision by making it more difficult to estimate the cross-correlation peaks that are utilised to calculate image displacements. Reliable feature matching is still difficult to achieve in low-SNR conditions, even with the use of speckle filtering techniques like the Lee or Frost filters [52, 53].

- Temporal and environmental changes: the sequential changes to the scene of interest that produce differences between the images obtained during different acquisitions (e.g., due to vegetation growth, urban development, or seasonal variations) can cause decorrelation of the images, lowering the coherence necessary for good coregistration [54]. Due to the decorrelation between images, standard feature matching, feature-based or area-based would no longer work or yield good results. Temporal decorrelation is a major issue in repeat-pass interferometry and in cases where phase coherence is essential for accurate displacement measure results [55, 56].

- Geometric distortions: generally, SAR images affected by geometric distortion resulting from the side-looking radar configuration. Foreshortening takes place when slopes facing the radar image appear compressed leading to misrepresentation of surface details in the image [57]. The inverse of foreshortening is layover and it is where linear features that are tall, such as buildings and mountains, are seen on the lower elevated surfaces, leading to overlapping features and bad geospatial resolution [57]. Shadowing is the case when objects that are behind tall structures or buildings do not receive any illumination from the radar beam hence resulting in gaps in the recorded data. Such distortions depend on the radar acquisition geometry and tend to increase the challenges of the alignment, especially in areas where the topography is complicated and three-dimensional in nature [57]. Such geometric distortions increase the difficulty level of the coregistration process as each image has its own features depending on their acquisition geometry [58]. A normal pixel-based matching approach may be sufficient to describe structures in flat/low areas but is unable to explain structures in rugged and urban areas. More elaborate approaches that exploit Digital Elevation Maps (DEMs) or model corrections, which are based on model fitting, are often used to correct geometric distortions [59].

- Inter-secondary misregistration: coregistration has been described as pairwise by most methods. The problem with this approach is that it does not account for

> possible relative motions that can occur between the secondary images, more specifically in multi-temporal setups where high rotational and translational discrepancies are present [23]. Inter-secondary misregistration is a result of the process in which secondary images do not fall perfectly within a bounding envelope which can be expected to create misalignments on the entire stack of images and lower the quality of the alignment. The problem of not being able to compensate for inter-secondary shifts could be particularly severe in airborne SAR systems or datasets collected when the platform moves considerably between acquisitions.

These issues are important in maritime scenarios, where sea clutter, variable sea state, and platform manoeuvres result in strong temporal decorrelation and significant inter-secondary misalignment. Chapter 3 proposes a joint cross-cross-correlation-based framework that directly addresses these limitations. Recent learning-based registration frameworks partially mitigate these issues by modelling radiometric differences and geometric distortions together, but performance is degraded under strong temporal decorrelation and low coherence [48, 49].

## 2.4 Micro-motions detection and analysis in SAR imaging

When a target is not moving, the frequency remains unchanged; that is to say, the frequency of the transmitted and received signal is the same. However, if the target is moving, the frequency of the received signal is changed due to the Doppler effect [6], which can be measured using the Fourier Transform (FT) of the received signal and follows the equation

$$f_D = f_R - f_c \approx \frac{(-)2v}{\lambda} \tag{2.8}$$

where $f_R$ is the receiving frequency, $f_c$ is the carrier frequency of the radar, $v$ is the radial velocity, $\lambda$ is the transmitted wavelength and the sign depends on the relative direction of the movement, being positive in the direction towards the radar and negative when moving away from it.

Moreover, if the object undergoes micro motion dynamics (rotations or vibrations), the micro-Doppler effect will be observed. As it varies with time, the micro-Doppler frequency follows the equation

$$fm_D[n] = \frac{4\pi f_c f_{vib} A}{c} \cos\left(2\pi f_{vib} n\right) \tag{2.9}$$

where $f_c$ is the carrier frequency, $f_{vib}$ is the frequency of the vibration, A is the displacement (in m) and $c$ is the speed of light. For extracting the target's micro-motions characteristics, the signal needs to be analysed in both time, and frequency domains. The observed micro-Doppler time-frequency representation is known as the micro-Doppler signature, and it can be used to distinguish moving and vibrating targets, such as ships [8,60,61].

### Vibration and rotation in SAR

The vibration of targets results in small Doppler frequency modulations in a radar signal. As the radars used for the surveillance of the sea generally operate in the X-band [62], these oscillations are around tens of hertz with few millimetres of amplitude. Millimetre wave radars allows for the tracking of targets with high sensitivity and accuracy without being disturbed by the surroundings. Millimetre wave SAR is ideal for image vibration due to its wavelength being close to typical vibration amplitudes and it has become important for the development of SAR technology as they are small, light and offer high resolution. This is why millimetre-wave radar systems are well suited to detect such small movements [63]. A moving target in a focused SAR image may appear smeared, unfocused and displaced in azimuth direction. The displacement depends on the radial velocity of the target and can be defined as

$$f_{D_n}(T) = \frac{\omega_c v_{r_n}(T)}{\pi c} \tag{2.10}$$

where $\omega_c$ is the carrier frequency, $v_{r_n}(T)$ is the instantaneous radial target velocity and $c$ is the speed of light. The instantaneous radial speed of the target can be expressed as:

$$v_{r_n}(T) = \frac{dQ_n(T)}{dT} \cdot \frac{r_n(T)}{|r_n(T)|} \tag{2.11}$$

where $r_n(T)$ is the radial look vector and $Q_n(T)$ is the position of the target $n$ at slow time T, defined by

$$Q_n(T) = \frac{x_n(T)}{y_n(T)} \tag{2.12}$$

where, using 2-D sensor flight geometry, $x_n(T)$ refers to the target at the x-coordinate, pointing down in the flight direction (azimuth), and $y_n(T)$ stands for the target at the y-coordinate, pointing in the radar look direction (range). Mechanical vibrations and rotations of a target produce phase modulation in the received echo. These modulations create sidebands in the Doppler frequency spectrum, the micro-Doppler effect that has previously been introduced. The micro-Doppler effect caused by both vibration and rotation is bigger as the wavelength $\lambda_c$ gets smaller or the carrier frequency $\omega_c$ increases.

- Vibrations: Vibrations result in a superimposed phase modulation on the received SAR signals scattered from the radar target. This phase modulation causes the effect known as paired echoes, ghost targets around the vibrating object because of the Doppler shifts. These paired echoes are an infinite series. Furthermore, if the SAR is carried in an airplane, some additional micro-Doppler frequencies can be originated by the vibration of the airplane.

  An interesting approach for vibrating targets in SAR imagery is described in [64], where a method to detect fatal damage in bridges and avoid their collapse is proposed by using micro-motion estimation, based on the analysis of modal properties. Sub-Pixel Offset Tracking (SPOT) is used as the tracking technique to apply to a single stripmap image where the acquisition time lasts for a few seconds. More recently, spaceborne SAR micro-Doppler techniques have been demonstrated for vibration-based Structural Health Monitoring (SHM) of real bridges, including a detailed assessment of space-borne micro-Doppler SAR (mDSAR) for measuring line-of-sight vibration time histories. These studies show that micro-Doppler information extracted from high-resolution SAR data can be used to retrieve natural frequencies and modal parameters with errors of the order of millimetres per second in velocity and on the order of 0.01 Hz in frequency [65, 66].

Such mDSAR-based SHM concepts are directly relevant to maritime applications, where vessel-induced structural vibrations or vibrations of offshore infrastructure can be exploited. In addition, recent work has analysed coupled micro-Doppler signatures from closely located vibrating scatterers, which is of interest when interpreting complex targets or multiple structures within a single SAR resolution cell [67].

- Rotations: Different rotating motions result in different backscattering characteristics but similar Doppler signature properties, closely linked to vibration. Rotating targets suffer a change of reflectivity as they rotate. As real SAR data often includes noise and clutter, the micro-Doppler effect might not be noticed in time or frequency domain. Consequently, atomic decompositions like the Gabor Transform can be successfully used to observe the micro-Doppler effects, as they are localised in both domains. As the vibration frequency increases, the target is more recognisable in a SARR image, due to the paired echoes being farther apart from it [63].

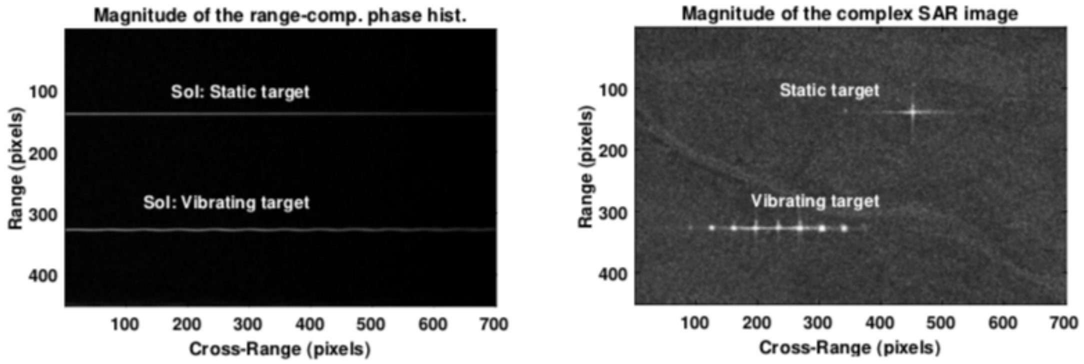
  To improve localisation of micro-motion within SAR images, recent work has proposed scale-invariant coherent change detection between sub-aperture images to locate micro-Doppler sources (e.g. vibrating scatterers) even when classical “ghost” artefacts are weak or masked by clutter [68]. This type of processing is particularly useful for identifying small vibrating components on otherwise bright stationary structures, such as parts of ships or bridge decks, directly from single-pass SAR imagery.

An approach to the detection and characteris ation of vibrating targets from SAR data is presented in [69]. The study proposes two different methods for detection and classification and it is based on three principles: first, the micro-Doppler response of a vibrating target suffers a non-linear transformation due to SAR sensing, which does not affects its information. Second, to classify vibrating objects, the Instantaneous Frequency (IF) of the SAR signal can be used. And third, a robust model can be developed to detect vibrating targets under various conditions, including scenarios with single or multiple objects in the presence of noise and clutter.

In order to prove these claims, a data-driven method and a model-based method are presented in the study. The data-driven method is based on applying the Discrete Fractional Fourier Transform (DFRFT) for the feature extraction from the IF of the slow-time SAR data, which previously has been obtained by time-frequency analysis of the SAR image. This approach is interesting for this study, as it could be adapted to work with SAR data before image formation for the on-board part of the framework. The step of obtaining the slow-time SAR data can be avoided if working with the SAR data directly and the vibration object can be identified from this data.

In Fig. 2.6, a static target and a vibrating target are shown in the form of the magnitude of the range-compressed phase history and the magnitude of the complex SAR image. Fig. 2.6 (a) shows the Signal of Interest (SoI) of the targets in a range-compressed phase history image. Fig. 2.6 (b) shows the reconstructed SAR image after applying azimuth compression to the phase history in 2.6 (a). The vibrating target produces side lobes, known as ghost targets [70]. The vibration-induced phase modulation is the micro-Doppler. Analysis tools other than the Discrete Fourier Transform (DFT) are required to estimate and characterise vibrations and non-stationary targets in general, and recent reviews summarise a broad spectrum of micro-Doppler analysis, extraction and mitigation techniques in radar and SAR imaging [60, 61].

The results of the study confirm that the approaches used can result in a robust classifier that can distinguish different types of machine vibrations and can be applied to differentiate types of vessels in the sea. When using the clutter suppression technique Hankel Rank Reduction (HRR), the performance rises and works for lower SNRs and lower Signal-to-Clutter Ratios (SCRs), working for over -10dB in both cases. The findings also confirm that SAR is suitable for characterising vibrations, as displacements remain distinguishable in the slow-time signals acquired by the sensor. Finally, comparing three pre-built classifiers (a linear classifier, a random forest classifier, and a fully connected neural network), the random forest classifier achieves the highest accuracy, especially when HRR is applied to the SAR data to suppress clutter noise before



(a) (b)

Figure 2.6: Reconstruction of a SAR image from its range-compressed phase history. (a) The magnitude of the range-compressed SAR phase history containing one static target and one vibrating target. The two targets are separated in range after range compression. (b) The reconstructed SAR image using the SAR phase history in (a) [69].

classification.

Beyond time-frequency analysis of the micro-Doppler signature, autocorrelation is widely used in vibration analysis to enhance periodic behaviour and estimate fundamental frequencies in noisy data. By correlating a signal with a delayed versions of itself, repeated patterns associated with oscillations are emphasised while uncorrelated noise is suppressed. Peaks in the autocorrelation at non-zero lags reveal the dominant vibration periods, which can then be mapped to the vibration frequency of the target under observation. This forms the basis for many condition-monitoring and fault-diagnosis techniques for rotating machinery and structural components, where autocorrelation is applied either directly to vibration time series, or to preprocessed representations (e.g. filtered, demodulated or time-frequency-projected signals) to extract periodic impulses or harmonics [71–73]. Similar correlation-based strategies are also employed in distributed and radar-based sensing, where the autocorrelation of demodulated phase or displacement signals is used to detect and localise periodic vibration sources and to discriminate them from noise [74–76].

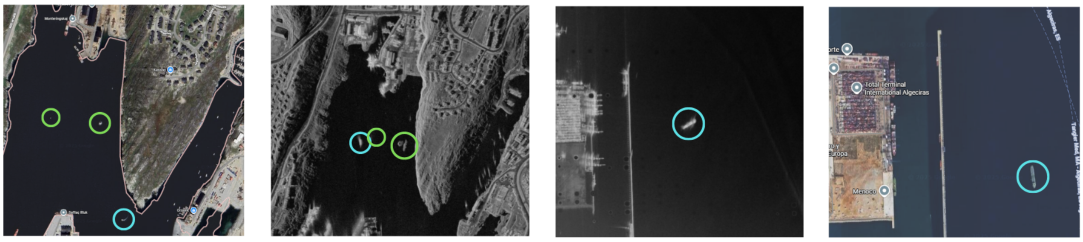
## 2.5 SAR Automatic Target Recognition for Maritime Applications

ATR from SAR imagery is crucial for maritime surveillance, allowing for the identification of vessels and other targets under varying environmental conditions. Maritime target surveillance can be performed effectively through the capability of SAR images to be acquired in any weather and light condition. The different SAR acquisition modes make it possible to image a wide area with a good trade-off in image resolution vs. coverage [15]. Moreover, the high contrast in SAR images between sea and targets, allows for reliable detection and segmentation of maritime objects [77]. Ships that surpass a certain length and width are required to share information about their position and characteristics at all times, through the Automatic Identification System (AIS), and when this system is deliberately off, they are likely to represent a threat and need to be identified as such. Different features can be used for the recognition and classification of ships from SAR data, depending on the information available and the requirements of the task.

Identifying maritime targets in SAR images typically involves two main stages: feature extraction, where discriminative characteristics are identified, and classification, where detected targets are assigned to categories based on their features. Traditional methods use feature-based techniques like edge detection and template matching models [78], and statistical methods for classification, like Bayesian classifiers [79] and Hidden Markov Models (HMMs) [80]. These classification methods perform well when big databases of labelled data are available, at the expense of significant power consumption and low-dimensional feature spaces. Early implementations of Multi-Layer Perceptrons (MLPs) and Radial Basis Functions (RBFs) have also been used for the classification of SAR images. These networks required careful feature extraction and struggled to generalise when the operating conditions were changed [81]. Feature-based ATR exploited SVMs, decision trees, and Linear Discriminant Analysiss (LDAs). Decision trees and SVMs provided interpretability and simplicity, but they were insufficient to handle complicated or overlapping target classes. SVMs needed careful kernel selection and parameter optimisation, however, they performed well for small datasets. LDA could not be used for non-linear scenarios, although it improved class separability by reducing feature dimensionality. [82].

Furthermore, ship-iceberg detection and classification has also been studied in different works using images from different satellites, like TerraSAR-X [83] and Sentinel-1 [84], mainly using focused SAR images for the distinction of the two groups. In [85], high-resolution Ground Range Detected (GRD) SAR images acquired with Sentinel-1 are used, involving land-masking processing and a detection threshold set for the classification. This study bases its classification on the backscatter values of the different objects obtained with different polarisations and the correlation of AIS data, achieving good detection rates for bigger vessels.

To visually illustrate the comparison between possible icebergs and potential vessels, Fig. 2.7 shows two SAR images from the Capella Space catalogue [86], and their corresponding optical images from Google Maps. The optical image (a) shows several vessels in the scene and what can potentially be two icebergs, highlighted by two green circles. In the corresponding SAR image (b), it can be observed that vessels and icebergs exhibit similar reflectivity patterns, making them difficult to differentiate based solely on SAR backscatter. By aligning the optical and SAR images, it is easier to discern which bright SAR points correspond to icebergs and which correspond to ships. The SAR image in (c) shows a ship off the coast of Algeciras, with its corresponding optical image (d). In the example in Algeciras, the possibility of mistaking a vessel for an iceberg does not exist, but in other parts of the planet, like the Antarctic or Greenland, the possibility of discerning between the two can be critical.



(a) (b) (c) (d)

Figure 2.7: (a) Optical image showing icebergs and a small vessel of the coast of Nuuk (Greenland), (b) the corresponding SAR acquisition, (c) optical image of the port of Algeciras (Spain), and (d) its corresponding SAR image with a clearly visible ship.

In this context, active vessels typically present vibrations generated by their engines, propellers, and onboard machinery. These vibrations modulate the returned radar signal and produce characteristic patterns in the Doppler domain. In contrast, static objects such as icebergs or rocks do not produce such signatures. Therefore, detecting and characterising these micro-motion patterns can be used as a discriminative feature for vessel-non-vessel classification.

In the field of detection and classification, ML and Deep Learning (DL) have significantly advanced ATR capabilities in the last few years. Convolutional Neural Networks (CNNs) have been employed to classify maritime targets in high-resolution SAR images, showing an improved accuracy over traditional approaches [82]. Moreover, some studies have shown that CNNs can be trained directly on complex-valued SAR data. In [87], a network that uses real-valued weights kernels to obtain spatial features from complex-valued polarimetric data is presented. As a result, complex-valued outputs are given at each layer. The method proposed was compared to its real-valued counterpart and showed higher accuracy than the latter.

Two main configurations can be discerned for ATR systems built on satellite data:

- On-Ground Processing: the SAR data is transmitted to ground stations where ATR algorithms are applied. While this allows for more complex models and computational power, it introduces delays due to data transmission and processing times. Ground-based processing is beneficial for detailed analysis and for applications that are not real-time dependent [88].

- On-Board Processing: ATR algorithms that are implemented directly on the SAR platform make real-time target detection and classification possible. This approach reduces the latency associated with data transmission to ground stations and is crucial for time-sensitive applications. For example, the HISEA-1 SAR satellite employs on-board real-time ship detection using a combination of the CFAR algorithm and lightweight DL algorithms, achieving efficient processing with limited computational resources [89].

Recent developments in ATR for maritime applications focus on integrating Artificial Intelligence (AI) techniques to enhance detection accuracy and operational efficiency. Two methods for SAR ATR are presented in [90] exploiting the phase history domain, rather than the traditional image domain. Both frameworks use CNNs for learning and classification. The study demonstrates that phase histories provide rich dependency patterns that CNNs may effectively use, avoiding the need for SAR ATR tasks to rely just on generated SAR images.

A wide variety of CNN architectures have been proposed and evaluated for SAR ATR [91]. While shallow networks initially led the way, deeper networks with attention mechanisms [92] and even capsule networks have emerged as significant models [93]. Researchers continue to innovate, ensuring high accuracy even when data is limited. For an ML or DL approach, a large varied dataset is needed, which is often a challenge when working with SAR data. Gathering a dataset of SAR data requires experts to work with the data and it can be very time-consuming. To tackle the issue of limited samples in SAR ATR, methods like data augmentation, Generative Adversarial Network (GAN) for generating new samples, and transfer learning, have been researched and implemented with varying degrees of success [94–96]. Each method offers unique advantages and challenges, emphasising the importance of tailored approaches for SAR data [97].

Motivated by the lack of annotated SAR data databases, a study was carried out in [98] for a deep transfer learning approach for ship detection based on transfer learning. Since labelling data in the electro-optical domain is easier, a deep Neural Network (NN) is trained by transferring knowledge from a similar problem in the electro-optical domain (source) to the SAR data domain (target). Two deep encoders are trained to map data from both domains to a shared cross-domain embedding space that is also discriminative for classification. SAR data points are matched to their corresponding class of electro-optical domain in the shared embedded space. In this space, the representation becomes domain-invariant, and samples from different sensing modalities are aligned and cannot be distinguished based on their original domain. Image processing is used to determine the areas of interest in SAR images before further action [98]. A similar approach based on transfer learning between different domains is looked at in [99], where transfer learning is performed between AIS and SAR image domains based on the extraction of Naïve Geometric Features (NGF). NGFs are useful to reduce the complexity of image processing to obtain the basic differences between different types of vessels. Both domains are combined, and transfer learning provides the SAR images domain with the information provided by the AIS domain. To connect both, the width and length of the vessels are selected and an Adaptive Support Vector Machine (A-SVM) classifier is proposed for improving the classification with the transferred learning.

In [100], an ATR study that is based in Krawtchouk moments to characterise targets in SAR images is proposed. It is demonstrated, through the use of the MSTAR dataset [101], that this method is reliable for the recognition and identification of military vehicles, even in noisy environments. Moreover, it has been proven that, compared to other image moments, Krawtchouk offers a compact representation of the targets' features for accurate and efficient target recognition [102]. They can effectively represented extended targets with a more reduced set of features, which facilitates the characterisation of the objects in the scene, very useful in complex SAR imaging scenarios [103].

CFAR detectors are a standard first stage in many SAR processing chains, providing adaptive thresholds that maintain a false alarm probability under varying clutter conditions. In maritime surveillance, CFAR-based ship detectors are widely used to separate bright vessel returns from sea clutter, including in near-real-time and on-board processing scenarios where CFAR is combined with lightweight deep-learning models or local refinement to improve detection accuracy and robustness [104–107]. Similar CFAR strategies, including two-parameter and correlation-based joint CFAR variants, have demonstrated to perform well in complex clutter and multi-target scenes by modelling the properties of sea clutter more accurately [108].

Beyond ship detection, CFAR detection has been extended to superpixel-level and feature-enhanced frameworks, where CFAR operates on segments or texture-intensity representations rather than individual pixels. These methods exploit spatial coherence and polarimetric or textural information to better distinguish small or nearshore targets from strong background clutter, and they integrate CFAR outputs with subsequent ML stages [109–112]. In this context, CFAR works as a main detector that filters large SAR scenes down to a set of candidate regions of interest, which can then be passed to more computationally demanding processing. As a result, CFAR is a good choice for micro-motion and vibration recognition in SAR, for a controlled selection of potential vibrating or moving targets.

### 2.5.1 Image Moments

Since their introduction by Hu, image intensity moments have gained prominence in fields such as image analysis, reconstruction, edge detection, and pattern recognition [113]. Broadly, moments refer to quantitative values derived from an image's pixels' intensities that encapsulate specific properties of said image. Depending on the foundational type of the moment function, they can be classified into orthogonal and non-orthogonal moments [114]. Non-orthogonal moments, including geometric, rotational, and complex moments, offer low computational demands but are notably vulnerable to noise, complicating their reconstruction [115]. In contrast, orthogonal moments like Legendre [116], Zernike [117], and pseudo-Zernike moments [118] exhibit reduced noise sensitivity and are efficient for image reconstruction. However, computing these moments necessitates approximating the continuous moment integrals, which can result in numerical errors and compromise analytical characteristics like invari-

ance and orthogonality [119]. To address the limitations associated with continuous orthogonal moments, discrete orthogonal basis moments have been suggested, such as Tchebichef moments [120], Krawtchouk moments [121], Hahn moments [122], and Racah moments [123]. Krawtchouk have gained prominence among the various invariant moments, as they offer benefits in terms of improved reconstructive capabilities and are orthogonal over a finite interval, which makes them especially well suited for image analysis. Furthermore, Krawtchouk moments exhibit a higher degree of robustness against noise and are demonstrated to be more computationally efficient, which expedites processing and analysing tasks [121].
The different image moments can be categorised depending on their mathematical properties, as shown in Fig. 2.8, from [124].

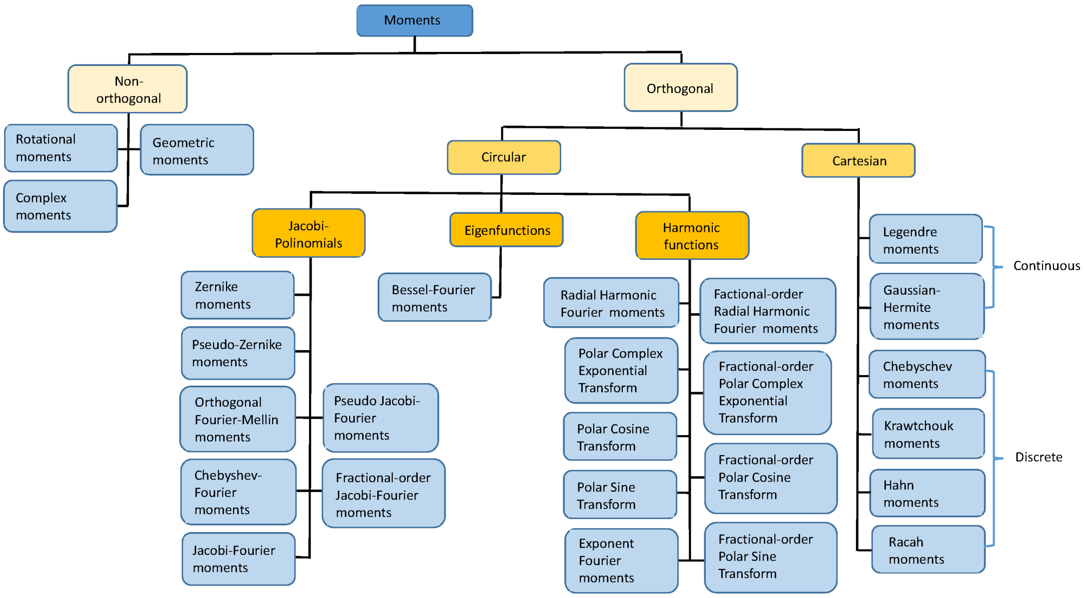


Figure 2.8: Schematic representation of image moments.

### 2.5.2 Krawtchouk Moments

Krawtchouk moments are based on discrete and weighted orthogonal polynomials known as Krawtchouk polynomials. They are well-fitted for digital image processing, as they can represent the information efficiently and in a compact manner, maintaining the original spatial structure [122]. Krawtchouk moments offer some key advantages for SAR image analysis, such as compact representation, translation and scaling invariance, and robustness to noise. Krawtchouk moments provide an efficient representation of the information contained in the images, useful when dealing with large matrices, which is the case for SAR data [125]. These moments can be used to extract invariant features under varying conditions, like different acquisition modes and SAR geometries, and moreover, they are less sensitive to noise compared to pixel-based features, making them suitable to be used on SAR data [126, 127]. The foundational elements of Krawtchouk moments are the discrete orthogonal Krawtchouk polynomials derived in [128], which are orthogonal with respect to a discrete weight function $j(x)$ over a finite range $0 \leq x \leq N$. When reconstructing or analysing an image, the orthogonality condition ensures that each polynomial represents independent information, preventing redundancy, and it is expressed as:

$$\sum_{x=0}^{N} j(x) k_n(x) k_m(x) = \rho(N, n, p)\delta_{nm}, \quad \text{for } 0 \leq n, m \leq N, \tag{2.13}$$

where:

- $k_n(x)$: The $n$-th order Krawtchouk polynomial.
- $j(x)$: The weight function, which is defined as $j(x) = \binom{N}{x} p^x (1-p)^{N-x}$.
- $\rho(N, n, p)$: The normalisation factor to ensure orthogonality.
- $\delta_{nm}$: The Kronecker delta, which is 1 if $n = m$, and 0 otherwise [129].

The orthogonality allows for an efficient decomposition of the information contained in the components that do not overlap.

Krawtchouk polynomials are explicitly defined as:

$$k_n(x) = \sum_{r=0}^{n} (-1)^r \binom{x}{r} \binom{N-x}{n-r} p^r (1-p)^{n-r}, \tag{2.14}$$

where: $n$ is the order of the polynomial, $x$ the pixel location, $N$, is the total number of spatial points, and $p$ is a parameter that determines the polynomials' centring within the interval $[0, 1]$, which controls the "focus" of the polynomials, allowing flexibility in the modelling of the spatial variations across an image. This gives Krawtchouk polynomials an advantage in capturing spatial structures in the images. Using Krawtchouk polynomials as a basis set, the Krawtchouk moments of an image $f(x, y)$ are computed as follows:

$$K_{nm} = [\rho(N, n, p)\rho(N, m, p)]^{-1} \sum_{x=0}^{N} \sum_{y=0}^{N} j(x)j(y)k_n(x)k_m(y)f(x, y), \tag{2.15}$$

where:

- $K_{nm}$: The $(n, m)$-th Krawtchouk moment.
- $f(x, y)$: The intensity value at pixel $(x, y)$ in the image.
- $k_n(x)$: The $n$-th order Krawtchouk polynomial in the $x$-dimension.
- $k_m(y)$: The $m$-th order Krawtchouk polynomial in the $y$-dimension.
- $j(x)$ and $j(y)$: Weight functions for the $x$ and $y$ dimensions, respectively.

Low-order moments capture global features like image intensity and smooth variations, while High-order moments capture fine details and textures [130]. In SAR image processing, Krawtchouk moments have been applied, especially for ATR.

## 2.6 Autocorrelation for pattern and vibration recognition

Autocorrelation is a classic tool for detecting periodic patterns and estimating fundamental frequencies in noisy time series, and is widely used across domains such as audio, vibration analysis, and radar. It enhances regular structures by correlating a signal with time-shifted versions of itself, so that recurrent patterns produce peaks at lags corresponding to their periods while uncorrelated noise averages out [131]. In audio and speech processing, autocorrelation has been used for pitch detection and voiced/unvoiced decision, where peaks in the short-time autocorrelation function directly reveal the fundamental period of quasi-periodic vocal vibrations. Similar principles are applied in mechanical vibration and condition monitoring, where autocorrelation of accelerometer or displacement signals helps identify dominant vibrations, detect periodic components associated with gear faults, and separate them from broadband noise. In these applications, autocorrelation-derived features (e.g., peak locations, peak heights, decay rates) are commonly fed into machine-learning classifiers for fault diagnosis [72, 132, 133].

In radar and related sensing systems, correlation-based processing is also central for extracting periodic micro-motions. For example, in distributed acoustic sensing and Phase-sensitive Optical Time-Domain Reflectometry (Φ-OTDR) systems, global phase demodulation followed by autocorrelation analysis is used to detect long-distance periodic vibrations and to estimate their frequencies with high sensitivity [74]. Correlation and autocorrelation measures have also been used in radar and ultra-wideband vital-signs detection to highlight small, periodic chest wall motions against clutter, and to stabilise period estimates in low-SNR conditions [76]. In airborne SAR and ground-based radar systems, correlation and autocorrelation of phase or displacement time series can be used as the basis of vibration monitoring and error compensation techniques. These are specially effective when the carrier frequency of the system is high, as even small mechanical motions can produce measurable phase modulations [134].

## 2.7 Recurrent Neural Networks in SAR image classification

In SAR ATR, the use of NNs shifts the approach from traditional statistical models to methods that can learn patterns directly from data. Unlike manually designed features, neural networks automatically find important features and structures in the data [135], making them suitable for dealing with complex and noisy data.

The application of NNs to SLC data is a significant advancement in SAR ATR. CNN-based models adapted for SLC data have been shown to learn phase-amplitude relationships, improving the detection and classification of targets in difficult environments [136]. RNNs are a type of NN architecture that are designed to process sequential data and have the ability to retain information from previous inputs through internal memory mechanism. This is achieved via recurrent connections in which neurons feed back into one another [137]. When working with SAR data and NNs, there are challenges to address:

- Dynamic range: The dynamic range of SAR images can be up to 90dB and the distribution of pixels is very asymmetric, being most of them in the low amplitude range. CNNs are not good at handling this and dynamic compression is usually performed [138].
- Signal statistics: the speckle is a multiplicative phenomenon that has to be taken into account when detecting features from SAR images. However, the speckle noise can be turned into additive random variable by using logarithmic mapping of the images before entering the CNN, which can also reduce the dynamic range. Speckle reduction filters, like non-local averaging, may be also applied to the images before feeding them to the network [138].
- Imaging geometry: data augmentation to have more extensive databases to train/ validate the network cannot be applied to SAR data. The SAR image coordinates, range and azimuth, are not arbitrary [138].
- Complex nature of SAR data: as the phase of SAR data contains the most important information, the system should be able to work with complex data. This is a problem for non-linear activation and loss function layers. The activation function can operate real and imaginary part individually or on the magnitude, but the phase will be distorted to some degree [138].
- Simulation-based training and validation data: using simulated data because of the lack of ground-truth for regression tasks can cause the system to learn on data that is too simplified [138].

In [139], a system based on an RNN is used to process received SAR signals instead of SAR images. It classifies targets from the slow-time and fast-time sampled signals based on sparse representation. This approach does not include image formation and it is based on sparse representation and a dictionary. The dictionary is formed by an optimal set of basis vectors in columns and the classes are the different subsets of those columns. The classification of the new samples is made by finding their sparse representations and their most suitable linear combination from the dictionary. This way, a new sample will belong to the closest linear combination of the columns, which most likely will be its class. Results show that for different SAR geometries and a variety of different backgrounds and target intensities, and even with limited SAR data and a small dictionary, this method results in promising outcomes.

### 2.7.1 Network combination to enhance RNN-based classification performance

#### Confidence-based combination

The confidence-based approach is a strategy used to improve decision-making in classification systems that involve multiple classifiers. Initially, the system evaluates an input using both classifiers, each providing its score or confidence level regarding the class the input belongs to. These scores are compared and the final decision is made. In the cases where an “unknown” class is included, this accounts for inputs that may not belong to any of the known categories. To decide whether an input belongs to the unknown class or not, a set of predetermined thresholds can be used. If no score surpasses the threshold, it is indicated that the system is not confident enough in its decision and the input is classified as belonging to the unknown class.

#### Logistic regression-based combination

A new model is trained using the outputs of the two classifiers as features. The logistic regression model comes into play as the meta-classifier. Instead of training on the original input features directly, the new model is trained on the outputs of the two base classifiers. It learns how to best combine the separate predictions to improve the overall prediction accuracy. The output scores from the base classifiers serve as input features for the logistic regression model. The final decision is based on its learned weights, effectively deciding how much trust to place in each of the base classifiers' predictions. The final output is a binary classification, outputting the final class for the combination of networks. This approach helps improve classification accuracy, as it leverages the strengths of individual classifiers and can compensate for their individual weaknesses [140].

## 2.8 Support Vector Machine for SAR micro-Doppler

SVMs are widely used in SAR for supervised classification and target recognition, especially when feature vectors are of low to moderate dimension and training data are limited. They fit well with pipelines that extract handcrafted features (e.g., from time-frequency or autocorrelation representations) and then perform binary or multi-class decisions [141]. In remote-sensing applications, SVMs have been extensively applied to land-cover mapping and change detection from SAR backscatter intensity and texture measures, including studies using Sentinel-1 data. Selecting the proper kernels and feature designs, SVM-based classifiers consistently achieve high accuracy in applications such as urban vegetation detection and post-disaster damage assessment. Their performance is often competitive with, or superior to, traditional statistical classifiers and comparable to other machine learning methods, especially when annotated data are not easily available. Similar advantages have been demonstrated in Polarimetric SAR (PolSAR) classification tasks [142, 143].

In SAR-based ATR, SVMs are commonly used to classify targets using feature vectors derived from Principal Component Analysis (PCA)/Kernel PCA (KPCA), Gabor-based descriptors, or other shape and scattering features; such approaches achieve good recognition performance on benchmark SAR datasets and are robust to moderate clutter and speckle. For micro-Doppler and motion-related applications, SVMs are frequently used when features that summarise time-frequency signatures are available [144, 145]. SVMs are well suited to operate on compact feature sets, manage high-dimensional but relatively small datasets, and support non-linear kernels to separate different motion patterns.

# Chapter 3

# A Cross-Cross-Correlation based Method for Joint Coregistration of Rotated Multitemporal SAR Images

Coregistration is a technique that ensures that multiple SARR images, often acquired at different points in time, under varying conditions, or by different sensors, are aligned such that each pixel in a main image corresponds to the same physical scatterer in a secondary image. From this point forward, the primary image will be referred to as the “reference” image, and the secondary image will be referred to as the ‘secondary’ image. Misregistration can significantly deteriorate the quality of SAR products, such as coherence maps or deformation models [146].

In the context of maritime surveillance, multitemporal SAR stacks can be used to detect and track vessels, analyse ship wakes, and monitor coastal and offshore infrastructures. These tasks rely on accurate geometric alignment of repeated SAR acquisitions, so keypoints in the images can be correctly associated over time. Misregistration can lead to false apparent motion, missed detections, and errors in estimated ship trajectories or wake patterns. The joint cross-cross-correlation-based coregistration algorithm proposed in this chapter addresses these issues by providing robust alignment under low SNR, decorrelation, and rotation, problems often found in typical maritime scenes affected by sea clutter and varying acquisition geometries [147–150]. However, maritime scenes are particularly challenging for coregistration due to the low SNR, strong temporal decorrelation, and the lack of persistent scatterers over time in open water.

This chapter introduces and extends a novel approach to coregistering multitemporal SAR images using a cross-cross-correlation-based method, called JCLS. Unlike conventional pairwise registration methods, the proposed technique jointly estimates displacement parameters, leveraging the interrelations among multiple secondary images during the optimisation process, reducing noise effects and enhancing registration accuracy. The proposed methodology extends the CLS optimisation framework [23] to account for rotational and translational misalignments, integrating additional information from the cross-cross-correlation of image patches, based on [1]. The main contributions are: a rigorous mathematical formulation of the joint coregistration problem, including the weighted CLS, an enhanced cross-cross-correlation-based algorithm, formalised into a completed and reproducible pipeline, including model construction and displacement estimation, a novel outlier rejection strategy, and an extensive experimental validation, including analyses of patch size, number of secondary images, and reference image selection. Experimental validation using airborne SAR datasets demonstrates significant improvements in registration accuracy, particularly for images affected by high rotation and low SNR.

## 3.1 Structure of the Chapter

This chapter is organised as follows: Section 3.2 presents an overview of the coregistration problem, focusing on the decomposition of translation and rotation in the context of multitemporal SAR coregistration. It includes challenges in pair-wise multitemporal SAR coregistration and presents the formulation of the Coregistration using Least Squares (LS) approach, along with its solution. Section 3.3 evaluates the displacement field through cross-correlations, detailing the procedure for constructing the cross-correlation model matrix. Section 3.4 discusses the cancellation of outlier equations to ensure data accuracy, and the experimental results are presented in Section 3.5, with a specific focus on the optimum selection of the reference image and the number of secondary images used. Finally, Section 3.6 provides conclusions and a discussion of the results presented, which concludes the findings and implications.

## 3.2 Problem Formulation

This section defines the coregistration problem as a geometric transformation estimation task. Specifically, given a set of SAR images $\{\boldsymbol{I}_0(z), \boldsymbol{I}_1(z), \ldots, \boldsymbol{I}_{K-1}(z)\}$ where each image is represented by a complex-valued matrix $\boldsymbol{I}_k(z) \in \mathbb{C}^{M \times N}$, and $z = x + jy$ represents the complex Cartesian coordinates [151]. The image indexed by $k = 0$ is selected as the *reference image* $\boldsymbol{I}_m(z)$, while the remaining $K-1$ images are designated as *secondaries* $\boldsymbol{I}_{s_k}(z)$, with $k = 1, \ldots, K-1$. Each secondary image $I_k$ can be represented as a transformed version of the reference image $I_0$, subjected to translation, rotation, and noise:

$$I_{sk}(z) = I_m\left(\frac{z - \delta_k}{\alpha_k}\right) + E_k(z), \quad \forall k \in \{1, 2, \ldots, K-1\} \tag{3.1}$$

where: $\mathbf{E}_k(z)$ denotes the error term, accounting for noise, image decorrelation, and differences in scattering properties between the reference and secondary images, and $\delta_k$ is the complex displacement vector, representing the translation of the secondary image $I_k$ with respect to the reference in the $x$- and $y$-directions. The parameters $\alpha_k = \gamma_k e^{j\theta_k} \in \mathbb{C}$ is the complex scaling factor, where $\gamma_k \in \mathbb{R}^+$ represents the zoom or scaling factor (typically $\gamma_k = 1$, assuming no scaling), and $\theta_k \in \mathbb{R}$ represents the rotation angle of the secondary image relative to the reference.

### Translation and Rotation Decomposition

The transformation parameters $\delta_k$ and $\alpha_k$ can be decomposed as follows:

- Translation:

  $$\delta_k = \delta_{xk} + j\delta_{yk} \tag{3.2}$$

  where $\delta_{xk}$ and $\delta_{yk}$ are the horizontal and vertical shifts, respectively.

- Rotation:

  $$\alpha_k = e^{j\theta_k} = \cos(\theta_k) + j\sin(\theta_k) \tag{3.3}$$

  where $\theta_k$ is the rotation angle. The effect of this rotation can be expressed as:

  $$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta_k) & -\sin(\theta_k) \\ \sin(\theta_k) & \cos(\theta_k) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

  where $(x', y')$ are the new coordinates after rotation.

The goal of the novel coregistration process presented is to jointly estimate the unknown parameters $\{\delta_k, \theta_k\}_{k=1}^{K-1}$ for all secondary images such that the transformed secondary images are geometrically aligned with the reference image. This can be formulated as a constrained optimisation problem:

$$\{\hat{\delta}_k, \hat{\theta}_k\} = \arg\min_{\delta_k, \theta_k} \mathcal{L}\left(I_0, I_k(\delta_k, \theta_k)\right) \tag{3.4}$$

where $\mathcal{L}(\cdot)$ is a loss function that measures the misalignment between the reference and secondary images, typically based on pixel-wise or patch-wise cross-correlation. The proposed procedure in this chapter provides an efficient solution in the estimation of the unknown parameters $\delta_k \in \mathbb{C}$ and $\theta_k \in \mathbb{R}$, enforcing the absence of zooming effect (i.e., setting $\gamma_k = 1$), for all the $K-1$ secondary images. More precisely, the proposed method jointly estimates the mentioned parameters also accounting for the respective misalignments between the secondary images. In this respect, the proposed method is based on the solution of the CLS problem developed in [23], after a proper selection of the areas of interest in the images, as in [152]. In [23], the translation and rotation parameters of each secondary image with respect to a reference image are calculated by jointly exploiting multiple patch-wise displacement measurements. This approach remains fundamentally pairwise, as each secondary image is independently aligned to the reference, and no explicit constraints are imposed on the relative alignment between different secondary images, and inter-secondary correlations are not taken into account. More in detail, once the patches from the reference and secondary are extracted, a cross-cross-correlation-based method, devised in [153] for delay estimation for 1-dimensional signals, is applied to obtain all the displacement fields for the secondary images in the considered stack with respect to the selected reference.

The proposed algorithm is described by the functional scheme shown in Fig. 3.1 whose main steps are detailed in the following.

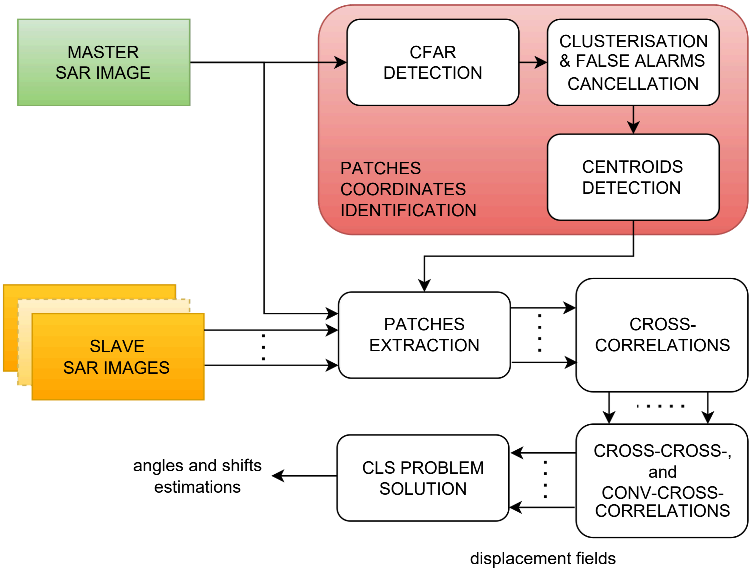


Figure 3.1: Block scheme of the multitemporal SAR images coregistration algorithm based on using the cross-cross-correlations.

The starting input of the algorithm is the image selected to be the reference that is used to apply the procedure developed in [152] to properly select the patches needed for the displacement field evaluation. In particular, the method of [152] applies a CFAR [154] detection scheme to the entire image to detect strong reflective areas, such as vessels, offshore platforms, buoys, or harbour structures in maritime scenes. The resulting binary detection map is then improved thanks to a clustering and false alarms cancellation procedure. The clustering, based on the use of an order filter, allows to improve the shape of the detected object/area exploiting the behaviour of the neighbouring pixels. The CFAR detector produces a binary map in which fragmented targets can appear due to thresholding effects and speckle. To recover missed pixels and enforce spatial continuity of detected objects, a rank-order filter is applied to the binary map. The filter updates each pixel based on the ordered values in its neighbourhood, effectively performing a majority-based clustering of detections. Meanwhile, the false alarm cancellation, based on a median filter, is performed to delete all noisy single detections belonging to the map. Once the extended objects are definitely obtained in the reference image, their centroids are evaluated as their centre of mass. Then, for each centroid in the reference, a patch in its centre is extracted from both the reference and all secondary images and all possible cross-correlations between these corresponding patches in the $K$ images are computed. The reason of performing the patch detection only on the reference image is twofold: from one side it allows to reduce the computation burden with respect to applying the detection process over more images. On the other side, by doing so, the risk of losing a matching between patches in each couple of images is mitigated, relevant in maritime time series where vessels may appear and disappear between acquisitions, while stable infrastructures, such as piers, provides persistent tie-points that anchor the registration. Additionally, the cross-correlations between each couple of patch cross-correlations are evaluated before constructing the overall displacement field needed to solve the CLS problem detailed in Section 3.2.1. As shown in [153] for Time Difference Of Arrival (TDOA) estimation, by exploiting the additional information provided by the cross-cross-correlation, it would be expected that the new method performs better than the classic competitor in the presence of correlated images and high-power noise. Moreover, random errors arising in the estimation of the cross-correlation peaks could be reduced since a higher number of equations is employed in searching the pseudosolution of the LS problem, creating a highly over-determined system by generating a much larger number of equations than unknowns, reducing random errors. Furthermore, the fourth-order correlation, which involves computing correlations from the existing couples of cross-correlations, can also be seen as a filtering operation on the incoming images. Finally, the procedure is repeated for each patch detected in the reference and extracted in all secondaries. By doing so, the displacement field is evaluated for each secondary as the vector containing the single displacements of each couple of tie-points in the reference-secondary solving the overall cross-cross-correlations problem.

### 3.2.1 CLS formulation and its solution

The CLS problem designed in [23] allows the registration of a secondary image $\boldsymbol{I}_s(z) \in \mathbb{C}^{M\times N}$, affected by rotation and shift, to a reference $\boldsymbol{I}_m(z) \in \mathbb{C}^{M\times N}$ representative of the same observed scene. To do this, the method needs to identify some tie-points (described in Section 3.3) in both the reference and secondary, indicated as $z_l = (x_{m,l} + jy_{m,l})$, and $\zeta_l = x_{s,l} + jy_{s,l},\ l = 1,\dots,L$, respectively, where $L$ is the number of tie-points in the image. Then, the displacement field between them can be obtained as the solution of an over-determined constrained linear system of $L$ equations in 3 unknowns (shift in $x$ and $y$ directions, and the rotation angle) [23], that is

$$\alpha z_l + \delta = \zeta_l,\ l = 1,\dots,L \tag{3.5}$$

Equivalently, (3.5) can be expressed in matrix form as

$$\begin{cases} \arg\min\limits_{\boldsymbol{p}} \|\boldsymbol{A}\boldsymbol{p} - \mathbf{diag}\,(\boldsymbol{w})\boldsymbol{\zeta}\|^2 \\ \text{s.t. } \boldsymbol{p}^\dagger \boldsymbol{D}\boldsymbol{p} - 1 = 0 \end{cases}, \tag{3.6}$$

with $\boldsymbol{p} = [\alpha, \delta]^T$ the vector containing the unknown parameters, and $\boldsymbol{\zeta} = [\zeta_1, \dots, \zeta_L]^T$ the measurement vector containing the coordinates $\zeta_1, \dots, \zeta_L$ of the displacement field expressed with respect to the reference system centred at the image centre. In (3.6), the weighting vector $\boldsymbol{w} = [w_1, \dots, w_L]^T$ is also considered to weigh the impact of each extracted patch in the LS problem. Finally, $\boldsymbol{A} = \mathbf{diag}\,(\boldsymbol{w})\boldsymbol{Z}$, with

$$\boldsymbol{Z} = \begin{bmatrix} z_1 & 1 \\ \vdots & \vdots \\ z_L & 1 \end{bmatrix}$$

the $L \times 2$ matrix containing the complex coordinates of the tie-points in the reference image, and

$$\boldsymbol{D} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

the constraint matrix in the CLS problem, to reinforce the unit-magnitude condition $|\alpha| = 1$, ensuring that the estimated transformation corresponds to a pure rotation with no scaling.

As shown in [23], an optimal solution to Eq. (3.6) can be found in [146]. Therefore, assuming that the matrix $\boldsymbol{A}$ is full column rank, the optimal solution to the CLS is

$$\boldsymbol{p}^{\star} = \left(\boldsymbol{A}^{\dagger}\boldsymbol{A} + \beta\boldsymbol{D}\right)^{-1} \boldsymbol{A}^{\dagger}\mathbf{diag}\left(\boldsymbol{w}\right)\boldsymbol{\zeta}, \tag{3.7}$$

with $\beta$ solution of $\varphi(\beta) = 0$, $\beta \in \mathcal{I} \subseteq \mathbb{R}$, and $\varphi(\beta) = \boldsymbol{p}^{\dagger}\boldsymbol{D}\boldsymbol{p} - 1$. The interval $\mathcal{I}$ consists of all $\beta$ for which $\boldsymbol{A}^{\dagger}\boldsymbol{A} + \beta\boldsymbol{D}$ is positive definite, which implies that

$$\beta \in \left(-\frac{1}{\lambda_1(\boldsymbol{D}, \boldsymbol{A}^{\dagger}\boldsymbol{A})}, +\infty\right). \tag{3.8}$$

## 3.3 Displacement field evaluation through the use of cross-cross-correlations

The proposed method JCLS addresses the limitations of conventional pairwise coregistration methods by using cross-cross-correlation and CLS. A standard procedure to estimate the displacement field (i.e., the values $\zeta_l$, $l = 1, \ldots, L$, contained in the vector $\boldsymbol{\zeta}$) used to form Problem (3.6) consists in evaluating the position of the peak in the magnitude of cross-correlation between corresponding patches in the reference and secondary as

$$[\hat{y}, \hat{x}]_k = \arg\max_{y,x} \left\{ \left|\boldsymbol{G}_{0,k}(y, x)\right| \right\}, \; k = 1, \ldots, K - 1. \tag{3.9}$$

where

$$\begin{aligned} \boldsymbol{G}_{0,k}(y,x) = & \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \boldsymbol{P}_0(m,n)\boldsymbol{P}^*_{s_k}(m-y,n-x), \\ & -(M-1) \leq y \leq (M-1), \\ & -(N-1) \leq x \leq (N-1) \end{aligned} \tag{3.10}$$

is the spatial cross-correlation of the quoted couple of patches $\boldsymbol{P}_0$ and $\boldsymbol{P}_{s_k}$ in the reference and $k$-th secondary, respectively. Note that, in the previous equations the subscript $s$ is omitted to indicate that $x$ and $y$ are the displacement in the $x$- and $y$-direction of the patch associated to a secondary image.

Beyond the classic cross-correlation $\boldsymbol{G}$ evaluated with respect to the reference, it is possible to consider all possible couples of images (discarding the auto-correlation). In this case, the total number of possible combinations of $K$ images is $Q = \left(K^2 - K\right)/2$. Additionally, it is useful to derive the cross-correlation and convolution (say conv-cross-correlation) between each couple of images cross-correlations, in order to obtain $T = K^4/4 - K^3/2 - K^2/4 + K/2$ combinations.

The cross-cross-correlation and the conv-cross-correlation (also denoted as flipped cross-cross-correlation) can be, respectively, defined as

$$\begin{aligned} \boldsymbol{C}_{ihlp}(\rho_y,\rho_x) = & \sum_{y=0}^{M+N-1}\sum_{x=0}^{M+N-1} \boldsymbol{G}_{ih}(y,x)\boldsymbol{G}^*_{lp}(y-\rho_y,x-\rho_x), \\ & -(M+N-1) \leq \rho_y \leq (M+N-1), \\ & -(M+N-1) \leq \rho_x \leq (M+N-1), \\ & i,h,l,p = 0,\ldots,K-1 \; (h > i \text{ and } p > l), \end{aligned} \tag{3.11}$$

and

$$\begin{aligned} \boldsymbol{F}_{ihlp}(\rho_y, \rho_x) = & \sum_{y=0}^{M+N-1} \sum_{x=0}^{M+N-1} \boldsymbol{G}_{ih}(y,x)\boldsymbol{G}_{lp}(\rho_y - y, \rho_x - x), \\ & -(M+N-1) \leq \rho_y \leq (M+N-1), \\ & -(M+N-1) \leq \rho_x \leq (M+N-1), \\ & i,h,l,p = 0,\dots,K-1 \; (h > i \text{ and } p > l). \end{aligned} \tag{3.12}$$

The apex of the magnitude of the cross-cross-correlation, $|\boldsymbol{C}_{ihlp}(\rho_y, \rho_x)|$, should be at the index

$$\left[y_i - y_h - y_l + y_p, x_i - x_h - x_l + x_p\right], \tag{3.13}$$

while that of $|\boldsymbol{F}_{ihlp}(\rho_y, \rho_x)|$ should be at the index

$$\left[y_i - y_h + y_l - y_p, x_i - x_h + x_l - x_p\right]. \tag{3.14}$$

Hence, the $K-1$ displacements can be estimated in the MMSE sense solving the overdetermined system made by the $T$ equations, consisting of the linear combination of the $2(K-1)$ unknowns (i.e., $K-1$ unknowns for each $x$- and $y$- coordinate) equal to the index of the maximum of the standard and flipped cross-cross-correlations considered in (3.11) and (3.12), that is

$$\begin{aligned} & [y_i - y_h - y_l + y_p, x_i - x_h - x_l + x_p] = [\bar{\rho}_y, \bar{\rho}_x]_{ihlp} \\ & \quad i,h,l,p = 0,\dots,K-1 \, (h > i \text{ and } p > l), \end{aligned} \tag{3.15}$$

and

$$\begin{aligned} & [y_i - y_h + y_l - y_p, x_i - x_h + x_l - x_p] = [\breve{\rho}_y, \breve{\rho}_x]_{ihlp} \\ & \quad i,h,l,p = 0,\dots,K-1 \, (h > i \text{ and } p > l), \end{aligned} \tag{3.16}$$

where

$$[\bar{\rho}_y, \bar{\rho}_x]_{ihlp} = \arg\max_{\rho_y,\rho_x} \left\{ \left| \boldsymbol{C}_{ihlp}(\rho_y, \rho_x) \right| \right\}, \tag{3.17}$$

and

$$[\breve{\rho}_y, \breve{\rho}_x]_{ihlp} = \arg\max_{\rho_y,\rho_x} \left\{ \left| \boldsymbol{F}_{ihlp}(\rho_y, \rho_x) \right| \right\}. \tag{3.18}$$

Resorting to a compact matrix form, (3.15)-(3.16) can be rewritten as

$$\boldsymbol{M}\boldsymbol{U} = \boldsymbol{\Xi}, \tag{3.19}$$

with

$$\boldsymbol{U} = [\boldsymbol{y}, \boldsymbol{x}] = \begin{bmatrix} y_1 & x_1 \\ \vdots & \vdots \\ y_{K-1} & x_{K-1} \end{bmatrix}, \tag{3.20}$$

$$\begin{aligned} \boldsymbol{\Xi} &= \left[\boldsymbol{\rho}_y, \boldsymbol{\rho}_x\right] \\ &= \begin{bmatrix} \bar{\rho}_{y_{0102}} & \bar{\rho}_{x_{0102}} \\ \vdots & \\ \bar{\rho}_{y_{(K-3)(K-1)(K-2)(K-1)}} & \bar{\rho}_{x_{(K-3)(K-1)(K-2)(K-1)}} \\ \breve{\rho}_{y_{0102}} & \breve{\rho}_{x_{0102}} \\ \vdots & \\ \breve{\rho}_{y_{(K-3)(K-1)(K-2)(K-1)}} & \breve{\rho}_{x_{(K-3)(K-1)(K-2)(K-1)}} \end{bmatrix} \end{aligned} \tag{3.21}$$

The model matrix $\boldsymbol{M}$ of size $T \times (K-1)$ depends only on the number of multitemporal SAR images $K$ and comprises several null elements and some non-zero elements equal to $\pm 1$ and $\pm 2$. More specifically, the values $\pm 1$ are related to measurements involving an image in a single operation, e.g., one cross-correlation. Whereas, the values $\pm 2$ are related to measurements where an image is involved twice, e.g., in both the cross-correlations used in the cross-cross. For this reason, it can be computed and stored reducing the computational complexity in real-time algorithms. Furthermore, the solution to (3.19) is obtained through the pseudo-inverse of $\boldsymbol{M}$, that is

$$[\hat{\boldsymbol{y}}, \hat{\boldsymbol{x}}] = \hat{\boldsymbol{U}} = \left(\boldsymbol{M}^T \boldsymbol{M}\right)^{-1} \boldsymbol{M}^T \boldsymbol{\Xi}. \tag{3.22}$$

Finally, repeating the procedure for all the $L$ patches, the entries of $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{x}}$ are used to construct the displacement field of each secondary, $\boldsymbol{\zeta}^{(l)}$, $l = 1, \ldots, L$, needed for rotation angle and displacement estimation in (3.6).

Algorithm 1 shows the pseudo-code that summarises the main steps involved in the proposed cross-cross-correlation based procedure for joint displacement estimation. This procedure is hence repeated $L$ times for each patch detected in the reference.

**Algorithm 1** Pseudo-code for the proposed framework for displacement fields evaluation

```
Input: Extracted patches from reference $\boldsymbol{P}_0$ and secondary images $P_{s_k}$, $k = 1, \ldots, K-1$;
Output: Estimated displacement fields associated with all the secondary images $[\hat{\boldsymbol{y}}, \hat{\boldsymbol{x}}]$.
Model Matrix Definition
Compute the model matrix $\boldsymbol{M}$ of size $T \times (K-1)$;
Measurements Acquisition
Evaluate the cross-correlation estimates through (3.10) for each couple reference-secondary and for all the extracted patches;
Compute the cross-cross- and conv-cross-correlations through (3.11)-(3.12) for each couple of cross-correlations;
Perform the measurements of the peaks' positions of the cross-cross-correlations and its flipped version through (3.17) and (3.18) and store them in the matrices $[\bar{\rho}_y, \bar{\rho}_x]_{ihlp}$ and $[\breve{\rho}_y, \breve{\rho}_x]_{ihlp}$;
Construct the measurements matrix $\boldsymbol{\Xi}$ through (3.21).
Solutions Computation
Compute the solution with the pseudo-inverse of $\boldsymbol{M}$ through (3.22).
```

### 3.3.1 Detailed procedure for cross-cross-correlation model matrix construction

For clarity, more details on how to construct the model matrix referred to above, as well as some practical examples, are shown here. The model matrix $\boldsymbol{M}$ of size $T \times (K-1)$ can be constructed following the procedure described in [153]. It starts considering all possible cross-cross-correlations $\boldsymbol{C}_{ihlp}$ that are sequentially re-numbered by the index $r$ with $1 \leq r \leq T/2$. The four corresponding vectors of size $T/2$, $\boldsymbol{l}_1$, $\boldsymbol{l}_2$, $\boldsymbol{l}_3$, and $\boldsymbol{l}_4$ containing respectively the values assumed by the indices $i$, $h$, $l$, $p$ associated with the $r$-th cross-cross-correlation, are hence introduced. Then, starting from $\boldsymbol{M} = \boldsymbol{0}$, the $(r, c)$-th element of $\boldsymbol{M}$ is obtained adding $+1$ if $c = \boldsymbol{l}_1(r)$ or $c = \boldsymbol{l}_4(r)$ and $-1$ if $c = \boldsymbol{l}_2(r)$ or $c = \boldsymbol{l}_3(r)$. The same four vectors of indices are used for the flipped cross-cross-correlations $\boldsymbol{F}_{ihlp} = \boldsymbol{C}_{ihpl}$ to fill the elements $T/2 + 1 \leq r \leq T$ of $\boldsymbol{M}$, after switching the rule by adding $+1$ if $c = \boldsymbol{l}_1(r)$ or $c = \boldsymbol{l}_3(r)$, and $-1$ if $c = \boldsymbol{l}_2(r)$ or $c = \boldsymbol{l}_4(r)$.

In the following, some numerical examples are reported to help in understanding the above procedure. In particular, assuming the availability of $K = 3$ SAR images, the resulting $Q = 3$ cross-correlations are $\{\boldsymbol{G}_{01}, \boldsymbol{G}_{02}, \boldsymbol{G}_{12}\}$. Then, the $T = 6$ normal and flipped cross-cross-correlations are $\{\boldsymbol{C}_{0102}, \boldsymbol{C}_{0112}, \boldsymbol{C}_{0212}\}$ and $\{\boldsymbol{F}_{0102}, \boldsymbol{F}_{0112}, \boldsymbol{F}_{0212}\}$, respectively. As a consequence the four vectors containing the subscripts of $\boldsymbol{C}$ and $\boldsymbol{F}$ are

$$\boldsymbol{l}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{l}_2 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{l}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{l}_4 = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}.$$

Each row of $\mathbf{M}$ can be interpreted by writing the corresponding peak equation and collecting the coefficients of $y_1, ..., y_{K-1}$ or $x_1, ..., x_{K-1}$, according to Eq. 3.13 for the cross-cross-correlation, and Eq. 3.14 for the flipped cross-cross-correlation. As an illustrative example, the first equation reads $y_0 - y_1 - y_0 + y_2$. Since $y_0$ refers to the reference image and is therefore not an unknown displacement, the corresponding terms cancel out, yielding $-y_1 + y_2$. Consequently, the first row of **M** is [-1, 1]. The subscripts $0, 1, 2, \ldots$ in the example are directly obtained from the vectors $\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3$, and $\boldsymbol{l}_4$, which collect the indices $(i, h, l, p)$ of each cross-cross-correlation measurement. Once the first 3 rows in **M** are derived, the following rows are obtained from the flipped cross-cross-correlation. Consider the fourth measurement, for which the associated indices are $(i, h, l, p) = (0, 1, 0, 2)$. In this case, the peak location is $y_0 - y_1 + y_0 - y_2$. Since $y_0$ corresponds to the reference image, it is removed, and the remaining terms result in $-y_1 - y_2$. Hence, the corresponding row of the model matrix is [-1, -1].

Applying the above-described procedure, the $6 \times 3$ model matrix results to be equal to

$$\boldsymbol{M} = \begin{bmatrix} -1 & 1 \\ -2 & 1 \\ -1 & 0 \\ -1 & -1 \\ 0 & -1 \\ 1 & -2 \end{bmatrix}.$$

Similarly, for $K = 4$ SAR images, there are $Q = 6$ cross-correlations, i.e., $\{\boldsymbol{G}_{01}, \boldsymbol{G}_{02}, \boldsymbol{G}_{03}, \boldsymbol{G}_{12}, \boldsymbol{G}_{13}, \boldsymbol{G}_{23}\}$. As a consequence $T = 30$ (i.e., 15 cross-cross-correlations and 15 flipped cross-cross-correlations). Hence, the four index vectors of size equal to

15 are

$$\boldsymbol{l}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{l}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ 2 \\ 3 \\ 3 \\ 3 \\ 2 \\ 2 \\ 3 \end{bmatrix}, \quad \boldsymbol{l}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 2 \\ 0 \\ 1 \\ 1 \\ 2 \\ 1 \\ 1 \\ 2 \\ 1 \\ 2 \\ 2 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{l}_4 = \begin{bmatrix} 2 \\ 3 \\ 2 \\ 3 \\ 3 \\ 3 \\ 2 \\ 3 \\ 3 \\ 2 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \end{bmatrix}.$$

Again, applying the above-described procedure, the $30 \times 4$ model matrix results to be equal to

$$\boldsymbol{M} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ -2 & 1 & 0 \\ -2 & 0 & 1 \\ -1 & -1 & 1 \\ 0 & -1 & 1 \\ -1 & 0 & 0 \\ -1 & -1 & 1 \\ 0 & -2 & 1 \\ -1 & 1 & -1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \\ 1 & -2 & 1 \\ 1 & -1 & 0 \\ -1 & -1 & 0 \\ -1 & 0 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & 1 & -1 \\ 0 & -1 & -1 \\ 1 & -2 & 0 \\ 1 & -1 & -1 \\ 0 & 0 & -1 \\ 1 & -1 & -1 \\ 1 & 0 & -2 \\ 0 & 1 & -2 \\ 2 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & -2 \end{bmatrix}.$$

The formation of the model matrix $\boldsymbol{M}$ provides a systematic framework to encode the relationships between cross-cross-correlations and their flipped counterparts for multitemporal SAR images. This matrix is essential for downstream applications, enabling robust alignment and integration of SAR data across different acquisitions.

## 3.4 Outlier equation cancellation

Once the displacement field is obtained through (3.22), and consequently, Problem (3.6) is formalised, the set of equations in it contained can be refined through the application of the iterative outlier cancellation procedure developed in [152, 155]. More precisely, this framework firstly evaluates the absolute error for each entry of the estimated parameters vector $\hat{\boldsymbol{p}}$, namely

$$\boldsymbol{\epsilon} = \left| \boldsymbol{A}\hat{\boldsymbol{p}} - \mathbf{diag}\,(\boldsymbol{w})\boldsymbol{\zeta} \right|, \tag{3.23}$$

with $|\cdot|$ the absolute value of each element in its vector argument. Then, it rejects all equations that share errors higher than a threshold set according to the so-called Median Absolute Deviation (MAD) criterion [156,157]. Finally, the resulting system of equation (characterised by having a reduced size) is solved to recover the final solution [152,155].

## 3.5 Experimental Results

The performance of the proposed algorithm for joint coregistration of multitemporal SAR images is assessed on the challenging full-polarimetric *Gotcha Volumetric SAR Data Set V1.0* [158], characterised by having a full azimuth coverage and eight different elevation angles with images acquired at different time instants. Although the experiments are conducted on the airborne Gotcha dataset, the demonstrated robustness to rotation, noise, and multitemporal decorrelation directly transfers to maritime surveillance scenarios, where similar conditions are given when monitoring vessel traffic and offshore structures over time. The sensor used for the acquisitions is located on a plane and operates at a carrier frequency of 9.6 GHz with a wide bandwidth of 640

MHz. The observed scene is a parking lot containing several civilian vehicles (cars, forklifts, tractors) and also calibration targets. For the conducted study, the aperture was divided into azimuth sub-apertures of 4°, providing approximately equal range-azimuth resolution cells of 23 cm. By doing so, the resulting dataset comprises 90 images (looks) of 501 × 501 pixels for each of the 8 circular passes (different elevations) in the four polarisations (viz., HH, VV, HV, VH). To better understand the observed scene, Fig. 3.2 depicts the span (expressed in dB) of the full-polarimetric Gotcha SAR image at 0 degrees in azimuth.

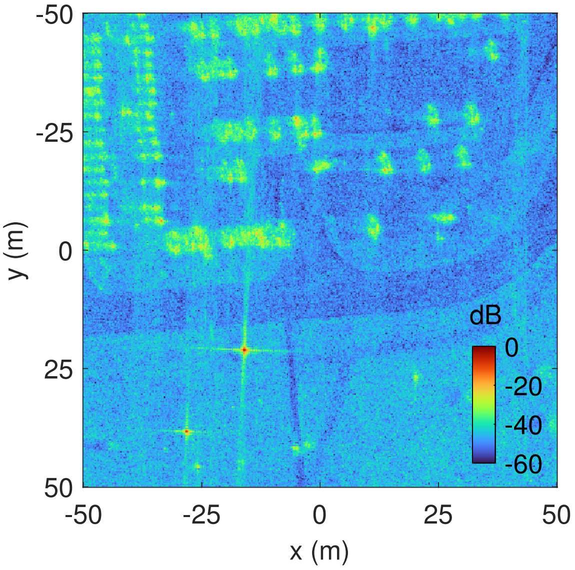


Figure 3.2: Span (dB) of the full-polarimetric Gotcha SAR image at 0° azimuth.

In the next tests, without loss of generality, focusing on the HH polarisation, considering all eight passes of acquisition once the azimuth angle has been fixed; then, the reference is chosen to be the image at pass 1, whereas the secondaries are those from pass 1 to 7. Since the Gotcha images are provided already registered to each other, in the devised tests, each secondary is clockwise or counterclockwise rotated by an angle $\theta_k$, $k = 1, \dots, K-1$, followed by a nearest neighbour interpolation to compensate the non-integer translation of the pixels.

The analyses are conducted considering as figure of merit the Root Mean Square

Error (RMSE) of the estimated angles

$$\text{RMSE} = \sqrt{\mathbb{E}\left[\left\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right\|^2\right]}, \tag{3.24}$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_7]^T$ is the vector containing the seven angles to estimate, whereas $\hat{\boldsymbol{\theta}}$ is the vector containing their estimates. Now, because of the lack of a close form expression for the RMSE, it is numerically evaluated resorting to the Monte Carlo simulation procedure. More precisely, at each Monte Carlo trial, $i = 1, \ldots, M_c$ (with $M_c$ the number of runs), each secondary image is rotated by an angle $\theta_k$, $k = 1, \ldots, 7$, randomly chosen in the interval $[-2°, 2°]$. The first test aims at empirically evaluating the optimum choice for the patch size. The size of the patches extracted from the imagery will directly impact the final coregistration performance. This size can be a priori set based on considerations about the overall image extent as well as the size of targets that are expected to be in it. Hence, Fig. 3.3 shows the RMSE (expressed in °) versus the patch size, having considered, without loss of generality, square shape patches. The curves are related to the sequence of 8 images for three different azimuth angles, viz. 0°, 176°, and 356°. Moreover, the RMSE is evaluated over a total of $M_c = 100$ Monte Carlo runs chosen the true rotation angles as above described.

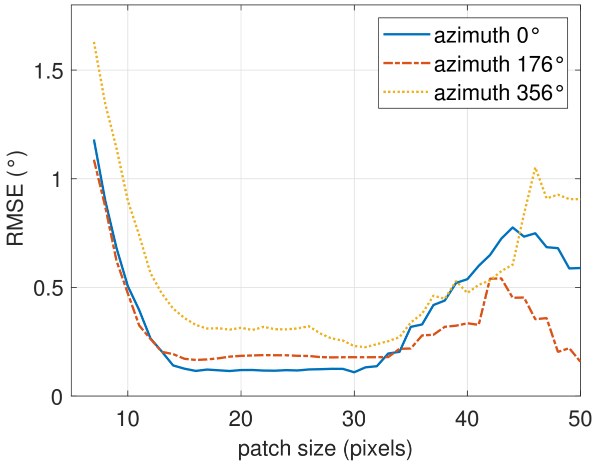
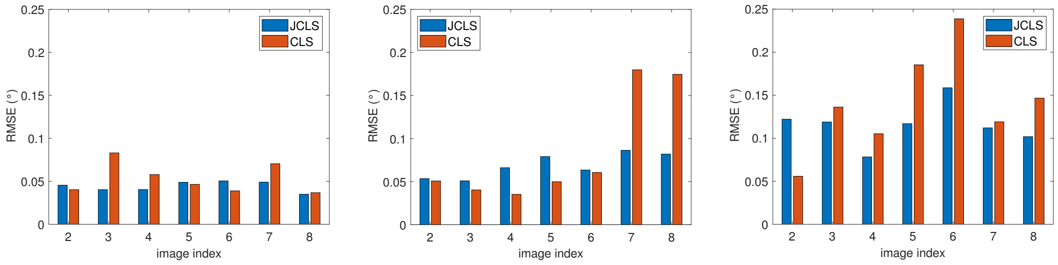


Figure 3.3: RMSE (°) versus patch size for three sequences of 8 SAR images acquired at different azimuth angles, viz. 0°, 176°, and 356°. A total of 100 Monte Carlo runs is performed randomly selecting the rotation angles in the interval $[-2°, 2°]$.

As expected the three curves show a coherent behavior, showing that the RMSE is higher when the patch size is chosen to be very small (in this case, possible extended targets are spread over more patches) and also when it is chosen to be too large (in this case, more than one target could be contained in the same patch). In particular, from the graph, it can be assured that the optimal patch size for these images is between 18 × 18 and 32 × 32 pixels.

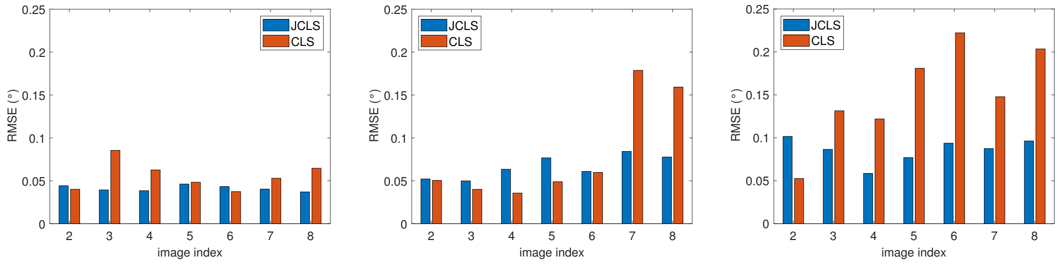
For the above-mentioned reason, in the successive tests, the patch size is set to 25 × 25 and 30 × 30 pixels. Therefore, Figs. 3.4 and 3.5 compare the proposed method, indicated as JCLS, with the CLS of [152] evaluating their achieved RMSE values for each secondary image. The tests are again conducted for the same settings as in the previous analyses with the results achieved for the three different azimuth angles (viz. 0°, 176°, and 356°) shown in the respective subplots.



(a) azimuth 0° (b) azimuth 176° (c) azimuth 356°

Figure 3.4: RMSE (°) for each secondary of the JCLS and CLS algorithms. The patch size is set to 25×25 pixels and subplots refer to 8 SAR images acquired at eight different elevations (marked by the image index). Moreover, three different tests are conducted fixing the azimuth angles to a) 0°, b) 176°, and c) 356°. A total of 100 Monte Carlo runs is performed randomly selecting the rotation angles in the interval $[-2°, 2°]$.

The bar graphs emphasise the superiority of the JCLS in jointly estimating the involved rotation angles. It can be noticed that, for some specific images, the RMSE of the JCLS is slightly higher than that of the CLS, but it compensates with much better results for the others. The RMSE values shown by the JCLS are mostly homogeneous, whereas those of the CLS are strongly unbalanced. Therefore, it can be concluded that the JCLS tends to mitigate the rotation angle estimation performance error to provide a more balanced situation between the involved stack of images.



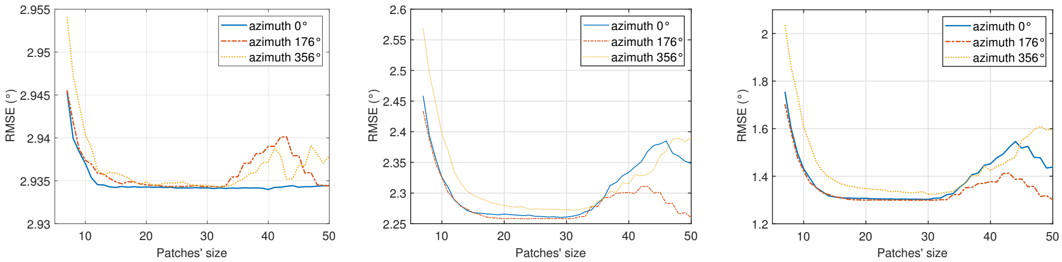
(a) azimuth 0° (b) azimuth 176° (c) azimuth 356°

Figure 3.5: RMSE (°) for each secondary of the JCLS and CLS algorithms. The patch size is set to $30 \times 30$ pixels and subplots refer to 8 SAR images acquired at eight different elevations (marked by the image index). Moreover, three different tests are conducted fixing the azimuth angles to a) 0°, b) 176°, and c) 356°. A total of 100 Monte Carlo runs is performed randomly selecting the rotation angles in the interval $[-2°, 2°]$.

### 3.5.1 Reference image and number of secondary selection

For a further evaluation of the most favourable choice for the patch size and the minimum number of secondary images needed, the experiment is repeated choosing the images from acquisition passes 0 to 7 to be the reference images, and choosing different number of secondary images for each experiment. This test allows to show the robustness of the proposed framework with respect to the selection of the reference image.

The behaviour of the curves obtained is consistent with the first experiment, where the reference image was the one from acquisition pass 1 and the rest acquisition images were chosen to be the secondaries. The RMSE decreases as the patch size increases until the optimal patch size is reached, increasing again thereafter. Fig. 3.6 shows the evolution of the RMSE, in degrees, versus the patch size when the reference is the image from pass 1 with three different number of secondary configurations. The RMSE becomes smaller as the number of secondary grows, the maximum values going from 2.95° when only one secondary is taken into account, to 2° when 6 secondaries are considered. Compared to the results shown in Fig. 3.3, the difference is quite noteworthy. The RMSE when 7 secondary are considered and the patch size is between $25 \times 25$ and $30 \times 30$, is under 0.5°. This trend has been proven to be uniform across experiments performed over the 8 acquisition passes.



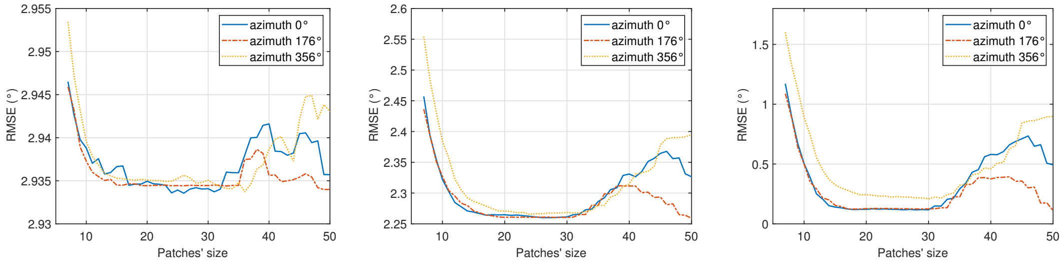
(a) Reference image pass 1, 1 secondary (b) Reference image pass 1, 4 secondaries (c) Reference image pass 1, 6 secondaries

Figure 3.6: RMSE (°) versus patch size for three configurations of Reference image (at pass 1) and different number of secondaries acquired at different azimuth angles, 0°, 176°, and 356°. a) shows the result for a single secondary image, b) the evolution with 4 secondary images, and c) the result for 6 secondary images. A total of 100 Monte Carlo runs is performed randomly selecting the rotation angles in the interval $[-2°, 2°]$.

It can be concluded that the number of secondaries cannot be reduced without compromising the performance of the algorithm, since the JCLS algorithm experiences a wide reduction in the number of equations as the number of secondary images decreases. From Fig. 3.6 it can be seen that the optimal patch size lies around 30x30 pixels, in accordance to the results shown in Fig. 3.3.

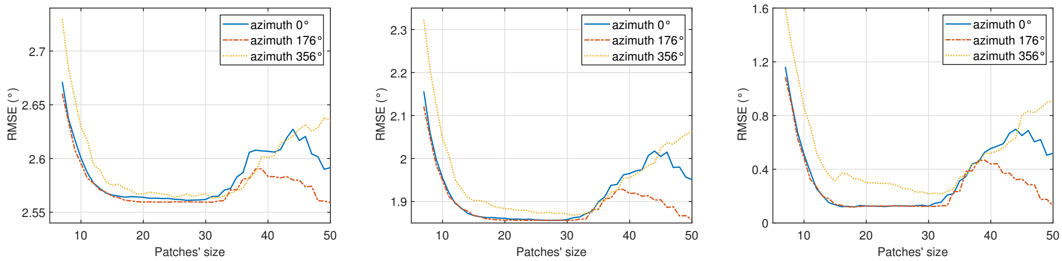
To show the consistency of the results of the experiments over different configurations of reference image and number of secondaries, three different examples are presented next. In Fig. 3.7, the case where the reference image is taken from pass number 4 is shown for 3 different numbers of secondary images. Results are presented for one, four, and seven secondary images. It can be seen that the RMSE (°) decreases as the patch number becomes bigger until it reaches the optimal patch size, around 30x30, it then increases again. The same trend as before is observed when taking into account the different number of secondary images selected. The RMSE decreases as the number of secondaries increases.

The second example can be seen in Fig. 3.8, where the image from pass 6 has been taken as the reference image. Again, the results follow a similar trend, obtaining lower RMSE (°) as the number of secondaries increases and reaching the optimal patch size at around 30x30 pixels. In this case, the number of secondary images is chosen to be three, five, and seven. For the last example, depicted in Fig. 3.9, the image from the last pass has been taken as a reference image and two, six, and eight secondaries have been chosen to illustrate the three different cases. Once again, the three graphs present a similar behavior compared to the examples seen before: as the number of secondaries grows, the RMSE decreases, and for the three cases, the optimal patch size is around 30x30 pixels, which is consistent with all the experiments performed.



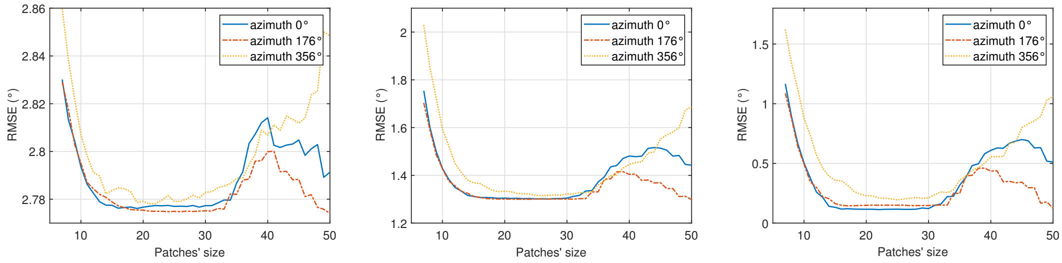
(a) Reference image pass 4, 1 secondary (b) Reference image pass 4, 4 secondaries (c) Reference image pass 4, 7 secondaries

Figure 3.7: RMSE (°) versus patch size for three configurations of Reference image (at pass 4) and different number of secondaries acquired at different azimuth angles, 0°, 176°, and 356°. a) shows the result for a single secondary image, b) the evolution with 4 secondary images, and c) the result for 7 secondary images. A total of 100 Monte Carlo runs is performed randomly selecting the rotation angles in the interval $[-2°, 2°]$.



(a) Reference image pass 6, 3 secondaries (b) Reference image pass 6, 5 secondaries (c) Reference image pass 6, 7 secondaries

Figure 3.8: RMSE (°) versus patch size for three configurations of Reference image (at pass 6) and different number of secondaries acquired at different azimuth angles, 0°, 176°, and 356°. a) shows the result for three secondary images, b) the evolution with 5 secondary images, and c) the result for 7 secondary images. A total of 100 Monte Carlo runs is performed randomly selecting the rotation angles in the interval $[-2°, 2°]$.



(a) Reference image pass 8, 2 secondaries (b) Reference image pass 8, 6 secondaries (c) Reference image pass 8, 7 secondaries

Figure 3.9: RMSE (°) versus patch size for three configurations of Reference image (at pass 8) and different number of secondaries acquired at different azimuth angles, 0°, 176°, and 356°. a) shows the result for two secondary images, b) the evolution with 6 secondary images, and c) the result for 7 secondary images. A total of 100 Monte Carlo runs is performed randomly selecting the rotation angles in the interval $[-2°, 2°]$.

## 3.6 Conclusion

The Cross-Cross-Correlation joint coregistration technique represents a significant advancement in the field of SAR image processing, offering improved accuracy and robustness in aligning multitemporal datasets. By simultaneously estimating the displacement fields of multiple secondary images with respect to a reference image, the method addresses several limitations inherent in conventional coregistration techniques.

One of the key strengths of the proposed approach is its ability to account for mutual displacements between secondary images during the estimation process. Traditional methods typically align each secondary image independently to the reference, which can lead to cumulative errors when multiple secondary images are involved. The proposed method mitigates these errors by considering the relative misalignments between secondary images, thus enhancing the overall accuracy and consistency of the registration.

SAR images are often characterised by low SNR due to speckle noise, atmospheric disturbances, and temporal decorrelation. The presented technique demonstrates robust performance in such conditions by leveraging higher-order correlations between cross-correlation functions. This redundancy in the estimation process helps reduce the impact of random noise, leading to more reliable displacement and rotation estimates. The inclusion of an outlier detection and refinement step further enhances robustness by eliminating erroneous measurements that could otherwise degrade the accuracy of the final solution.

The performance of the method is influenced by the choice of patch size and the number of secondary images used in the coregistration process. As demonstrated in the experimental results, smaller patch sizes may lead to higher errors due to insufficient information, while excessively large patches can introduce inaccuracies by encompassing multiple scatterers with different motion characteristics. An optimal patch size of 25×25 to 30×30 pixels was identified as a balance between these competing factors.

Additionally, the method benefits from a larger number of secondary images, as this increases the number of cross-cross-correlation measurements available for parameter estimation. However, it is important to note that an excessive number of secondary images can increase computational complexity, necessitating a careful balance between accuracy and efficiency.

The proposed joint coregistration framework provides more accurate and homogeneous alignment across multitemporal SAR stacks, even under challenging conditions typical of maritime surveillance. By robustly handling low SNR, temporal decorrelation, and rotational/geometric mismatches that appear in sea-clutter-dominated scenes with varying acquisition geometries, it establishes a reliable foundation for downstream maritime monitoring tasks such as vessel tracking, change detection of coastal infrastructures, and fusion of detections across multiple passes and sensors. However, the lack of temporally persistent keypoints in open-sea environments limits the applicability of the proposed approach, as reliable cross-correlation-based estimation requires stable reference points. Consequently, the method is more suitable for scenarios where consistent scatterers are present, such as coastal regions and harbour areas, where either multitemporal images or subapertures can count on fixed scatterers. Moreover, the proposed JCLS framework is derived under the assumption that the underlying scatterers correspond to quasi-static physical structures observed across the stack of SAR images. In this context, the estimated transformation parameters represent a global rigid motion between acquisitions, valid for stationary scenes such as urban infrastructures, coastal facilities, or persistent maritime installations. As a result, moving targets are not explicitely tracked within the current formulation but are instead treated as non-coherent contributions that are rejected during estimation.

The accuracy gains provided by the algorithm make it a compelling choice for applications where precision is critical, maritime surveillance in coastal areas, InSAR processing, earthquake monitoring, and urban change detection.

# Chapter 4

# Maritime Target Classification from SLC SAR Data Based on Spectral Profiles and Invariant Features

In this chapter, a novel approach for ATR in SAR imaging is presented, with an emphasis on the extraction and application of spectral information present in SAR data. The goal of the presented method is to improve the capacity to differentiate between different types of ships, with a focus on recognising and categorising ships operating in the open ocean. Traditional target recognition techniques in maritime surveillance frequently depend on observable physical characteristics, including the vessel's length, width, or general shape. Although these traits are helpful, they often overlap across many ship categories, including fishing boats, cargo ships, and naval ships, making precise classification difficult. In this chapter, their distinctive spectral profiles, generated by their engines, are exploited for accurate classification. The presented methodology focuses on the development of a binary classifier and a multiclass classifier that exploit spectral information and Krawtchouk moments derived from SLC SAR images.

## 4.1 Structure of the Chapter

The outline of this chapter is structured as follows: Section 4.2 introduces the methodology for the generation of spectral profiles from SAR SLC data, to be used for target characterisation. Section 4.3 introduces the use of the Krawtchouk moments to enhance the performance of the classification, and Section 4.4 offers an overview of the dataset used for the experiments. In Section 4.5, the classification model, including the architecture of the network, training parameters, and details on the implementation is discussed. In Section 4.6, the binary classification based on the spectral profiles is proposed, along with the dataset used and the results obtained, while in Section 4.7, the combination of spectral features and Krawtchouk moments for binary classification is presented. In Section 4.7.4, the multiclass classification using the spectral profiles is explained, along with the dataset and the results obtained, and in Section 4.8 and 4.9, an overview and a discussion of the proposed methodology can be found.

## 4.2 Spectral profile features from SLC data

SLC images are the result of the compression and focusing, through a series of processing steps, of the raw packetised SAR data, known as Level-0 data. Once the raw I/Q data from the sensor is extracted, a 2D array of data samples can be formed, with fast time $\tau$ along one axis and slow time $\eta$ along the other. In order to get the final image, the 2D array needs to be focused along range and azimuth dimensions.

To extract the information from the SLC data before and after calibration, the steps shown in Fig. 4.1 are performed: first, an FFT is applied in the azimuth direction to defocus the image in this dimension to revert it to a state where it represents the Doppler histories of the scatterers (domain transformation). This effectively spreads out the phase history of the target within the SAR image, representing the target's temporal response in the azimuth direction. The same procedure is applied to the radiometrically calibrated set of images, in order to analyse the differences in performance and accuracy of the classification.

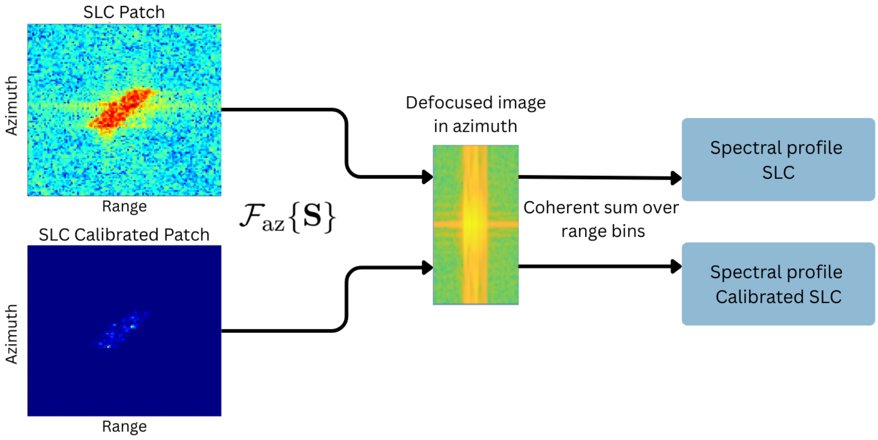


Figure 4.1: Process to obtain spectral profiles from SLC images and calibrated SLC images.

Given the SLC SAR-focused image $\mathbf{S} \in \mathbb{C}^{N_r \times N_a}$, where $N_r$ denotes range dimension and $N_a$ denotes azimuth, the defocusing of the SLC image is calculated as

$$\mathbf{I_{RD}} = \mathcal{F}_{\text{az}}\{\mathbf{S}\} \tag{4.1}$$

where $\mathcal{F}_{\text{az}}$ denotes the FFT operation and $\mathbf{I_{RD}}$ is the image defocused in azimuth direction. Second, once the representation of the Doppler histories of the scatterers is obtained, a coherent sum is performed over all range bins resulting in a combined signal for each azimuth time instance. By collapsing the range information, the azimuth-time variation is obtained, where changes in frequency due to vibrations manifest:

$$S_{\text{az}}(k) = \sum_{r=0}^{N_r-1} I_{RD}(r,k) \tag{4.2}$$

As a result, the entire range content of the image for each azimuth bin is compressed into one single value, forming the vector $S_{\text{az}}(k)$, which represents the azimuth spectral profile of the target. The spectral profile of the azimuthal content provides insight into the dominant azimuthal frequencies and the underlying scatterer structures of features within. To understand how the classifier performance varies when different aspects of the data are used, the following features are derived from the spectral profiles:

- Normalised and not normalised spectral features: Normalisation helps the model focus on the pattern of the signal rather than the differences in magnitude, useful when using samples with different scales. However, unnormalised features preserve magnitude information, which can be important when the differences in intensity levels are an inherent distinction in the input data. The z-score normalisation is applied;
- Normalised and not normalised logarithmic spectral features: The logarithmic scaled spectral values help in emphasising smaller variations and reduce the impact of outliers, compressing the range values. This approach is useful for capturing correlations, changes in orders of magnitude, and highlighting details in regions of the image where the intensity is lower or obscured by linear scales. Comparing results of linear and logarithmic representations allows for the identification of patterns across different scales and dynamic ranges of the data. SAR data is largely affected by a dynamic range resolution of up to 90 dB with a very asymmetric distribution [138];
- Normalised and not normalised autocorrelation features: Useful for capturing repetitive patterns, periodicity, and consistencies between samples. Normalising the autocorrelation helps the model focus on patterns and the underlying structure of the features, while the unnormalised version retains intensity information.

Given the spectral profile $S_{\mathrm{az}}(k)$, the extracted feature sets are defined as follows. For any feature vector $f(k)$, z-score normalisation is applied as

$$\tilde{f}(k) = \frac{f(k) - \mu_f}{\sigma_f} \tag{4.3}$$

where $\mu_f$ and $\sigma_f$ denote the mean and standard deviation of $f(k)$, respectively.

The considered feature categories are:

- **Spectral features:**

$$f_{spec}(k) = |S_{\mathrm{az}}(k)|, \qquad \tilde{f}_{spec}(k) = \frac{|S_{\mathrm{az}}(k)| - \mu_{spec}}{\sigma_{spec}} \tag{4.4}$$

- **Logarithmic spectral features:**

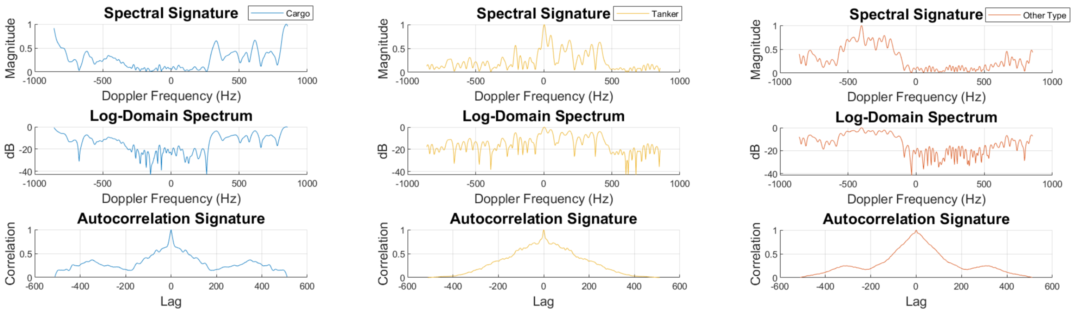
$$f_{log}(k) = \log(|S_{\mathrm{az}}(k)| + 1)\,, \qquad \tilde{f}_{log}(k) = \frac{f_{log}(k) - \mu_{log}}{\sigma_{log}} \tag{4.5}$$

- **Autocorrelation features:**

$$R(\ell) = \sum_{k} S_{\mathrm{az}}(k)\, S^{*}_{\mathrm{az}}(k-\ell), \qquad \tilde{R}(\ell) = \frac{R(\ell) - \mu_R}{\sigma_R} \tag{4.6}$$

Where $\ell$ denotes the discrete lag variable.

To explore the extent to which the spectral features capture engine-induced vibration, Fig. 4.2 compares representative spectra across vessel classes (Cargo, Tanker, and Other). Differences in the spectral content are observed, revealing the class-dependent micro-motion effects of each class. It should be noted that the extracted spectral profiles may reflect not only engine-induced vibrations, but also other motion sources, such as roll, pitch, propeller rotation, or rotating components. Consequently, the observed spectral content should be interpreted as a combined micro-motion signature of the target.



(a) Cargo (b) Tanker (c) Other

Figure 4.2: Spectral features derived from Cargo (a), Tanker (b), and Other (c), the logarithmic representation, and the autocorrelation for each class

## 4.3 Krawtchouk moments features from SLC data

Krawtchouk moments are of special interest for this study due to their inherent rotation and position-invariant properties. When targets are centred in the images, but their orientations are not aligned, Krawtchouk moments-based features can account for these variations removing the need for further alignment processing steps. To compute the Krawtchouk moments and obtain features of a consistent length, it is important that the input images are of the same dimensions. If images in the dataset are smaller than the target size, a frame of Gaussian noise is included, in order for the Krawtchouk moments not to take a more contrasting frame (for example all zeros) as a distinct feature that could introduce unwanted biases in the results. In Fig. 4.3, two examples of SLC images with Gaussian padding are shown, one for a Cargo ship, and one for a Tanker. The foundational elements of Krawtchouk moments are the discrete orthogonal Krawtchouk polynomials $k_n(x; p, N)$ [121,122], defined by their orthogonality property:

$$\sum_{x=0}^{N} w(x; p, N) k_n(x; p, N) k_q(x; p, N) = \rho(n; p, N)\delta_{nq}, \quad 0 \leqslant n, q \leqslant N \tag{4.7}$$

where $\delta_{nq}$ is the Kronecker delta function, defined as $\delta_{nq} = 1$ if $n = q$ and 0 otherwise, and the weight function is:

$$w(x; p, N) = \binom{N}{x} p^x (1-p)^{N-x} \tag{4.8}$$

and $k_n(x)$ are the Krawtchouk polynomials of order $n$, with $N$ as the image dimension and $p \in [0, 1]$ controlling polynomial location (typically $p = 0.5$ for centred features), and

$$\rho(n; p, N) = (-1)^n (\frac{1-p^n}{p}) \frac{n!}{(-N)_n} \tag{4.9}$$

The 2D Krawtchouk moments for an $N \times N$ image $f(x, y)$ are then computed as:

$$K_{nm} = [\rho(n; p, N)\rho(m; p, N)]^{-1} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} k_n(x; p_1, N-1) k_m(y; p_2, M-1) f(x, y) \tag{4.10}$$

for $n, m = 0, 1, \ldots, N$, where $k_n(x; p_1, N-1), k_m(y; p_2, M-1)$ are given by Eq. 4.7.

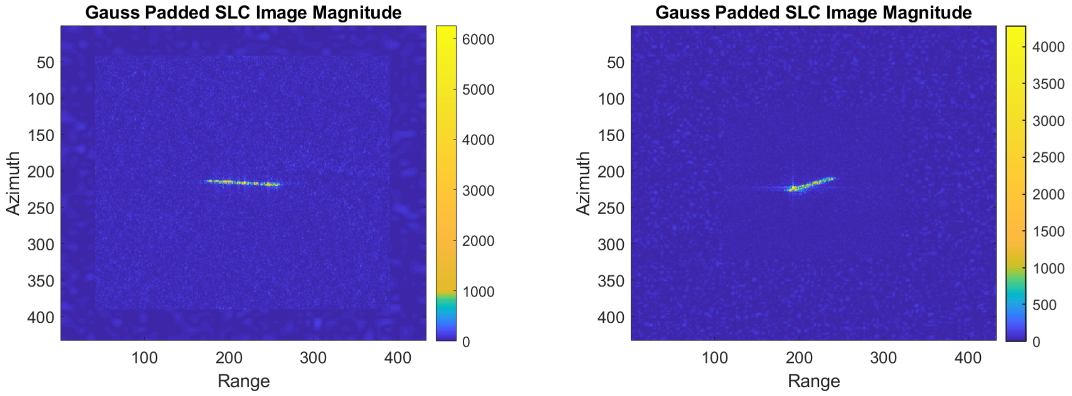
Krawtchouk moments of order 10 and 5 are computed for the experiments performed in this chapter, as a 2-D combination of the orders $n$ and $m$, where they represent the degrees of the polynomials in the $x$- and $y$-dimensions. The total number of moments to all possible combinations of $n$ and $m$, up to and including the targeted order, is given by:

$$\text{Total Moments} = (n_{max} + 1)^2 \tag{4.11}$$

which results in a flattened feature array of length 121 for order 10, and 36 for order 5, providing a standardised set of features for further processing. These moments capture directional texture information in the image, with the order controlling how fine or coarse the captured patterns are. Moments of orders 10 and 5 were selected to provide a balance between feature richness and computational efficiency while avoiding excessive sensitivity to noise and numerical instability, associated with high-order moments [159]. It is important to note that the Krawtchouk moments are computed directly from the original SLC image intensities, without prior normalisation of the image. This choice is made to preserve the physical meaning of the SAR backscatter, as normalising each image independently would remove absolute intensity differences that may carry discriminative information between ship classes. Instead, normalisation is applied at the feature level, where the resulting Krawtchouk moment vectors are standardised prior to classification (z-score). This ensures numerical stability and prevents features with larger magnitudes from dominating the learning process, while retaining the underlying physical characteristics of the SAR data.

## 4.4 Dataset Description

For the classification experiments conducted herein, the OpenSAR ship dataset [160] was used. This dataset provides Sentinel-1 SLC products (slant-range/azimuth oriented, complex I/Q format) in VV and HH polarisations, containing multiple ship chips per image with associated AIS data, which provides information about the geolocation, vessel attributes, unique identification numbers, and additional salient param-



(a) (b)

Figure 4.3: SLC images padded with a frame of Gaussian noise added where (a) is a Cargo ship, and (b) is a Tanker ship, showcasing the preprocessing technique for consistent input dimensions.

eters, among other attributes.

SLC products contain both original complex data and radiometrically calibrated versions, enabling direct comparison of preprocessing effects. The objective of calibration is to produce an image where the pixel values are physically meaningful backscatter measurements $\sigma^0$, enabling quantitative analysis and comparison across sensors, acquisition times, or modes. Sentinel-1 SLC products include look-up tables for the radiometric correction, accounting for range-dependent gain and absolute calibration constants. Nevertheless, raw SLC data are valid for qualitative analysis and feature extractions, and are used to investigate how spectral features and Krawtchouk moments respond to calibration and how different preprocessing ultimately affects classification and feature extraction results [161].

For this study, the VV polarisation images were used, as it has been demonstrated in previous studies that for SAR incidence angles between 30° and 45°, typical of Sentinel-1 IW acquisitions, ships tend to produce stronger backscatter in the VV polarisation channel [162, 163]. The dataset originally consisted of 1,750 Cargo chips and 489 Tanker chips. For binary classification, the data were balanced to 476 Cargo and 485 Tanker samples. Additionally, a third class, labelled “Other”, was constructed by merging several minor categories for multiclass experiments. Image dimensions vary up to 433×433 pixels and Gaussian noise padding was applied for Krawtchouk moment computation (see Section 4.3). Table 4.1 summarises the distribution of product types across classes. To mitigate class imbalance, the final experimental dataset contains 1,311 images, distributed as 476 Cargo, 485 Tanker, and 350 Other samples.

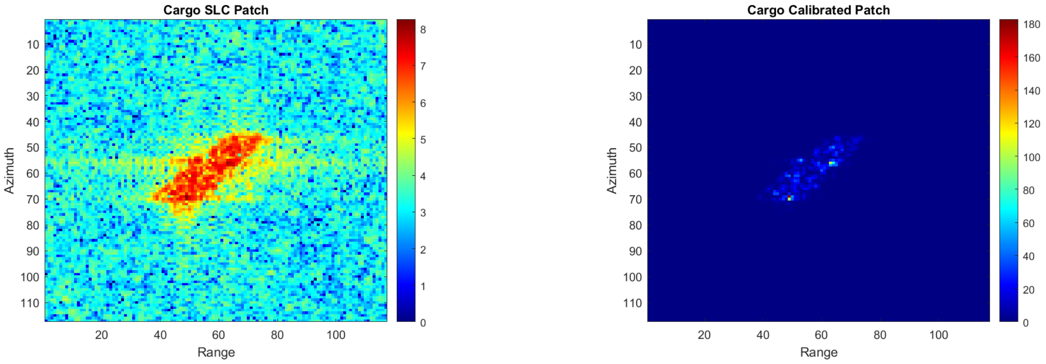
Table 4.1: Classification categories extracted from OpenSAR dataset: Cargo, Tanker, and Other.

| **Product type** | **Number of ships** | **Number of Ships for classification** |
|---|---|---|
| SLC and radiometric calibrated products | 1750 Cargos<br>489 Tankers<br>350 Other | 476 Cargos<br>485 Tankers<br>350 Other |

In Fig. 4.4, an example for each selected class can be seen in both formats, SLC and SLC calibrated images. Calibrated data provides a more accurate representation of the radar backscatter, which in turn can lead to more distinct and recognisable features for classification tasks. The reflectivity values are consistently aligned and scaled through calibration.

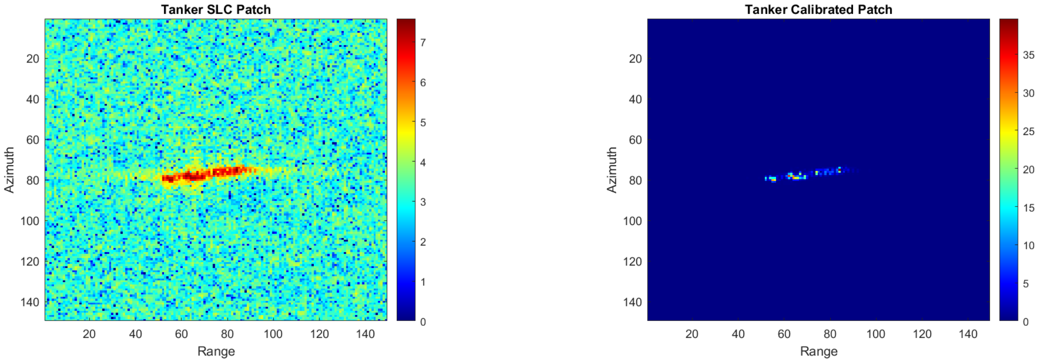
## 4.5 Classification Model

A combination of two RNNs is explored for classification, as they are widely used for the prediction of time series, containing hidden states and loops that store information to be used in future data and outputs. They are especially interesting for signal processing, as signals recorded over a period of time are naturally sequential. Raw features and processed data can be fed into the network to focus on specific features, such as time dependency or frequency components [164]. Hence, understanding the nature of the input data, the dimensionality, and the application is crucial. In this chapter, two distinct types of features are extracted (spectral and Krawtchouk moments) and classification is evaluated on both independently and in combination. Both networks share the same architecture, shown in Fig. 4.5, with the difference between them residing in



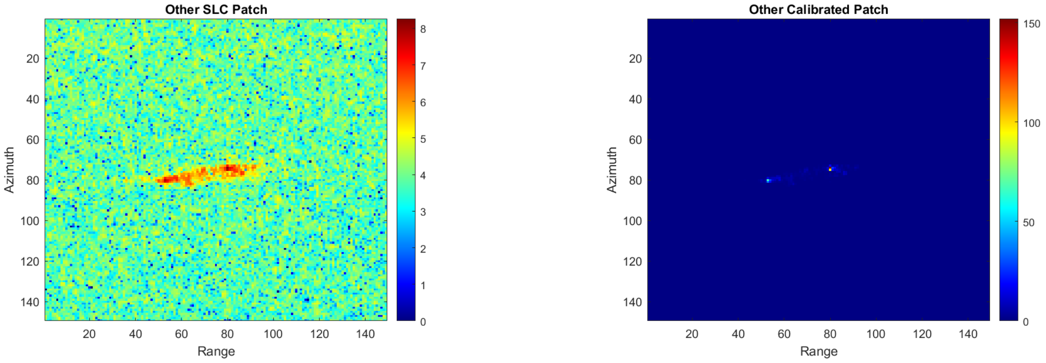
(a) Cargo SLC image

(b) Cargo Calibrated image



(c) Tanker SLC image

(d) Tanker Calibrated image



(e) Other SLC image

(f) Other Calibrated image

Figure 4.4: Comparison of ship data in original SLC and calibrated formats. (a), (c), and (e) show the SLC images before calibration for the Cargo, Tanker, and Other classes, respectively, while (b), (d), and (f) present the corresponding images after calibration.

the input layer.

The model is a composite of 16 layers such as: Bi-directional Long Short-Term Memory

(BiLSTM), Dropout, Batch Normalisation, Long Short-Term Memory (LSTM), Self-Attention, Fully Connected, and Softmax layers. Each layer plays a role in feature extraction, sequence pattern identification, and classification, so the network can effectively capture sequential dependencies and discriminative features. In the following, a breakdown of their functions is provided:

1. Sequence/Feature input Layer: Spectral features are fed into the network as sequential data, given their resemblance to time series signals. In contrast, features derived from Krawtchouk moments are processed through a feature input layer, which is designed for non-sequential data or fixed-size vectors representing specific characteristics of the data. This type of layer can be found in standard feed-forward neural networks or the initial layers of CNNs. For example, it can accept a fixed-length feature vector that encapsulates attributes of an image or a data sample.

2. BiLSTM Layer: The first layer is a BiLSTM layer with 128 hidden units (number of neurons in each LSTM) and 256 hidden states (the concatenation of forward and backward LSTM outputs) that processes the input sequences from both forward and backward directions. By considering past (forward pass) and future input features (backward pass) for each time step, it enhances pattern recognition, capturing dependencies that a standard LSTM might fail to detect [165];

3. Dropout Layer: Immediately following the BiLSTM, a Dropout layer with drop out probability of 0.5 is used to prevent overfitting by randomly setting a fraction of input units to 0 during training. A dropout layer is also added after every LSTM layer, with the same dropout probability and for the same purpose;

4. Batch Normalisation Layer: To speed up training and improve generalisation, this layer is added, which normalises the activations of the nodes, ensuring a mean close to 0 and a standard deviation close to 1;

5. LSTM Layer: The LSTM layer, with 128 hidden units and 128 hidden states, is used to remember patterns over long sequences, crucial for understanding the

context in sequence data [166]. The architecture includes 3 LSTM layers in total, enabling the model to extract deeper patterns and refine the representation of the input data;

6. Self-Attention Layer: This layer allows the model to weigh the importance of different parts of the input spectral vector differently, providing a form of dynamic weighting. The parameters of this layer are 12 heads, meaning that the input is processed in 12 parallel subspaces, each learning to capture different patterns or relationships. Each head uses 96 key channels, which define the dimensionality of the keys, values, and query matrices. Each set produces different weightings and captures different aspects of the input data. Key channels refer to the dimension of the keys, value, and queries in each head [92]. This layer enables the model to assign varying levels of importance to different parts of the feature vector based on their relevance to the classification task;

7. Fully Connected Layer: A dense layer is added to aggregate the extracted features and begin the classification process [167];

8. Softmax Layer: Converts the output of the previous layer into probability scores for each class [168];

9. Class Output: The final layer provides the classification results based on the highest probability score obtained from the Softmax layer [169].

### 4.5.1 Implementation

For the implementation of both networks, the data was randomly partitioned into training (70%), validation (5%), and testing (25%) sets. To ensure a robust and statistically meaningful evaluation, a Monte Carlo validation strategy was adopted, where the data splitting process is repeated over 50 independent random realisations. This allows the assessment of the variability of the classification performance and reduces the dependency on a single data partition. An adaptive moment estimation (Adam) optimiser with the following parameters was chosen:

- Maximum number of epochs: The number of full passes of the training data, selected to be 30 to provide the model sufficient opportunity to learn patterns while limiting excessive iterations that could increase the risk of overfitting. A batch size of 22 was chosen to balance computational efficiency and gradient estimation stability;
- Shuffle: The training and validation data were selected to be shuffled once before training, to ensure randomisation of data order, preventing the model from learning misleading patterns;
- Initial learning rate (0.001): Standard starting point for Adam, to provide stable and efficient convergence.

To evaluate the networks' performances, the accuracy, precision, recall, and F1-score parameters are used, further discussed in Section 4.6.

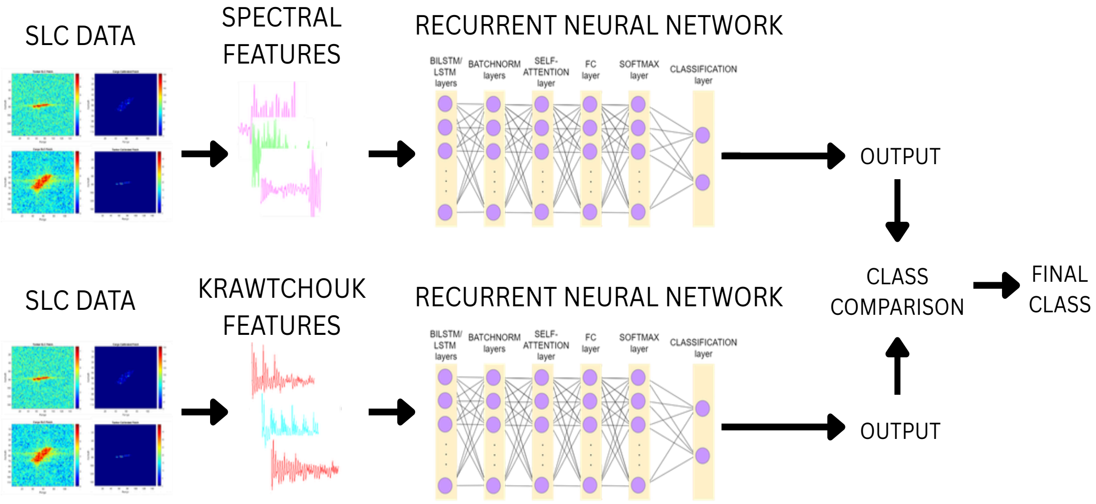


Figure 4.5: Illustration of the RNN architectures used for classification, where the network processes vibrational profile feature vectors as sequential input, Krawtchouk moment feature vectors as non-sequential input, and ultimately combines the independent outputs of both RNNs for final classification.

## 4.6 Binary classification of SLC SAR spectral profiles

For binary classification, the main two classes in the dataset (Cargo and Tanker) are selected and the different feature arrays were extracted, as detailed in Section 4.2, to enhance model performance and allow for a comparative analysis:

- Normalised features: Z-score normalisation is used to standardise feature values, adjusting the mean and standard deviation of the feature arrays. This ensures comparability and reduces the effects of extreme values;
- Auto-correlation features: Auto-correlation measures the correlation of a signal with a delayed version of itself. This processing was applied to normalised and non-normalised feature arrays to help capture repetitive patterns in the signal to distinguish between different patterns that arise in different types of ships in SAR data;
- Logarithmic transformation features: This processing step mitigates the skewness inherent in SAR data distributions. By adjusting the distributions toward symmetry, the transformation aims to enhance the suitability of the data for ML models.

The proposed model was evaluated on the spectral feature vectors extracted from the SLC images and the radiometric-calibrated SLC images. As previously indicated in this chapter, the feature vector was generated by applying an FFT in the azimuth direction of the images and performing a coherent sum of all range values per azimuth bin, yielding the spectral profile arrays. The number of points (N) for the FFT is chosen based on the biggest image patch in the set, which is 433x433:

$$N = 2^{nextpow2(433)} \tag{4.12}$$

resulting in N=512.

The performance of the RNN is evaluated using the following metrics, where TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively:

- Accuracy: A widely used metric that measures the proportion of correct predictions out of the total predictions made by the model. It is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.13}$$

- Recall: This metric evaluates how many of the actual positive samples are correctly identified by the model. It is particularly useful when minimising false negatives is crucial. It is given by:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.14}$$

- Precision: It assesses the reliability of positive predictions by determining the proportion of correctly predicted positive samples out of all predicted positives. It is especially relevant when false positives have a high cost. The formula is:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.15}$$

- F-1 score: A function of recall and precision, used to balance the results of these two metrics, defined as:

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN} \tag{4.16}$$

Calculating these four metrics, a more insightful view of the classifier performance can be obtained [170].

The results are assessed by comparing raw SLC data with radiometrically calibrated SLC data to determine which format offers better classification results. Additionally, the study evaluates the classifier's performance using both normalised and non-normalised SAR SLC images for both cases. The following subsections present the results of the binary classification between Cargo and Tanker, Cargo and Other, and Tanker and Other

### 4.6.1 Classification Results: Cargo and Tanker

To evaluate the classifier's performance, feature arrays derived from both raw and radiometrically calibrated SLC images were independently input to the network, with additional analysis of the effect of image normalisation. The results are shown in Fig. 4.6 for raw SLC images and Fig. 4.7 for calibrated SLC images. In each case, classification performance is compared between non-normalised and normalised inputs, represented by green and blue bars, respectively. As indicated on the horizontal axis, multiple feature extraction techniques were evaluated, as described in Section 4.2. The availability of both raw and calibrated SLC products in the OpenSAR dataset enables a direct comparison of classification performance under different preprocessing conditions.

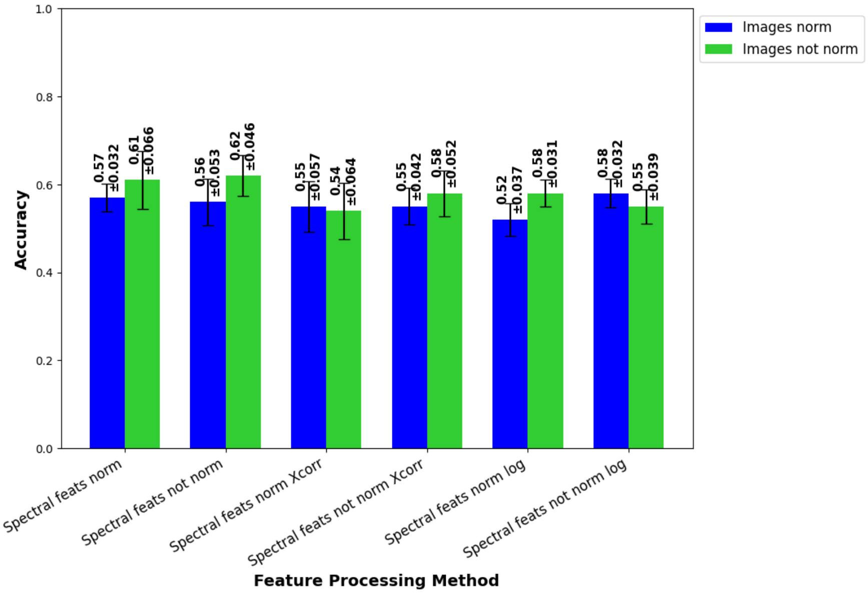


Figure 4.6: Comparison of the network performance for six different feature configurations and for the case where SLC images are not normalised (green bars) and normalised (blue bars) before processing and feature extraction Cargos vs. Tankers.

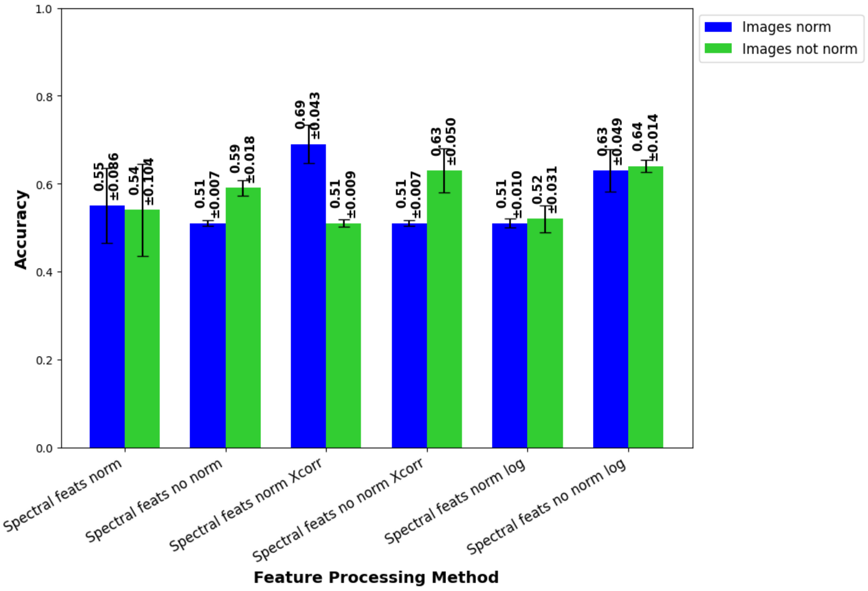


Figure 4.7: Comparison of the network performance for six different feature configurations and for the case where SLC radiometric calibrated images are not normalised (green bars) and where the images are normalised (blue bars) before processing and features extraction Cargos vs. Tankers.

The results from Fig. 4.6 and Fig. 4.7, compare the accuracy and standard deviation of the binary classifier, highlighting several key observations:

- Auto-Correlation Features: The highest accuracy is achieved when the normalised feature array is correlated with itself (auto-correlation) on radiometrically calibrated data (Configuration 3 on Calibrated Data), with an accuracy of 69.19%, precision of 66.59%, recall of 75.34%, and F-1 score of 70.69%, which can be seen in bold in Table 4.2. Furthermore, it shows a relatively low standard deviation of 0.04289. This suggests that the calibrated data, which has been corrected for sensor distortions and noise, provides a more consistent signal for the classifier to work with. The consistency of the calibrated backscatter signal allows the auto-correlation operation to more effectively capture the relevant information that differentiates between the two ship types.
- Normalisation: Z-score normalisation was employed, however, non-normalised features generally resulted in higher accuracy. Notably, when using SLC images

for the classification, the highest accuracy result (62.22%) is achieved with a combination of non-normalised images and their corresponding non-normalised feature vectors. However, the best classification performance is observed when the model processes features derived from the auto-correlation of normalised spectral vectors obtained from normalised calibrated SLC images. Normalisation reduces signal variability and aligns the data within a consistent range, which enhance machine learning model performance, while non-normalised features may retain useful raw information.

- Radiometric Calibration: The classifiers using radiometrically calibrated data tend to show higher accuracies and lower standard deviations compared to SLC raw data. This reinforces the importance of calibration, which improves the physical interpretability of SAR data and enhances the classifier's ability to discern between the two classes.
- Logarithmic Transformation: The performance of the classifier is not improved by using the logarithmic transformation, however it drastically reduced the skewness of the Cargo ships data from an average of 1.17 to 0.4015, and for Tankers, from 5.41 to 0.4759, indicating a shift towards symmetry for both datasets.

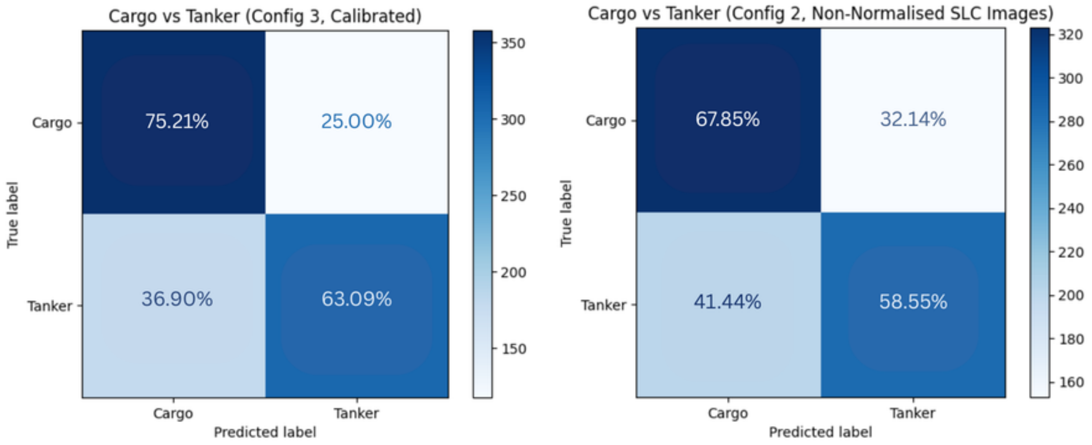
SAR systems are sensitive to the dielectric properties and geometries of objects, and preprocessing steps like calibration and normalisation prove to be useful in this case to better extract and distinguish the relevant features.

Table 4.2: Evaluation metrics comparisons for SLC data and SLC radiometric calibrated data where 1 is for spectral features normalised, 2 for spectral features not normalised, 3 for the auto-correlation of the normalised features, 4 for the auto-correlation of the not normalised features, and 5 and 6 for the logarithms of the normalised and not normalised features respectively.

| **Metric** | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|
| **Comparison Cargo vs Tanker SLC** | | | | | | |
| **Precision** | 0.6332 | 0.6154 | 0.5749 | 0.6131 | 0.5813 | 0.5748 |
| **Recall** | 0.5534 | 0.6795 | 0.3260 | 0.4603 | 0.5973 | 0.4630 |
| **F1-score** | 0.5906 | 0.6458 | 0.4161 | 0.5258 | 0.5892 | 0.5129 |
| **Accuracy** | 0.6111 | **0.6222** | 0.5361 | 0.5792 | 0.5778 | 0.5542 |
| **Comparison Cargo vs Tanker Norm Images SLC** | | | | | | |
| **Precision** | 0.5749 | 0.5439 | 0.5724 | 0.5365 | 0.5170 | 0.5524 |
| **Recall** | 0.5890 | 0.8493 | 0.4658 | 0.7644 | 0.6247 | 0.8959 |
| **F1-score** | 0.5819 | 0.6631 | 0.5136 | 0.6305 | 0.5658 | **0.6834** |
| **Accuracy** | 0.5708 | 0.5625 | 0.5528 | 0.5458 | 0.5139 | 0.5792 |
| **Comparison Cargo vs Tanker Norm Images Calibrated** | | | | | | |
| **Precision** | 0.5744 | 0.4932 | 0.6659 | 0.4932 | 0.4979 | 0.5906 |
| **Recall** | 0.4548 | 0.4000 | 0.7534 | 0.4000 | 0.6521 | 0.8219 |
| **F1-score** | 0.5076 | 0.4418 | **0.7069** | 0.4418 | 0.5647 | 0.6873 |
| **Accuracy** | 0.5649 | 0.5014 | **0.6919** | 0.5014 | 0.5040 | 0.6311 |
| **Comparison Cargo vs Tanker Calibrated** | | | | | | |
| **Precision** | 0.5769 | 0.5627 | 0.4898 | 0.7056 | 0.5141 | 0.6253 |
| **Recall** | 0.2877 | 0.7863 | 0.1973 | 0.4137 | 0.6000 | 0.6630 |
| **F1-score** | 0.3839 | 0.6560 | 0.2813 | 0.5216 | 0.5537 | 0.6436 |
| **Accuracy** | 0.5446 | 0.5932 | 0.5027 | 0.6257 | 0.5230 | 0.6378 |

To provide a deeper insight into the classifier's performance for the Cargo vs. Tanker classification task, confusion matrices are included in Fig. 4.8. These matrices complement the evaluation metrics shown in Table 4.2 by highlighting the distribution of true and false predictions across the two classes.

For Configuration 3 on Calibrated Data, which achieved the highest accuracy of 69.19%, the confusion matrix (Fig. 4.8 (a)) is obtained by averaging the results over 50 Monte Carlo realisations, and it shows that 75.21% of the Cargo samples were correctly identified, while the 36.90% were misclassified as Tankers. Similarly, the classifier correctly identified 63.09% of the Tanker samples, misclassifying 25.00% as Cargo. This indicates that while the model performs well in distinguishing these classes overall, it struggles more with Tanker samples due to overlapping characteristics. On the other hand, the confusion matrix for Configuration 2 on non-normalised SLC Images (Fig. 4.8 (b)), a similar averaging procedure over the Monte Carlo realisations shows a different trend, with a higher false positive rate for Cargo ships. This suggests that the lack of calibration and normalisation introduces noise that confuses the classifier. Furthermore, error analysis reveals that the features extracted from these unprocessed images retain raw but inconsistent patterns, producing worse misclassification.



(a) Calibrated data

(b) Non-normalised SLC images

Figure 4.8: Confusion matrices comparing classification performance under different data preprocessing configurations.

### 4.6.2 Classification Results: Tanker and Other

The results of the binary classification of Tanker and Other are presented in Fig. 4.9 for raw SLC images, and in Fig. 4.10 for calibrated SLC images. Each graph compares classification performance for non-normalised and normalised images, respectively represented by green and blue bars.

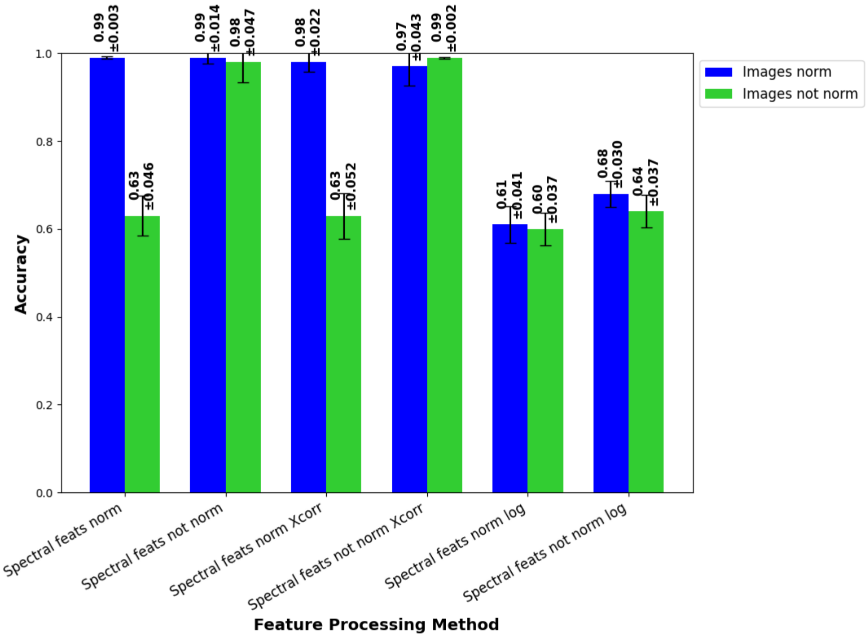


Figure 4.9: Comparison of the network performance for six different feature configurations and for the case where SLC images are not normalised (green bars) and normalised (blue bars) before processing and feature extraction Tanker vs. Other.

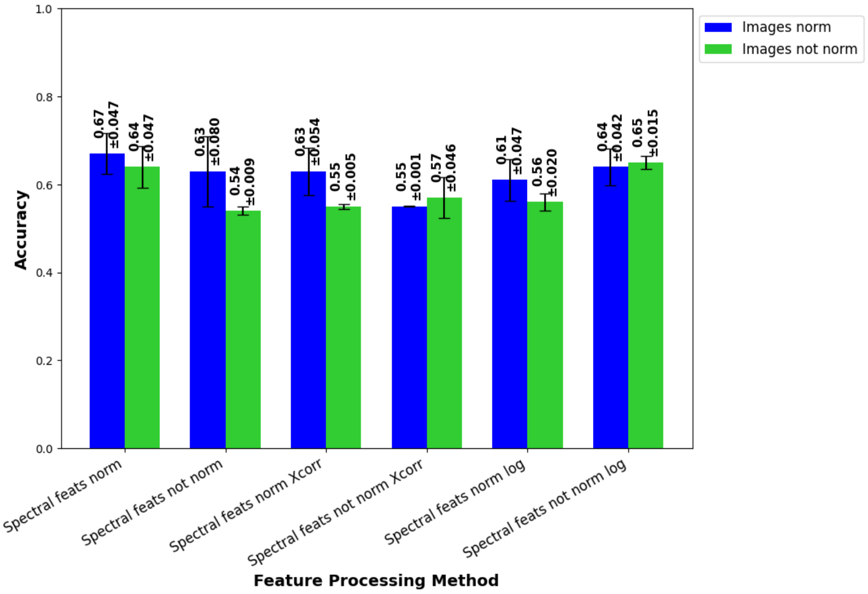


Figure 4.10: Comparison of the network performance for six different feature configurations and for the case where SLC radiometric calibrated images are not normalised (green bars) and where the images are normalised (blue bars) before processing and features extraction Tankers vs. Other.

The results from Fig. 4.9 and Fig. 4.10, further summarised in Table 4.3, reveal key differences in classification accuracy and variability, leading to several important observations:

- Auto-Correlation Features (Xcorr): The use of auto-correlation features generally degrades classification performance. In particular, when applying auto-correlation to normalised feature vectors derived from non-normalised SLC images, the model exhibits extremely poor recall, correctly identifying only 23.27% of actual positive cases. This indicates a failure to effectively separate the classes. Despite achieving a relatively high accuracy of 63.12%, this suggests that the classifier is biased toward one class while struggling to correctly identify the other.

  On the other hand, when using non-normalised auto-correlation features from non-normalised SLC images, the classifier performs very well, achieving an accuracy of 99.92%, the highest in this category. Similarly, both variations of auto-correlation applied to normalised SLC images also produce strong results, with accuracies of 98.72% and 97.60%.

However, when applied to calibrated data, auto-correlation features lead to significantly lower accuracy, with values between 55% and 63%. These results indicate that while auto-correlation can be effective in specific cases, such as with non-normalised SLC data, it remains unreliable overall for classifying Tankers vs. Other, weakening the classifier's effectiveness;

- Normalisation: The normalisation of the spectral features produces the best results for the normalised SLC images and the normalised calibrated images. The classification results for the normalised spectral features of the normalised SLC images and the auto-correlation of the not-normalised features seem to yield a very similar result, highlighted in table 4.3. However, the standard deviation for the latter is 0.001 better, as can it be seen in Fig. 4.9.

- Radiometric Calibration: calibration consistently reduces classification accuracy across all feature configurations, not exceeding 66% in any case, whereas non-calibrated results surpass 99%. These results indicate that calibration complicates the classification of Tankers vs. Others, making it more challenging for the model to distinguish between the two classes;

- Logarithmic Transformation: Logarithmic Transformation: In most cases, applying a logarithmic transformation to spectral features results in a decline in classification performance. This trend is particularly evident in SLC and normalised SLC images, where accuracy drops significantly compared to other feature configurations. For instance, in SLC data, applying logarithmic transformation to normalised and non-normalised features (columns 5 and 6) results in lower accuracy (59.60% and 64.08%) compared to the highest accuracy in this category (99.92% for auto-correlation of non-normalised features). Similarly, in normalised SLC images, log-transformed features yield lower accuracy (61.36% and 67.76%) compared to the best-performing configurations.

  However, for calibrated data, logarithmic transformation helps mitigate the performance loss introduced by calibration. Accuracy increases from 54.63% (non-normalised features without logarithm) to 65.45% (log-transformed non-normalised features, column 6), marking the highest classification accuracy for calibrated data. This suggests that logarithmic scaling helps stabilise the feature distribution, reducing distortions introduced by radiometric calibration and improving separability between classes.

Thus, while logarithmic transformation generally degrades the performance for raw SLC and normalised SLC images, it proves beneficial for calibrated data, balancing the negative impact of calibration on classification accuracy.

Table 4.3: Evaluation metrics comparisons for SLC data and SLC radiometric calibrated data where 1 is for spectral features normalised, 2 for spectral features not normalised, 3 for the auto-correlation of the normalised features, 4 for the auto-correlation of the not normalised features, and 5 and 6 for the logarithms of the normalised and not normalised features respectively.

| **Metric** | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|
| **Comparison Tanker vs Other SLC** | | | | | | |
| **Precision** | 0.5775 | 0.9940 | 0.6612 | 1.0000 | 0.5263 | 0.5989 |
| **Recall** | 0.4442 | 0.9635 | 0.2327 | 0.9981 | 0.2885 | 0.4135 |
| **F1-score** | 0.5022 | 0.9785 | 0.3442 | 0.9990 | 0.3727 | 0.4892 |
| **Accuracy** | 0.6336 | 0.9824 | 0.6312 | **0.9992** | 0.5960 | 0.6408 |
| **Comparison Tanker vs Other Norm Images SLC** | | | | | | |
| **Precision** | 1.0000 | 0.9867 | 0.9941 | 0.9471 | 0.5837 | 0.6364 |
| **Recall** | 0.9981 | 0.9962 | 0.9750 | 0.9981 | 0.2481 | 0.5250 |
| **F1-score** | 0.9990 | 0.9914 | 0.9845 | 0.9719 | 0.3482 | 0.5753 |
| **Accuracy** | **0.9992** | 0.9928 | 0.9872 | 0.9760 | 0.6136 | 0.6776 |
| **Comparison Tanker vs Other Norm Images Calibrated** | | | | | | |
| **Precision** | 0.6465 | 0.6087 | 0.6078 | 0.5489 | 0.5948 | 0.6623 |
| **Recall** | 0.8616 | 0.9014 | 0.9192 | 1.0000 | 0.8808 | 0.6849 |
| **F1-score** | 0.7387 | 0.7267 | 0.7317 | 0.7087 | 0.7101 | 0.6734 |
| **Accuracy** | **0.6654** | 0.6278 | 0.6301 | 0.5489 | 0.6053 | 0.6428 |
| **Comparison Tanker vs Other Calibrated** | | | | | | |
| **Precision** | 0.6225 | 0.5459 | 0.5460 | 0.5607 | 0.5572 | 0.6328 |
| **Recall** | 0.8630 | 0.9945 | 1.0000 | 0.9493 | 0.9342 | 0.8712 |
| **F1-score** | 0.7233 | 0.7049 | 0.7063 | 0.7050 | 0.6981 | 0.7331 |
| **Accuracy** | 0.6403 | 0.5463 | 0.5470 | 0.5672 | 0.5597 | **0.6545** |

For the highest accuracy and robustness in Tanker vs. Other binary classification, the best-performing feature sets are:

1. Normalised spectral features extracted from normalised raw SLC images;
2. Non-normalised auto-correlation applied to non-normalised raw SLC images.

These configurations provide the most reliable classification performance, regarding accuracy and stability.

### 4.6.3 Classification Results: Cargo and Other

To asses the classifier's performance for the classes Cargo and Other, feature arrays extracted from SLC raw images and calibrated SLC images are fed into the network, evaluating the impact of image normalisation. The results are presented in Fig. 4.11 for raw SLC images and in Fig. 4.12 for calibrated SLC images. Each graph compares classification performance for non-normalised and normalised images, represented by green and blue bars respectively. The evaluation includes a range of feature extraction techniques used on raw SLC data and calibrated SLC data, making a direct assessment of the network's classification efficacy under different input conditions possible.

The results presented in Fig. 4.11 and Fig. 4.12 compare the accuracy and standard deviation of the binary classifier, highlighting key observations:

- Auto-Correlation Features: When auto-correlation is performed on the features for classification, it can be seen that the accuracy worsens across all configurations. However, when combined with calibrated non-normalised images, it achieves better results within this specific setting. Among all configurations, the best accuracy using auto-correlation is obtained with raw SLC non-normalised images, reaching 79.62%. For calibrated images, auto-correlation still performs relatively well, achieving an accuracy of 78.43%, suggesting that while calibration affects classification, auto-correlation can retain useful discriminative features in this setting.
- Normalisation: Normalisation generally reduces classification accuracy across most settings. For raw SLC images, the accuracy for normalised spectral fea-

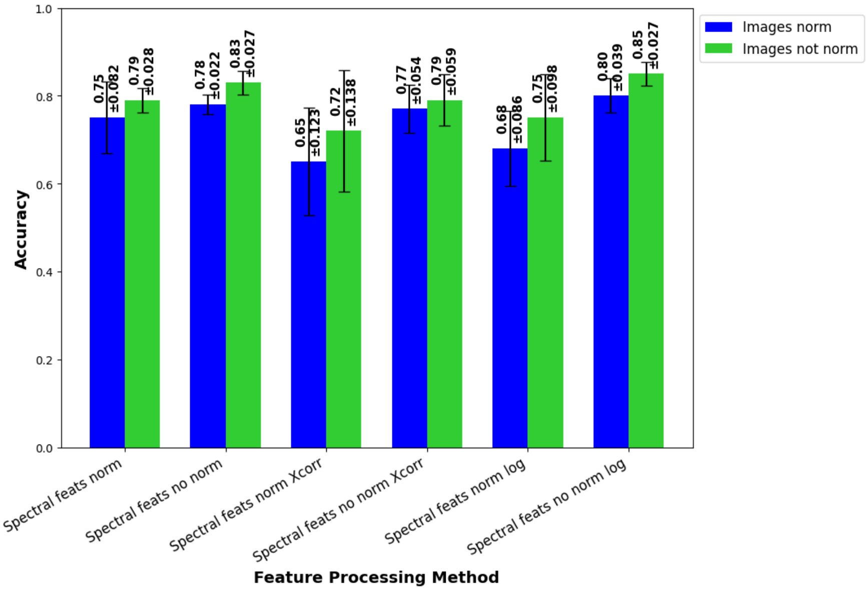


Figure 4.11: Comparison of the network performance for six different feature configurations and for the case where SLC images are not normalised (green bars) and normalised (blue bars) before processing and feature extraction Cargo vs Other.

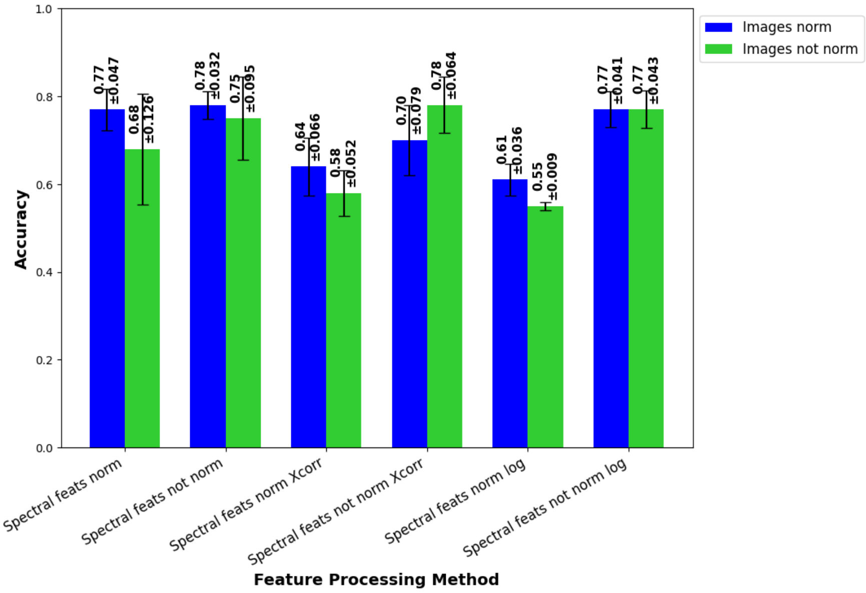


Figure 4.12: Comparison of the network performance for six different feature configurations and for the case where SLC radiometric calibrated images are not normalised (green bars) and where the images are normalised (blue bars) before processing and features extraction Cargo vs Other.

tures is 79.43%, which is lower than the 82.83% achieved with non-normalised features. Similarly, for normalised SLC images, non-normalised features yield 78.11% accuracy, outperforming the 75.28% obtained with normalised features. This suggests that normalisation may remove useful variability in the spectral features.

For auto-correlation features, normalised feature vectors result in lower accuracy across all cases. In raw SLC images, the auto-correlation of normalised features reaches 71.89% accuracy, whereas non-normalised auto-correlation improves performance to 79.62%. A similar trend is observed in normalised raw SLC images, where normalised auto-correlation achieves 64.53%, compared to 77.26% for non-normalised auto-correlation.

In radiometrically calibrated data, the impact of normalisation is less consistent. While normalised features achieve a slightly lower accuracy compared to non-normalised features, the difference is smaller than in SLC data. On the other hand, applying auto-correlation, normalisation significantly reduces performance, with an accuracy drop from 78.43% for non-normalised features to 57.57% for normalised features.

- Radiometric Calibration: Calibration affects negatively the performance of the classification, reducing the overall accuracy across all configurations. Although the performance worsens, it still shows accuracies up to 78.43%, which is not dramatically low when compared to the best accuracy for the non-calibrated image settings (85.28%).

- Logarithmic Transformation: logarithmic transformation improves accuracy for Cargo vs. Other when the raw SLC images are used for extracting the feature vectors, with the highest accuracy obtained for both normalised and non-normalised images (85.28% and 80.09% respectively). However, after calibration, the logarithm transformation harms the accuracy, suggesting that log-based scaling amplifies distortions in calibrated data for this case scenario.

The most notable drop in accuracy occurs when normalisation is applied to the logarithmic transformation of non-normalised calibrated images. In this case, accuracy is only 55.05%, highlighting that normalisation can further amplify performance degradation introduced by logarithmic scaling. Overall, these results indicate that while normalisation can help standardise data distribution, it may also remove crucial discriminative information, particularly in raw SLC and auto-correlation-based features.

Table 4.4: Evaluation metrics comparisons for SLC data and SLC radiometric calibrated data where 1 is for spectral features normalised, 2 for spectral features not normalised, 3 for the auto-correlation of the normalised features, 4 for the auto-correlation of the not normalised features, and 5 and 6 for the logarithms of the normalised and not normalised features respectively.

| **Metric** | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|
| **Comparison Cargo vs Other SLC** | | | | | | |
| **Precision** | 0.8356 | 0.8340 | 0.6796 | 0.7754 | 0.7561 | 0.9252 |
| **Recall** | 0.7231 | 0.8115 | 0.8077 | 0.8231 | 0.7154 | 0.7615 |
| **F1-score** | 0.7753 | 0.8226 | 0.7381 | 0.7985 | 0.7352 | 0.8354 |
| **Accuracy** | 0.7943 | 0.8283 | 0.7189 | 0.7962 | 0.7472 | **0.8528** |
| **Comparison Cargo vs Other Norm Images SLC** | | | | | | |
| **Precision** | 0.7276 | 0.8182 | 0.6090 | 0.7591 | 0.6638 | 0.8207 |
| **Recall** | 0.7808 | 0.7269 | 0.7308 | 0.8000 | 0.6000 | 0.7923 |
| **F1-score** | 0.7808 | 0.7699 | 0.6643 | 0.7790 | 0.6303 | 0.8063 |
| **Accuracy** | 0.7528 | 0.7811 | 0.6453 | 0.7726 | 0.6830 | **0.8009** |
| **Comparison Cargo vs Other Norm Images Calibrated** | | | | | | |
| **Precision** | 0.7626 | 0.7581 | 0.6485 | 0.8024 | 0.5872 | 0.7467 |
| **Recall** | 0.7067 | 0.7581 | 0.3567 | 0.4533 | 0.3817 | 0.7467 |
| **F1-score** | 0.7336 | 0.7473 | 0.4602 | 0.5793 | 0.4626 | 0.7467 |
| **Accuracy** | 0.7719 | **0.7785** | 0.6381 | 0.7074 | 0.6059 | 0.7748 |
| **Comparison Cargo vs Other Calibrated** | | | | | | |
| **Precision** | 0.6891 | 0.7391 | 0.6394 | 0.8045 | 0.4000 | 0.7797 |
| **Recall** | 0.5098 | 0.6749 | 0.1240 | 0.6858 | 0.0044 | 0.6885 |
| **F1-score** | 0.5861 | 0.7055 | 0.2078 | 0.7404 | 0.0086 | 0.7313 |
| **Accuracy** | 0.6770 | 0.7473 | 0.5757 | **0.7843** | 0.5505 | 0.7730 |

After the evaluation of multiple preprocessing strategies is carried out, including radiometric calibration, normalisation, autocorrelation, and logarithmic transformations, the results do not indicate a single consistently optimal configuration across all classification tasks. The impact of preprocessing is highly dependent on both the dataset and the specific class pair under consideration. For instance, while radiometric calibration improves performance in the Cargo vs. Tanker case, it degrades classification accuracy in the Tanker vs. Other and Cargo vs. Other scenarios. Similarly, normalisation does not systematically enhance performance and, in several cases, appears to reduce discriminative information contained in the spectral features. These variations suggest that the preprocessing steps affect the separability of the classes in a non-uniform manner.

Overall, the moderate classification performance observed in some cases highlights the intrinsic difficulty of the problem, likely due to overlapping spectral characteristics between vessel types and the complex nature of SAR backscatter. The results indicate that, although certain preprocessing and feature extraction combinations can improve performance under specific conditions, no single approach generalises across all scenarios.

## 4.7 Integration of Krawtchouk moments for classification

To investigate whether combining both classifiers improves performance, the outputs of the individual models and their combined predictions are compared (Fig. 4.5). Multiple feature configurations are evaluated, including normalised and non-normalised spectral features, logarithmic transformations, auto-correlation, and both calibrated and non-calibrated SLC images. For each experiment, the dataset is split into 80% for training, 15% for testing, and 5% for validation, adopting a Monte Carlo validation strategy where the data is divided over 50 independent random realisations.

Two methods are used to combine the classifier outputs: the first is a confidence-based approach, where the final prediction is selected based on the highest confidence score from the individual classifiers, while the second method employs a logistic regression model, trained using the outputs of the individual classifiers as input features.

In the logistic regression-based integration approach, the outputs of the individual classifiers are combined into a single feature vector for each sample, concatenating the outputs of both classifiers. For multi-class classification, a one-vs-rest scheme is employed: a separate binary logistic regression model is trained for each class to predict the probability that a given sample belongs to that class versus all others. In the case of binary classification, only a single logistic regression model is trained. Each model produces a probability score for each class, which reflects the model's confidence in assigning a sample to that class.

To further improve reliability and reduce misclassifications, a rejection mechanism is applied. For each sample, the highest predicted probability across all classes is compared against a predefined threshold $T \in [0, 1]$. If the maximum probability is below the threshold, the sample is assigned to an "Unknown" class rather than forcing a potentially erroneous classification. By adjusting the threshold, a trade-off can be achieved between accuracy and coverage: higher thresholds result in fewer misclassifications but a larger number of rejected samples, whereas lower thresholds increase coverage at the risk of misclassification. The rejection rate is defined as:

$$\text{Rejection Rate} = \frac{N_{\text{rejected}}}{N_{\text{total}}} \times 100\% \tag{4.17}$$

where $N_{\text{rejected}}$ is the number of rejected samples and $N_{\text{total}}$ is the total number of test samples. Three thresholds are evaluated: $T = 0.5$, $T = 0.6$, and $T = 0.7$. A higher threshold results in more conservative classifications, leading to fewer misclassifications but a higher number of rejected samples. This trade-off is particularly useful in scenarios where misclassification has significant consequences, such as in defence applications, where an incorrect classification could lead to operational risks. Krawtchouk moments are computed directly from the original images and normalisation is applied to the resulting moment vectors, to ensure numerical stability. In the following, the results for the combined classifier for different combinations of ship classes are presented and discussed, focusing on the best-performing configurations. Including all results would be excessive and unnecessary for the analysis.

### 4.7.1 Combined Classifier Results: Cargo and Tanker

In Table 4.5, a comparative analysis is shown for the classification of Cargo and Tanker, to evaluate the performance of classifiers based on Krawtchouk moments, spectral features, and a combined approach, showing a summary of the most interesting results.

The combination of Krawtchouk moments (order 5) with normalised spectral features significantly improves classification accuracy, achieving a peak accuracy of 90.53%. However, this result comes with the cost of a high rejection rate of 55.77% at a threshold of 0.7 out of 1 (being 1 the maximum accuracy score for a class). Using autocorrelation and logarithm spectral features with Krawtchouk moments (orders 10 and 5) shows promising results, especially when combined with logarithm spectral features, achieving accuracies up to 86% with moderate rejection rates (∼20%).

The normalisation of spectral features, when used in conjunction with Krawtchouk moments, clearly enhances the classification accuracy when distinguishing between Cargos and Tankers. This is evident from the improvement in results with normalised features compared to non-normalised spectral features. The selection of spectral features plays a crucial role in the performance of the classifiers, and the tailored combination of these features with Krawtchouk moments significantly impacts the classification results.

Krawtchouk moments, particularly of order 5, emerge as a powerful tool for classification alone and when combined with specific spectral features, indicating their effectiveness in capturing essential shape and texture information from the SLC images. In some cases, order 10 provides a slight improvement and comparable performance to order 5.

Table 4.5: Summary of the best binary classification results for Cargos vs. Tankers in SLC format. The comparison includes classifiers using Krawtchouk moments, spectral features, and the combination.

| **Feature Set** | **Threshold** | **Best Classifier** | **Accuracy (%)** | **Rejection (%)** | **Data Format** |
|---|---|---|---|---|---|
| Normalised / Non-Normalised Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 5) | 79.50 | 0.00 | Non-Normalised SLC |
| | Th = 0.6 | Krawtchouk (Order 5) | 83.18 | 10.97 | |
| | Th = 0.7 | Krawtchouk (Order 5) + Normalised Spectral Features | 90.53 | 55.77 | |
| Auto-Correlation + Logarithmic Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 10/5) | 79.31 / 79.03 | 0.00 | Non-Normalised SLC |
| | Th = 0.6 | Krawtchouk (Order 5) + Logarithm Spectral Features / Krawtchouk (Order 5/10) | 86.00 / 83.40 / 83.50 | $\sim$20.00 / $\sim$10-15 | |
| | Th = 0.7 | Krawtchouk (Order 5/10) | 86.00 / 85.00 | $\sim$20 | |
| Normalised / Non-Normalised Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 5) + Normalised Spectral Features (Regression Model) | 78.30 | 0.00 | Normalised Calibrated SLC |
| | Th = 0.6 | Krawtchouk (Order 5) + Normalised Spectral Features (Regression Model) | 82.95 | $\sim$10.00 | |
| | Th = 0.7 | - | - | - | |

### 4.7.2 Combined Classifier Results: Tanker and Other

In Table 4.6, a comparative analysis is presented for the classification of Tanker vs. Other, extending previous findings by incorporating Krawtchouk moments alongside spectral features, as well as an unknown class. The evaluation considers various feature extraction techniques, including normalised/non-normalised spectral features, auto-correlation, and logarithmic transformation, to determine the optimal classifier configuration.

The combination of Krawtchouk moments (Order 10) with auto-correlation of non-normalised spectral features achieves the highest classification accuracy, reaching 99.98% at a threshold of 0.7. This result demonstrates the effectiveness of auto-correlation in unprocessed SLC data, which retains the full backscatter signal variation crucial for distinguishing Tankers from other ship types. Notably, this setup also maintains a low rejection rate ($\sim$1.5%) for the unknown class, ensuring a balance between accuracy and sample retention. Nonetheless, this result is slightly lower than the accuracy obtained previously, taking into account only the spectral features for classification, where the highest accuracy was ∼99.92% for auto-correlation features, which can be recalled from Table 4.3. This suggests that combining classifiers in this case is not necessary, as relying solely on spectral features leads to better performance while reducing computational complexity.

Other combinations also perform well, like normalised spectral features combined with Krawtchouk moments (Order 5), which consistently achieve high accuracy (approximately 99.85%), particularly for non-normalised SLC images. This suggests that while normalisation aids classification in processed images, it is less critical in raw SLC data, where unnormalised features excel. The findings align with the Cargo vs. Other and Cargo vs. Tanker cases, where normalisation was beneficial primarily in calibrated and processed data.

Auto-correlation features, while highly effective in certain cases, display significant variability depending on the normalisation method and dataset type, as shown in Table 4.6. For non-normalised SLC images, auto-correlation combined with Krawtchouk moments achieves near-perfect classification, reaching 99.92% at Th = 0.5 and further improving to 99.98% at Th = 0.7. This trend suggests that auto-correlation enhances the classifier's ability to capture relevant spatial relationships in raw SLC backscatter, particularly when paired with Krawtchouk moments. The rejection rate remains low, only increasing from 0% at Th = 0.5 to 1.5% at Th = 0.7, demonstrating that the classifier maintains high accuracy without excessively rejecting samples.

However, when auto-correlation is applied to calibrated SLC images, accuracy drops significantly to the range of 55-63%, even when combined with Krawtchouk moments. This is particularly evident in the calibrated auto-correlation of normalised features, where accuracy is 98.72% in raw SLC but drops to 63.01% in calibrated SLC. This suggests that calibration introduces distortions that reduce the effectiveness of auto-correlation. The rejection rates also rise for calibrated SLC auto-correlation features, reaching 3% at Th = 0.7, indicating that while calibration makes the data more interpretable, it also increases classification uncertainty.

Overall, combining classifiers does not yield better results than using the spectral classifier alone, making it unnecessary to allocate time and resources to this approach for the Tanker vs. Other classification. The most effective method remains the combination of auto-correlation of non-normalised features in raw SLC images with Krawtchouk moments of order 10, achieving a classification accuracy of 99.98% with minimal rejection. Normalised spectral features in normalised SLC images also perform well, offering a strong alternative. As observed previously, for the Tanker vs. Other classification, radiometric calibration consistently reduces performance, while logarithmic transformation fails to provide meaningful improvements, highlighting the importance of selecting features based on the specific characteristics of the data.

Table 4.6: Summary of the best binary classification results for Tankers vs. Other in SLC format. The comparison includes classifiers using Krawtchouk moments, spectral features, and their combinations.

| **Feature Set** | **Threshold** | **Best Classifier** | **Accuracy (%)** | **Rejection (%)** | **Data Format** |
|---|---|---|---|---|---|
| Auto-Correlation of Non-Normalised Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 5) + Auto-Corr Non-Normalised (Regression Model) | 99.92 | 0.00 | Non-Normalised SLC |
| | Th = 0.6 | Same as above | 99.95 | 0.50 | |
| | Th = 0.7 | Krawtchouk (Order 10) + Auto-Corr Non-Normalised (Regression Model) | 99.98 | ∼1.50 | |
| Normalised Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 5) + Normalised Spectral Features (Regression Model) | 99.50 | 0.00 | Normalised SLC |
| | Th = 0.6 | Same as above | 99.70 | 1.00 | |
| | Th = 0.7 | Krawtchouk (Order 10) + Normalised Spectral Features (Regression Model) | 99.85 | ∼2.00 | |
| Non-Normalised Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 5) + Non-Normalised Spectral Features (Regression Model) | 99.28 | 0.00 | Normalised Calibrated SLC |
| | Th = 0.6 | Same as above | 99.45 | ∼1.00 | |
| | Th = 0.7 | Same as above | 99.55 | ∼2.50 | |
| Auto-Correlation of Normalised Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 5) + Auto-Corr Normalised (Regression Model) | 98.72 | 0.00 | Normalised Calibrated SLC |
| | Th = 0.6 | Krawtchouk (Order 10) + Auto-Corr Normalised (Regression Model) | 98.85 | ∼1.50 | |
| | Th = 0.7 | Same as above | 99.00 | ∼3.00 | |
| Logarithm of Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 5) + Logarithm Features (Regression Model) | 67.00 | 0.00 | Normalised Calibrated SLC |
| | Th = 0.6 | Krawtchouk (Order 10) + Logarithm Features (Regression Model) | 70.00 | ∼5.00 | |
| | Th = 0.7 | Same as above | 72.00 | ∼10.00 | |

### 4.7.3 Combined Classification Results: Cargo and Other

Table 4.7 presents a comparative analysis of the best classification results achieved using different feature combinations and classifier combination for the classes Cargo and Other. The study incorporates spectral features and Krawtchouk moments applied to both SLC and calibrated SLC image formats. The classifiers differ based on their feature sets (normalised/non-normalised spectral features, logarithmic transformation, and auto-correlation spectral features) and the order of Krawtchouk moments used. For all feature sets, applying a stricter threshold for the unknown class enhances classification accuracy however leading to a higher rejection rate. The trade-off between accuracy and rejection rate depends on the specific application and operational requirements.

The combination of classification results from Krawtchouk moments order 10 and the result from the non-normalised spectral features from non-normalised SLC images achieves the highest accuracy (91.93%) among the SLC format datasets, with a moderate rejection rate of around 5%. This indicates that for non-normalised SLC format data, higher-order moments combined with non-normalised spectral features provide the best balance between accuracy and rejection rate.

For normalised vs. non-normalised spectral features, Table 4.7 shows that not normalised features tend to provide slightly higher accuracy, especially when combined with higher-order Krawtchouk moments. From the results, it can be drawn that the combination of classifiers performs differently on SLC vs. calibrated SLC data formats, and overall enhances the performance of the final classification. While the highest accuracy for SLC format reaches 91.93%, calibrated SLC data sees a maximum accuracy range of 87-90% with a corresponding increase in rejection rates ($\sim$10-15%). This might indicate that while calibrated SLC data might require more refined or different feature selection and classification strategies, it presents a more challenging scenario for achieving high accuracy for the case of Cargo vs. Other.

When considering the results obtained solely from the spectral classifier, without introducing an unknown class, the classification performance varies depending on the selected feature configurations, with the highest accuracy being 85.25%.

Comparing the two approaches, the combination of Krawtchouk moments classification with spectral features classification generally provides better results than spectral classification alone. The introduction of an unknown class further refines classification reliability by reducing misclassifications, ensuring a more confident classification process. However, this comes at the expense of a higher rejection rate. The decision to integrate both classifiers depends on the specific application and resources available.

Table 4.7: Summary of the best binary classification results for Cargo vs. Other. The comparison includes classifiers using Krawtchouk moments, spectral features, and their combinations.

| **Feature Set** | **Threshold** | **Best Classifier** | **Accuracy (%)** | **Rejection (%)** | **Data Format** |
|---|---|---|---|---|---|
| Normalised / Non-Normalised Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 5) + Non-Normalised Spectral Features (Regression Model) | 90.00 | 0.00 | Non-Normalised SLC |
| | Th = 0.6 | Same as above | 90.21 | 2.38 | |
| | Th = 0.7 | Krawtchouk (Order 10) + Non-Normalised Spectral Features (Regression Model) | 91.93 | ∼5.00 | |
| Auto-Correlation + Logarithmic Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 10) + Auto-Correlation Features (Regression Model) | 85.19 | 0.00 | Non-Normalised SLC |
| | Th = 0.6 | Same as above | 89.35 | ∼10.00 | |
| | Th = 0.7 | Same as above | 91.67 | ∼15.00 | |
| Normalised / Non-Normalised Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 5) + Normalised Spectral Features (Regression Model) | 84.81 | 0.00 | Normalised Calibrated SLC |
| | Th = 0.6 | All classifier combinations show similar results. Higher accuracy for non-normalised, but higher rejection. | 85-87 | ∼5-10 | |
| | Th = 0.7 | Same as above | 87-90 | ∼10-15 | |
| Auto-Correlation + Logarithmic Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 5) + Non-Normalised Log Features (Regression Model) | 85.60 | 0.00 | Normalised Calibrated SLC |
| | Th = 0.6 | Krawtchouk (Order 10) + Auto-Correlation Features (Regression Model) | 88.41 | ∼5.00 | |
| | Th = 0.7 | Same as above | 91.20 | ∼15.00 | |

### 4.7.4 Multiclass classification

A comprehensive analysis has been conducted on multiclass classification involving the three distinct classes. Similar to the binary classification approach, various methods were employed and evaluated for their effectiveness. For this purpose, Krawtchouk moments of order 5 and order 10 were used to achieve classification outcomes independently. Additionally, different spectral features were examined within their respective networks. In the process of combining both networks, a comparison was made between a confidence-based method and a Multinomial Logistic Regression (MLR) to determine the most effective strategy for classification. MLR is an extension of binary logistic regression that allows for the prediction of outcomes with more than two categories, estimating the association between a set of predictors (independent variables) and a multicategory nominal (unordered) outcome. This type of regression is particularly useful when the outcomes are categorical and do not have a natural order [171].

In Table 4.8, a comparison of the best classification results for classifying Cargo, Tanker, and Other types of vessels based on different feature sets and strategies is presented, emphasising the effectiveness of combining Krawtchouk moments with spectral feature analysis using MLR. In all trials, the combination of Krawtchouk moments (order 10) with spectral characteristics - normalised or not - outperforms individual classification methods. This suggests that integrating spatial with spectral features provides a more robust framework. Results are presented based on whether spectral features were normalised or not and on different types of spectral features (normalised and non-normalised spectral features, auto-correlation of spectral features, and logarithm of spectral features). Interestingly, the highest accuracy is observed when using non-normalised spectral features from normalised calibrated SLC data (81.93%), however, the rejection rate is over 20% which means that the classification is not as successful, indicating that for this specific classification task, retaining the original scale of spectral features might be beneficial. In Table 4.8 it can be seen that increasing the threshold for classifying an observation as Unknown from 0.5 to 0.6 generally leads to higher accuracy. This implies that being more conservative about classifying uncertain observations can improve overall performance by reducing false positives, at the cost of potentially increasing false negatives (as seen in the increased rejection rates).

In conclusion, combining classifiers offers better performance over single-feature classifiers for multiclass vessel classification. It also highlights the effects of feature normalisation, the selection of spectral features, the setting of classification thresholds, and the importance of data format on the overall classification performance. These findings can guide future research and practical applications in vessel classification and similar remote sensing tasks.

Table 4.8: Summary of the best results for the multiclass classification of Cargos vs. Tankers vs. Other. The comparison includes classifiers using Krawtchouk moments, spectral features, and their combinations. The spectral features column specifies the spectral features used for each classification trial.

| **Feature Set** | **Threshold** | **Best Classifier** | **Accuracy (%)** | **Rejection Rate (%)** | **Data Format** |
|---|---|---|---|---|---|
| Normalised / Non-Normalised Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 10) + Non-Normalised Spectral Features (MRM) | 74.38 | 3.34 | Non-Normalised SLC |
| | Th = 0.6 | Krawtchouk (Order 10) + Non-Normalised Spectral Features (MRM) | 79.72 / 81.93 | 19.67 / 21.2 | Non-Nnormalised SLC/ Normalised Calibrated SLC |
| Auto-Correlation of Normalised / Non-Normalised Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 10) + Auto-Correlation of Non-Normalised Spectral Features (MRM) | 73.28 | 3.41 | Normalised Calibrated SLC |
| | Th = 0.6 | Krawtchouk (Order 10) + Auto-Correlation of Non-Normalised Spectral Features (MRM) | 79.71 | 23.48 | Normalised Calibrated SLC |
| Logarithm of Normalised / Non-Normalised Spectral Features + Krawtchouk (Order 5/10) | Th = 0.5 | Krawtchouk (Order 10) + Non-Normalised Logarithm of Spectral Features (MRM) | 70.24 | 0.87 | Normalised Calibrated SLC |
| | Th = 0.6 | Krawtchouk (Order 10) + Non-Normalised Logarithm of Spectral Features (MRM) | 77.32 | 17.26 | Normalised Calibrated SLC |

A summary of the best result for each format of spectral features can be seen in Fig. 4.13, where the threshold value Th=0.5 is represented by the blue bars and Th=0.6 by the orange bars, with their corresponding rejection rates in green and red.

## 4.8 Discussion

The presented methodology demonstrates significant potential in advancing ATR in SAR imaging, specifically for maritime applications. By leveraging spectral features and Krawtchouk moments derived from SLC SAR data, the method effectively differentiates between ship types, such as Cargo, Tanker, and Other vessel classes, even when physical characteristics are ambiguous.

A key advantage of the approach lies in its use of spectral profiles and invariant features like Krawtchouk moments. Spectral profiles capture the unique signatures of

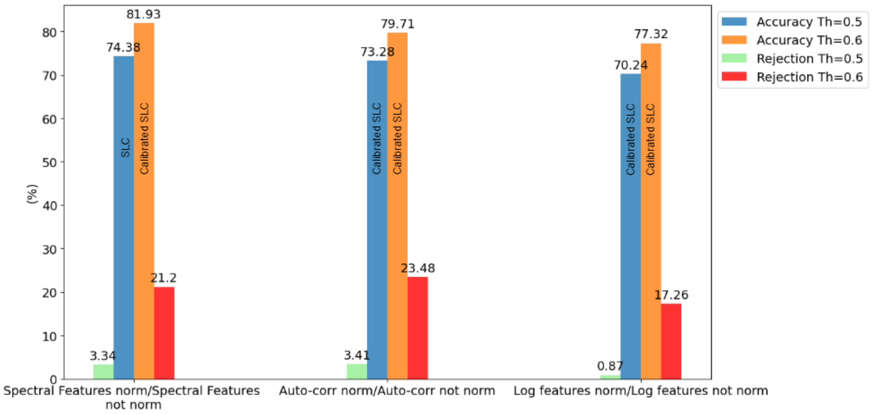


Figure 4.13: Bar graph illustrating multiclass classification accuracy across various thresholds for the unknown class, along with their rejection rates, highlights how varying the threshold impacts the system's ability to identify and reject ambiguous or unclassified instances.

ship engines and motions, while Krawtchouk moments offer robust rotation and position invariance, addressing alignment issues in SAR images. These complementary features mitigate the challenges posed by overlapping physical characteristics across different vessel categories. The study highlights the effects that different data preprocessing, including radiometric calibration and normalisation, have in the classification performance. Radiometric calibration ensures consistent backscatter values, enabling the classifier to extract more meaningful patterns, while normalisation aligns feature scales, reducing model sensitivity to outliers.

The integration of two independent classifiers-one utilising spectral profiles and the other leveraging Krawtchouk moments-proves to be a robust solution. The combined approach outperforms individual classifiers in some cases, achieving higher accuracy and stability, particularly in challenging scenarios like multiclass classification. However, the choice of feature configuration and thresholds must be carefully tailored to the specific use case, as evidenced by the trade-offs between accuracy and rejection rates.

## 4.9 Conclusion

This chapter presents a novel ATR framework that combines spectral profiles and invariant features for maritime target classification in SAR imaging. The approach addresses the limitations of traditional classification techniques by focusing on spectral and invariant properties, enabling reliable differentiation between vessel classes. The contributions shown in this chapter can be summed up as follows:

- Innovative use of spectral features: by extracting and analysing spectral profiles, the method can capture engine-induced signatures that are difficult to discern using physical features alone, like width and length;
- Integration of Krawtchouk moments: incorporating position- and rotation-invariant moments enhances the robustness of classification, especially in scenarios with misaligned targets;
- Dual-Classifiers combination: The fusion of spectral and Krawtchouk-based classifiers achieves superior performance compared to individual methods, underscoring the benefits of complementary feature sets.

The results, with classification accuracies exceeding 80% in multiclass scenarios and over 99% in binary classification trials, validate the effectiveness of the proposed methodology. This framework not only improves the detection and recognition of maritime targets but also sets the basis for future advancements in SAR-based ATR systems.

In contrast to the dominant intensity-based and deep learning approaches, this work investigates the use of spectral features derived from SLC SAR data, explicitly leveraging the coherent nature of the signal. By applying Fourier analysis in azimuth and aggregating range information, spectral profiles are generated that may encode dynamic characteristics of the observed vessels. Furthermore, the integration of Krawtchouk moment-based features introduces a complementary representation that captures spatial distributions in a mathematically compact and orthogonal basis. This combination allows the proposed framework to bridge spatial and spectral domains, offering a richer description of the target compared to conventional methods.

While state-of-the-art CNN approaches generally achieve higher classification accuracies, they rely heavily on large labelled datasets and may lack physical interpretability. In a study performed on the OpenSARShip dataset, the authors train a three-class model to distinguish between tanker, container ship, and bulk carrier, using a dual polarised network based on a pretrained visual geometry group model, obtaining a 88% accuracy that drops to 67% when 5 classes are used [172]. However, this is done on amplitude only, ignoring the phase information in SAR images and results compare to the multiclass classification shown in this chapter, where highest accuracy reached around 81%.

The approach proposed in this work prioritises interpretability and the exploration of alternative feature spaces, which may be particularly valuable in scenarios with limited training data or when understanding the underlying scattering mechanisms is required. The classification performance observed in this study reflects both the inherent difficulty of distinguishing between similar vessel types and the exploratory nature of the proposed features. Nevertheless, the results demonstrate that spectral representations contain discriminative information that could be further exploited, combined with established spatial-domain methods.

# Chapter 5

# Automatic Location, Characterisation, and Classification of Vibrating Maritime Targets using High-Order Features

The objective of this chapter is to develop and assess a methodology for the detection, classification, and characterisation of potential maritime targets in SLC SAR imagery, exploiting the complex data to retrieve specific information. Candidate targets are first located in the Range-Doppler (RD)-image domain using a CFAR detector together with STFT analysis, while auto-correlation signatures extracted from the selected range bins are employed as discriminative descriptors. An SVM model is then used to perform multi-class classification among three categories:

1. **Static targets** (no induced vibrations), representing static objects;
2. **Vibrating targets**, vessels with active propulsion systems and working machinery;

3. **Clutter**, areas that do not contain any target, only noise and sea clutter.

To this end, simulated synthetic SAR SLC images containing static and vibrating targets embedded in clutter are created to train and test the processing pipeline, while real SAR images are used to demonstrate the extrapolation of the method to real-world use cases.

## 5.1 Structure of the Chapter

This chapter presents a methodology for discriminating between vibrating and non-vibrating maritime targets using detection techniques, autocorrelation feature extraction, and an ML classifier. The chapter begins with the motivation for vibration-based discrimination and the problem formulation. A detailed description of the feature extraction pipeline, including SLC to RD image conversion, target location, and final classification, is then provided. An SVM classifier is introduced to perform multiclass classification between clutter, vibrating, and static targets. The chapter concludes with experimental results obtained on synthetic SAR datasets and the application to real SAR data, followed by a discussion of the implications for maritime surveillance.

## 5.2 Methodology

The contribution of this chapter is a processing pipeline to identify possible vibrating objects in SLC SAR images to distinguish between vibrating targets, non-vibrating objects, and clutter on the ocean, plus a further estimation of the identified vibrations. Moreover, a detection and classification pipeline is proposed and tested on both simulated and real data, and its performance is discussed. This section presents the methodology for the detection, characterisation, and classification of the targets in three complementary levels: (i) a high-level description of the processing steps, (ii) a block diagram to facilitate clarity, and (iii) a mathematical formulation.

### 5.2.1 High-level description of the processing pipeline

The proposed processing pipeline transforms a focused SLC SAR image into a compact feature representation that captures micro-vibration signatures, which are later used for target classification. The main processing stages are summarised below:

1. **Input SLC image**: The process starts from a focused SLC SAR image containing potential maritime targets embedded in sea clutter. These images are widely available to the public and open code is accessible to work with them. They preserve phase and amplitude while being a focused representation of the scene, which reduces its size with respect to other representations from which vibrations can be estimated, such as phase history images. Compared to SLC, phase history images are heavier and require more knowledge of SAR theory to work with.

2. **Range-Doppler transformation**: The SLC image is transformed into the RD domain to expose micro-Doppler signatures linked to target motion and micro-vibrations. This is done to emulate a step back in the focusing process, where vibrations can manifest, effectively defocusing the image in the azimuth direction and obtaining the cross-range profiles. Micro-motion effects are a key physical phenomenon in this processing stage. The periodic vibrations of active vessel machinery induce slow-time phase modulations in the SAR signal, which alter the coherent integration process used in SAR image formation. As a consequence, originally compact scatterers exhibit azimuthal energy spreading, structured replicas (ghost-like responses), and micro-Doppler signatures in the Doppler domain. Ghost targets can sometimes be evident in the focused SLC image, but they become analysable once the data is re-expressed in the RD domain, where slow-time phase variations can be explored explicitly for vibration-related information. It should be noted that this operation is applied to an already focused SLC image and does not correspond to the SAR image formation (focusing) stage; instead, it re-represents the focused data in the RD domain to enable slow-time phase analysis and the extraction of micro-motion-induced phase modulations.

3. **Candidate detection**: Potential target range bins are identified using a CFAR detector, which suppresses background clutter while preserving strong scatterers.
4. **Range bin aggregation and averaging**: After CFAR detects candidate range bins, contiguous bins likely corresponding to the same physical target are grouped together. The signals within each group are averaged to form a single representative range-compressed signal. This reduces the effect of bin-to-bin variability and improves the robustness of STFT-based vibration detection.
5. **Candidate refinement**: The detected range bins, along with any aggregated bins resulting from contiguous detections, are further processed to suppress false alarms and isolate cross-range profiles associated with potential targets by applying the STFT.
6. **Cross-range profile selection**: Following refinement, a single representative range bin is selected for each detected target. In the case of contiguous detections, only the central range bin of the group is retained, rather than the averaged signal, to preserve the intrinsic micro-motion characteristics and avoid spectral smearing or leakage effects. The corresponding cross-range profiles extracted from these selected bins are then used for subsequent processing.
7. **Autocorrelation analysis**: The cross-range profiles are processed by applying autocorrelation to enhance periodic components and suppress noise, highlighting vibration-induced patterns in the signals. Moreover, the estimation of the vibration frequency is drawn from the autocorrelation result.
8. **Feature extraction**: A one-dimensional feature vector is derived from the resulting autocorrelation signal, highlighting the presence and characteristics of micro-vibrations.
9. **Multiclass classification**: Finally, the extracted features are classified using a SVM to distinguish between vibrating targets, non-vibrating targets, and clutter.

These steps define a hierarchical processing framework that progressively reduces the dimensionality of the data while preserving discriminative micro-motion information, shown in Fig. 5.1. The following subsections describe each processing stage in detail.

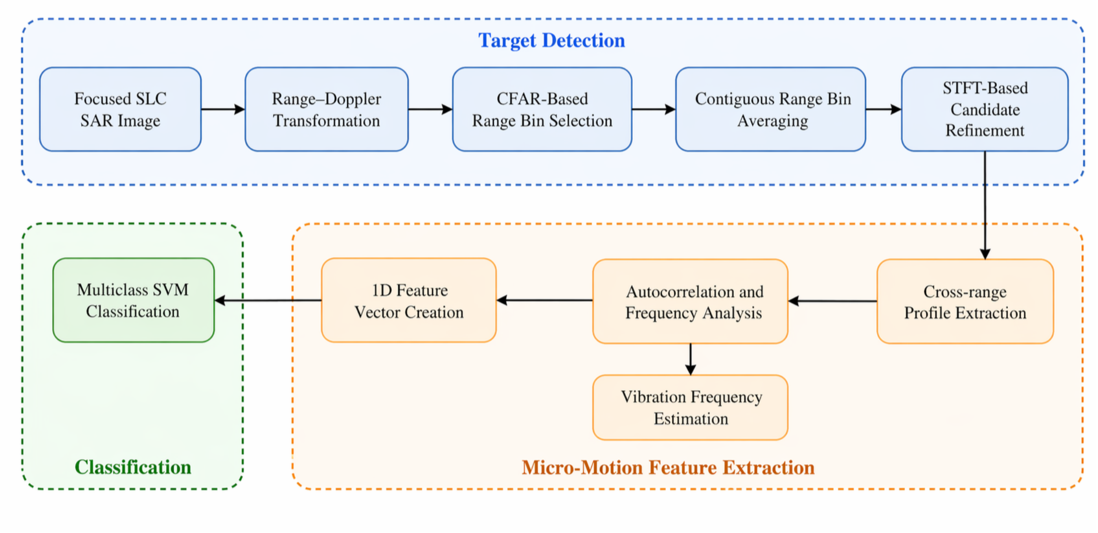


Figure 5.1: Block diagram of the proposed processing pipeline for vibrating target detection, vibration frequency estimation, and multiclass classification from SLC SAR imagery.

### 5.2.2 Mathematical Formulation of the Processing Pipeline

This section presents the mathematical formulation of the proposed processing pipeline for the detection and refinement of vibrating scatterers in SLC SAR images.

#### Signal Model and Assumptions

Let the focused SLC SAR image be denoted as $\mathbf{S} \in \mathbb{C}^{N_r \times N_a}$, a complex matrix where $N_r$ and $N_a$ are the number of range bins and azimuth samples, respectively.

The following assumptions are made throughout the section:

- The SLC images are motion-compensated and phase-preserving.
- Noise in the RD domain is modelled as complex Gaussian, leading to exponentially distributed energy measurements.

- Vibrating targets induce periodic micro-motion along the radar line of sight, resulting in slow-time phase modulations.
- The vibrating target undergoes small-amplitude sinusoidal vibration exclusively along the z-axis (vertical direction in the acquisition geometry), with no lateral motion.
- Static or rigid scatterers exhibit negligible residual phase modulation after detrending.

### Detection and Refinement of Vibrating Scatterers

The detection and refinement stage aims to identify candidate vibrating scatterers in the RD domain while suppressing non-vibrating clutter. The procedure consists of a CFAR-based detection step followed by an STFT refinement step exploiting and enhancing true vibrations.

**Range-Doppler Representation** An FFT is applied along the azimuth dimension to defocus the image and obtain the RD representation of it:

$$I_{RD}(r,k) = \sum_{n=0}^{N_a-1} S(r,n)\, e^{-j2\pi kn/N_a}, \tag{5.1}$$

where $k$ denotes the Doppler frequency index. In Fig. 5.2, an example of a synthetic SLC image side by side with its corresponding RD image can be seen, after the process described in this section. After FFT, the energy is spread over the azimuth axis, and the visual differentiation of targets is not possible at this stage.

**CFAR-Based Candidate Detection** For each range bin $r$, a Cell-Averaging CFAR (CA-CFAR) detector is applied independently across the range dimension. For a Cell Under Test (CUT) at range bin $r$, the noise power is estimated from the training cells on either side of the CUT, excluding the guard cells. The CA-CFAR detector is the most used CFAR variant, and can be implemented as a baseline for other CFAR techniques

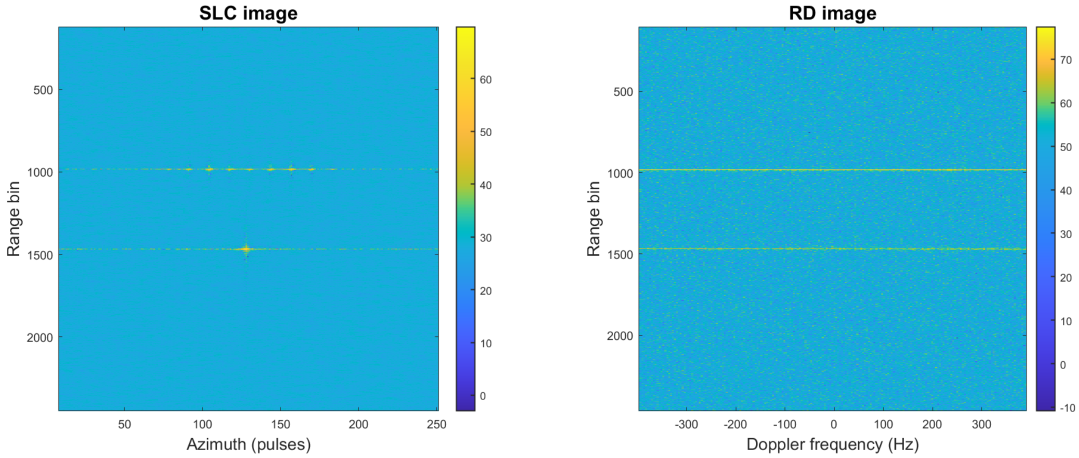


Figure 5.2: Synthetic SLC image of a vibrating and a static point (left), and its corresponding RD image (right).

[173]. In a typical radar scene, targets are embedded in background noise and clutter, which may show statistical homogeneity depending on the physical characteristics of the environment. Nevertheless, the local scattering characteristic can sharply change and create sudden boundaries between scattering regions. The approach followed to estimate the background noise power for detection in a specific cell, is to average the measured power in the surrounding cells [174], which is the CA-CFAR method used in this pipeline, producing a set of detected bins: $\mathcal{R}_{\text{CFAR}} = \{r_1, r_2, \ldots, r_M\}$.

**Contiguous Range bin Averaging** Following CFAR detection, contiguous range bins are grouped in clusters and complex-valued coherent averaging is performed to form a single representative candidate per physical scatterer, reducing redundant detections caused by energy spreading across adjacent range bins. let $\mathcal{G}_m$ denote the $m$-th contiguous bin cluster obtained after CFAR detection. The aggregated cross-range profile is then defined as:

$$\bar{I}_{RD}(r_m, k) = \frac{1}{|\mathcal{G}_m|} \sum_{r \in \mathcal{G}_m} I_{RD}(r, k), \tag{5.2}$$

where $k$ is the Doppler index, $|\mathcal{G}_m|$ is the number of bins in cluster $\mathcal{G}_m$, and $r_m$ is the representative bin of the cluster, chosen as the mean of the cluster indices:

$$r_m = \left\lfloor \frac{1}{|\mathcal{G}_m|} \sum_{r \in \mathcal{G}_m} r \right\rfloor . \tag{5.3}$$

This is done under the assumption that contiguous detected bins correspond to the same dominant scattering centre, with approximate coherent phase behaviour. Following candidate selection refinement, autocorrelation and feature extraction are performed on the representative range bin signal to preserve localised micro-motion characteristics for classification and avoid degradation of the signal content that can later affect the SVM performance.

**STFT-Based Vibration Refinement** To discriminate vibrating targets from static scatterers, the refinement stage is applied to the spatially averaged cross-range profile associated with each detected scattering cluster. The slow-time complex sequence used for vibration analysis is defined as

$$x_m(k) = \bar{I}_{RD}(r_m, k), \tag{5.4}$$

where $\bar{I}_{RD}(r_m, k)$ denotes the averaged Range-Doppler signal over the contiguous CFAR detection group $\mathcal{G}_m$ associated with representative range bin $r_m$.

Mean removal, phase extraction, and unwrapping are applied:

$$\phi_r(n) = \mathrm{unwrap}\left(\arg\{x_r(n) - \mathbb{E}[x_r(n)]\}\right). \tag{5.5}$$

In coherent radar systems, the measured phase of the return signal is limited to the interval $[-\pi, \pi)$, meaning that the argument of the complex number “wraps around” the unit circle when it goes over $+\pi$ or lower than $-\pi$. This causes phase ambiguity, as different delays can correspond to the same observed phase shift, and phase unwrapping is needed to recover the number of $2\pi$ cycles to reconstruct the original delay unambiguously [175,176]. The unwrapped phase is then linearly de-trended to suppress residual trends unrelated to vibration:

$$\tilde{\phi}_r(k) = \phi_r(k) - \mathcal{T}\{\phi_r(k)\}, \tag{5.6}$$

where $\mathcal{T}\{\cdot\}$ denotes least-squares trend removal. The reason for de-trending the unwrapped phase is to reduce the leakage from low-frequency signal components, revealing the underlying character of the data at low frequencies. Linear trends manifest strongly in the low-frequency part of the signal and can dominate the frequency content when an FT-based transform is applied [177].

A STFT is computed:

$$\mathrm{STFT}_r(f,\tau) = \sum_k \tilde{\phi}_r(k)\, w(k-\tau)\, e^{-j2\pi f k/f_s}, \tag{5.7}$$

where $w(\cdot)$ is a Hann window, $f_s$ is the pulse repetition frequency, and $\tau$ indexes the subapertures.

The energy magnitude spectrogram is defined as

$$S_r(f,\tau) = \left|\mathrm{STFT}_r(f,\tau)\right|^2. \tag{5.8}$$

To enhance the frequency at which the resulting signal appears stronger, the spectrogram is collapsed over time:

$$SP_r(f) = \sum_\tau S_r(f,\tau). \tag{5.9}$$

The spectrum $SP_r(f)$ represents the dominant vibrational behaviour of the de-trended phase signal at range bin $r$. Fig. 5.3 illustrates the result of this processing applied to the synthetic example shown in Fig. 5.2. In this simulated scenario, the target vibration was generated with a known frequency of 50 Hz. The peak observed in the frequency profile corresponds to this vibration frequency, validating the proposed processing chain.

The right images in Fig. 5.3 show the time–frequency representations obtained via the STFT of the phase history signal extracted from the range bin containing the vibrating target. By integrating the STFT energy across all time windows, a time-collapsed micro-Doppler spectrum is obtained. This operation emphasizes persistent spectral components over the observation time, allowing the dominant vibration frequency to be identified.

Spectral peaks are obtained and sorted by amplitude, and the strongest one is selected. As detrending has been applied to minimise the impact of the DC component, it reduces the sensitivity to the residual low-frequency influence, and captures the dominant vibrational component.

To quantify the prominence of the detected vibration component, a spectral dominance score is computed. After detecting the strongest peak at frequency $\hat{f}_{\text{vib}}$, the corresponding peak amplitude is denoted by

$$P_{\text{peak}} = P_r(\hat{f}_{\text{vib}}). \tag{5.10}$$

The refinement-based score is then defined as

$$\Gamma_r = \left( \frac{P_{\text{peak}}}{\sum_{f \in \Omega} P_r(f)} \right) \left( \frac{P_{\text{peak}}}{\text{median}\big(P_r(f)\big)} \right). \tag{5.11}$$

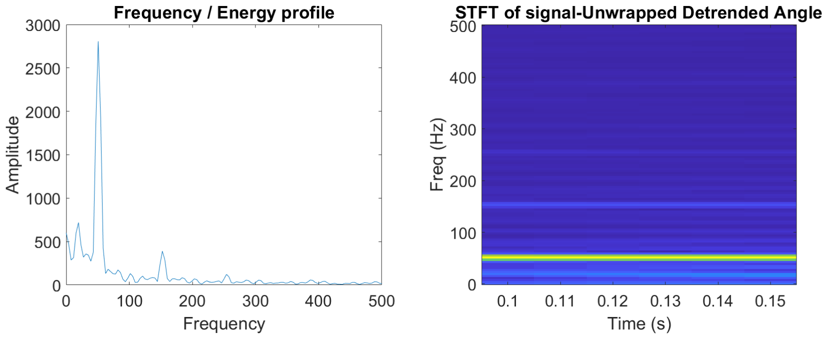
where $\Omega$ is the valid vibration frequency band (if known). If the valid vibration frequency band is not known, the sum is performed over all frequencies. This metric measures together the spectral concentration and peak-to-baseline contrast. A high value of $\Gamma_r$ indicates that the detected frequency component is both dominant within the spectrum and stands out above the spectral background, indicating the possible presence of a true vibrating target [178]. Finally, each candidate is compared to the strongest detected vibration signature. Let

$$\Gamma_{\max} = \max_{r \in \mathcal{R}_{\text{CFAR}}} \Gamma_r. \tag{5.12}$$

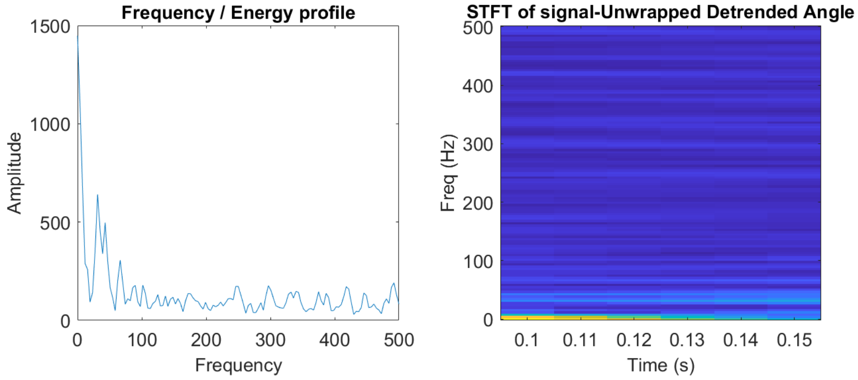
The refined candidate set is then defined as

$$\mathcal{R}_{\text{STFT}} = \{ r \in \mathcal{R}_{\text{CFAR}} \mid \Gamma_r \geq \eta\, \Gamma_{\max} \}, \tag{5.13}$$

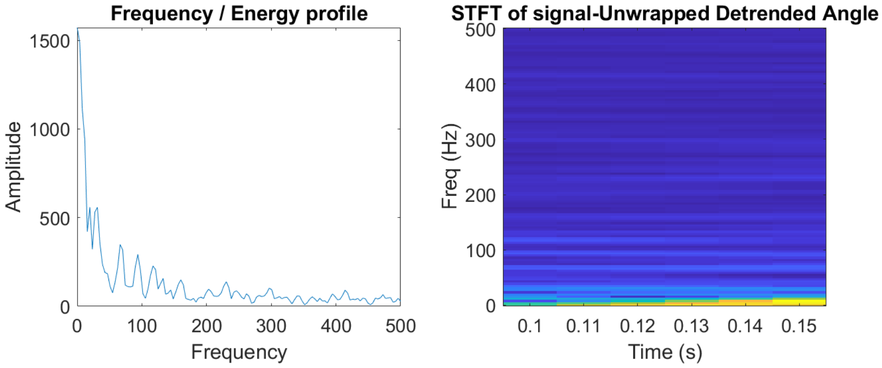
where $\eta \in (0, 1)$ is a relative threshold parameter. In this work, $\eta = 0.5$ is used,



(a) Vibrating target.



(b) Static target.



(c) Clutter.

Figure 5.3: STFT-based analysis of the slow-time SAR signal. The spectrogram and its derived representations illustrate the distribution of vibration-related energy across frequency and time.

to retain cross-range profiles whose vibration score remains within a fixed percentage of the strongest detected response, suppressing weak or noise-dominated detections. These refined range bins correspond to the most likely vibrating scatterers and are forwarded to subsequent feature extraction and classification stages.

The proposed vibration detection pipeline is composed of four stages: (i) Range-Doppler processing, (ii) CFAR-based candidate detection, (iii) STFT-based vibration refinement, and (iv) final candidate selection.

### Micro-motion Feature Extraction

SAR data, specially real SAR images that cover kilometers and where the resolution achieved is in the range of centimetres, is inherently of very high dimensionality. By leveraging CFAR detection and STFT refinement, the problem is reduced to a smaller subset of relevant range bins, making it possible to transform a very high power and time consuming task into a computationally efficient extraction of signatures.

After reducing the dimensionality of the problem, from potentially hundreds of range bins to test, to only a few, features are extracted from the final detected range bins, and the frequency of the vibration, is estimated. This involves computing the autocorrelation of each slow-time signal derived from each detected range bin. For each refined candidate range bin $r \in \mathcal{R}_{\text{STFT}}$, features are extracted from the associated slow-time complex signal $x_r(k) = I_{RD}(r, k)$, where $k$ indexes the pulses.

Two avenues to compute the autocorrelation have been studied:

- **Single autocorrelation:** computing the autocorrelation of the signal directly, defined as:

$$R_r(\ell) = \sum_k x_r(k)\, x_r^*(k - \ell), \tag{5.14}$$

  where $\ell$ denotes the lag index and $(\cdot)^*$ is the complex conjugate.

- **Double autocorrelation:** computing a second-order autocorrelation, defined as:

$$\tilde{R}_r(p) = \sum_\ell R_r(\ell)\, R_r^*(\ell - p), \tag{5.15}$$

where $p$ is the lag of the second autocorrelation stage.

Depending on the signal characteristics, like SNR, amplitude of the vibration, or clutter level, one approach may yield better feature representations than the other. To account for this, both options are studied and compared, as shown in an example in Fig. 5.4, where both single and double autocorrelations result in a very distinguishable vibrational pattern.

Instead of using the raw autocorrelation signals directly, features are extracted to avoid overfitting and lack of generalisation, and to improve robustness. Raw signals contain too much information, including noise and irrelevant variations, and outliers can be confused with important data. By extracting features, the focus is on those characteristics of the signal that capture the underlying structure and patterns. This reduces dimensionality and improves performance in new unseen data. The extracted *structure features* include:

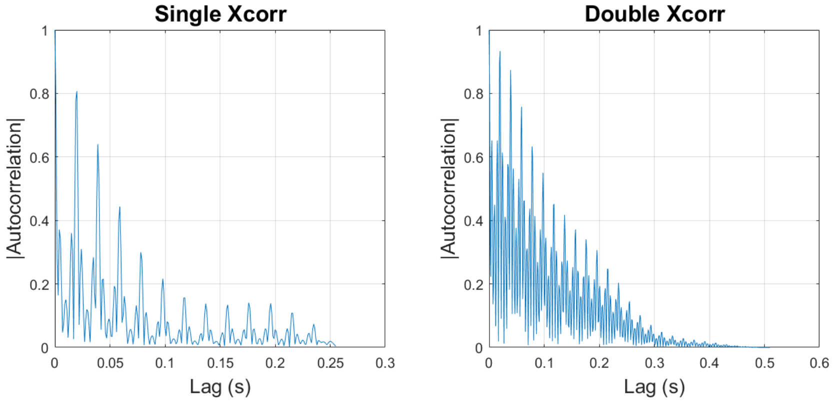
- **Dominant frequency ratio (F/R):** it measures how narrowband the signal is in the frequency domain [179]. Vibrating targets have a strong dominant frequency rate, while static targets and clutter are expected to exhibit low and moderate values for this feature, respectively.

  $$f_1 = \frac{\max_f |S_R(f)|}{\frac{1}{K}\sum_{f=1}^{K} |S_R(f)|}, \tag{5.16}$$

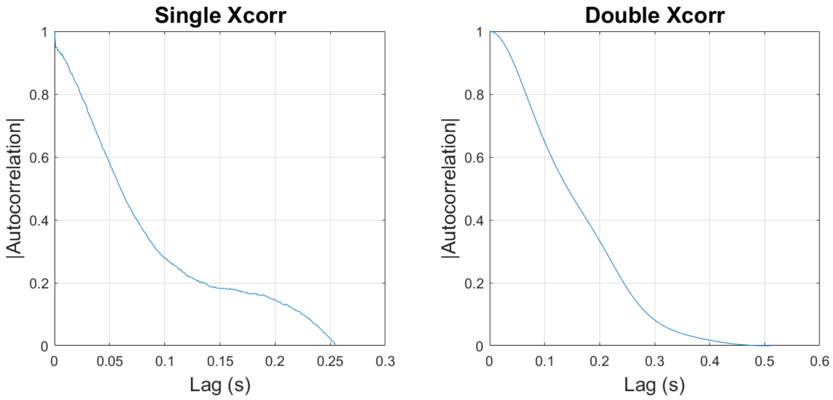
  where $S_R(f) = \mathcal{F}\{R(\ell)\}$ denotes the spectrum of autocorrelation signal, and $K$ is the number of discrete frequency bins considered.

- **Peak prominence:** it measures how dominant the strongest autocorrelation peak is relative to the mean energy [180]. Vibrating targets are expected to have sharper peaks than a non periodic target, whose autocorrelation is flatter. Clutter is expected to present an intermediate and inconsistent peak sharpness.

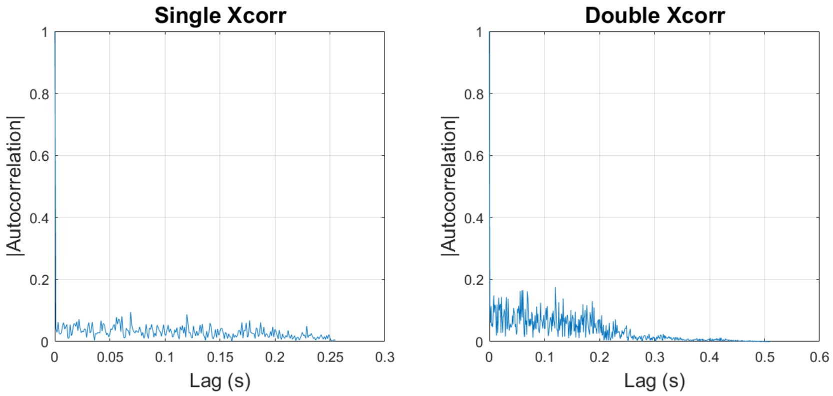
  $$f_2 = \frac{\max_{\tau>0} |\tilde{R}(\ell)|}{\sqrt{\frac{1}{T}\sum_{\tau=1}^{T} \tilde{R}^2(\ell)}}, \tag{5.17}$$



(a) Vibrating target.



(b) Static target



(c) Clutter.

Figure 5.4: Autocorrelation of the range compressed signal for a vibrating target (a), a static target (b), and clutter (c), accounting for both representations: single and double autocorrelations.

The function $\tilde{R}(\ell)$ denotes the normalised autocorrelation of the signal at lag $\ell$, with $\ell = 0$ corresponding to the zero-lag component (excluded).

- **Peak spacing consistency:** this feature measures how regular the space between the strongest autocorrelation peaks is, indicating whether a stable oscillation is present in the signal [181]. The coefficient of variation of peak spacing is computed and converted to a 0-1 metric. The vibrating target should result in a coefficient close to 1.

  $$f_3 = \exp\left( -h \cdot \frac{\sigma_{\Delta R}}{\mu_{\Delta R}} \right), \tag{5.18}$$

  The quantity $\Delta R$ represents the spacing between consecutive dominant autocorrelation peaks, while $\mu_{\Delta R}$ and $\sigma_{\Delta R}$ denote the mean and standard deviation of these spacings, respectively. The parameter $h$ is a positive tuning constant controlling the sensitivity of the spacing consistency metric.

- **Phase linearity score:** to measure how coherent the signal is, how close the evolution of it is to a linear model across lags, capturing temporal coherence [182]. Vibrating target should show coherent phase change, static targets mostly flat, and clutter should result in irregular phase.

  $$f_4 = 1 - \frac{\sigma\Big(\phi(\ell) - \hat{\phi}(\ell)\Big)}{\sigma(\phi(\ell))}, \quad \hat{\phi}(\ell) = a\ell + b, \tag{5.19}$$

  The function $\phi(\ell) = \arg\{R(\ell)\}$ denotes the unwrapped phase of the autocorrelation signal, and $\hat{\phi}(\ell)$ is the least-squares linear approximation of the phase as a function of lag, with coefficients $a$ and $b$. The operator $\sigma(\cdot)$ denotes the standard deviation.

**Sub-aperture-based autocorrelation features** In addition to the global structure features extracted from the full autocorrelation of the signal, a set of four more features is computed based on sub-apertures of the autocorrelation [183]. The motivation for using sub-apertures is to capture the temporal evolution of the signal's correlation structure, to acquire further discriminative metrics [184, 185]. These were proved to better separate vibration and clutter classes in scenarios where discrimination was specially difficult.

A sub-aperture is a short, overlapping window of the signal, covering a part of the full aperture length. Given the signal $x[n]$ of length $N$, a sub-aperture is defined as

$$x_w[n] = x[n_0 : n_0 + L - 1] \cdot w[n], \quad n_0 = 0, H, 2H, \dots \tag{5.20}$$

where $L$ is the window length, $H$ is the hop size between consecutive windows, and $w[n]$ is a tapering function such as a Hamming window.

This approach allows to detect changes in the autocorrelation structure of the signal that occur over time [186]. The idea is that vibrating targets will produce a consistent periodic signature across sub-apertures, but clutter will produce inconsistent peaks. By analysing the autocorrelation within each sub-aperture and comparing them, metrics can be extracted that can quantify temporal consistency.

For each sub-aperture, the normalised autocorrelation is computed and the dominant lag $\ell_{\text{dom}}[k]$ is identified as the lag corresponding to the maximum magnitude of $R_k[\ell]$ (excluding zero lag). With $K$ denoting the number of sub-apertures, the sequence of dominant lags $\{\ell_{\text{dom}}[k]\}_{k=1}^{K}$ forms the basis for the sub-aperture features.

The four features derived from this approach are:

1. **Agreement Ratio (AR):** The fraction of sub-apertures whose dominant lag matches the mode of the dominant lag across all sub-apertures:

   $$f_5 = \frac{1}{K} \sum_{k=1}^{K} \mathbf{1}\big(\ell_{\text{dom}}[k] = \ell_{\text{mode}}\big). \tag{5.21}$$

   This measures the temporal consistency of the dominant autocorrelation peak. Vibrating targets exhibit high agreement ratios, while static targets and clutter exhibit lower consistency [187].

2. **MAD Normalised:** The robust variability of the dominant lag:

$$f_6 = \frac{\mathrm{MAD}(\ell_{\mathrm{dom}})}{\ell_{\mathrm{mode}}}. \tag{5.22}$$

   This feature captures the spread in the periodicity across sub-apertures, providing robustness to outliers and noise. Analogous to spectral spread measures, widely used in audio information literature [188].

3. **Normalised Lag Entropy:** The Shannon entropy of the histogram of dominant lags:

$$f_7 = -\frac{1}{\log M} \sum_{i=1}^{M} p_i \log p_i, \tag{5.23}$$

   where $p_i$ is the fraction of sub-apertures with dominant lag equal to the $i$-th unique lag, and $M$ is the number of unique lags. Low entropy indicates that the dominant lag is consistent across sub-apertures, while high entropy indicates irregular behaviour [189].

4. **Lag Spread:** The mean absolute deviation of the dominant lag relative to the mode:

$$f_8 = \frac{1}{K} \sum_{k=1}^{K} |\ell_{\mathrm{dom}}[k] - \ell_{\mathrm{mode}}| \Big/ \ell_{\mathrm{mode}}. \tag{5.24}$$

   This feature complements the MAD and agreement ratio, providing a normalised measure of temporal dispersion in the autocorrelation structure. This feature can be considered analogous to pitch variability measures in audio [190].

These features quantify the stability and consistency of the autocorrelation across sub-apertures, effectively capturing the temporal evolution of the signal's periodic structure. By combining the sub-aperture-based features with the global structure features, the feature set is robust to amplitude variations, noise, and clutter. It is worth mentioning that, for the case of the double autocorrelation, the only change to the process is the performance of a second autocorrelation per each sub-aperture.

The resulting features are then used to train and test the SVM classifier, allowing different combinations of autocorrelation strategies (e.g., single-single, single-double,

double-double) to be evaluated. By focusing on structured features instead of raw signals, the system achieves robust generalisation across diverse signal conditions while maintaining interpretability of the extracted characteristics.

**Feature space analysis via Principal Component Analysis** To assess the discriminative structure of the proposed feature space, PCA is applied to the feature vectors derived from simulated data. PCA provides a low-dimensional representation that preserves the maximum variance of the data, enabling a qualitative evaluation of class separability in the feature space [191]. Fig. 5.5 shows the projection onto the first two principal components. These components are linear combinations of the original features, however, an inspection of the component coefficients indicates that the first principal component is primarily influenced by the agreement ratio feature, while the second principal component is mainly driven by the peak spacing consistency. This projection highlights the degree of separability between classes in the proposed feature space.

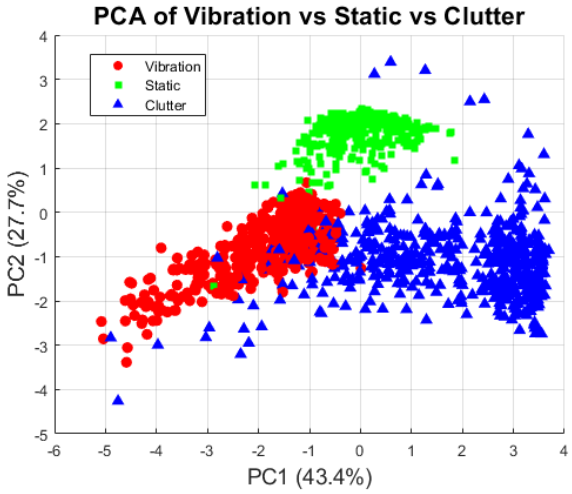


Figure 5.5: PCA projection of the feature space.

### Classification

For the classification experiments, an SVM was employed due to its effectiveness in low-dimensional, structured feature spaces and strong generalisation capability. A multiclass strategy was implemented using Error-Correcting Output Codes (ECOC) with One-vs-One (OvO) binary learners, as provided by the MATLAB Classification Learner framework [192]. The classifier operates on a three-dimensional feature space derived from the proposed descriptors, and derives the optimal hyperplane that best separate the different classes, as reflected in Fig. 5.6. The extracted features from each autocorrelation variant (e.g., single, double) are used to train two separate SVM models, to distinguish between micro-motion patterns. An SVM is chosen as it has been widely applied in many pattern recognition studies, and it can offer an efficient classification model [194–196].

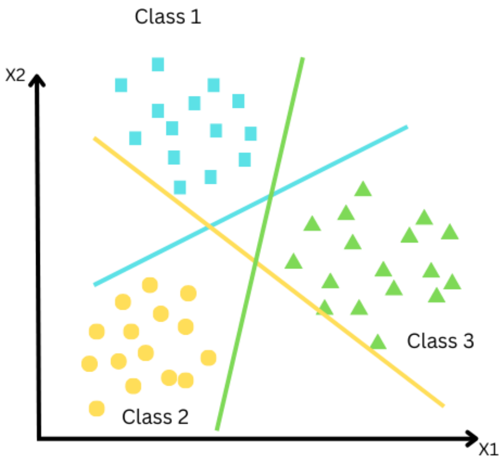


Figure 5.6: Multiclass classification based on ECOC with binary SVM, OvO multiclass strategy.

During training, features derived from each autocorrelation strategies are incorporated, allowing the classifier to learn decision boundaries that remain valid across different signal representations. During testing, features from each autocorrelation variant are independently evaluated by their corresponding trained classifier. In addition to class labels, the classifier outputs confidence scores, which are subsequently used in the final detection stage to rank and validate candidate micro-motion detections.

The analysis of different kernels for the model and the final selection is presented in 5.4.2. The impact of different autocorrelation strategies, feature combinations, and classifier performance is evaluated in Section 5.4. In Algorithm 2, a high-level pseudocode is presented to accompany the proposed classification framework.

**Algorithm 2** Classification using structure features and SVM classification

```
Initialize $count \leftarrow 0$, $detections \leftarrow []$
for $r = 1$ to length(refined_bins) do
    $count \leftarrow count + 1$
    $sig \leftarrow IM\_RD(refined\_bins[r], :)$
    $[R, \tau_{pos}, freq\_est] \leftarrow$ Autocorr_function($sig, fs, xcorr\_value, plotif$)
    $feat[count, :] \leftarrow$ extract_structure_features($R, fs$)
    $[label[count, :], score1[count, :]] \leftarrow$ predict($final\_model\_svm, feat[count, :]$)
    if $label[count, :] = final\_model\_svm.ClassNames[2]$ then
        Append to $detections$:
  range_bin $\leftarrow refined\_bins[r]$
  score $\leftarrow score1[count, 2]$
  freq $\leftarrow freq\_est$
  candBin $\leftarrow refined\_bins[r]$
    end if
end for
$detection\_final[kk] \leftarrow detections$
```

## 5.3 Synthetic Dataset Simulation

This section details the synthetic SAR acquisition geometry, the transmitted waveform, and the maritime scene configuration used to generate the SLC images for training and testing the proposed classification methodology. The goal is to simulate stripmap mode SLC images of a simple scenario containing a vibrating target and a quasi-static non-vibrating target in a cluttered environment to emulate the sea.

### 5.3.1 Simulated Acquisition Geometry and Radar Parameters

A monostatic, side-looking SAR platform is modelled as flying at a constant altitude and velocity along the azimuth (along-track) direction. The radar is located at (0,0,500) m at the start of the aperture and moves with constant speed $v_p = 100$ m/s in azimuth direction. The scene contains two point-like scatterers representing maritime targets located at fixed ground-range positions (1000,0,0) m and (1200,0,0) m, respectively. The transmitted waveform is a LFM chirp with a carrier frequency $f_c = 10$ GHz, a bandwidth of $B = 200$ MHz, and a pulse duration of $T_p = 10\ \mu s$. The PRF is set to 1 kHz, and 256 pulses are transmitted, defining a synthetic aperture in slow time of 0.256 s and a maximum unambiguous range of 150 km. The fast-time sampling frequency is $f_s = 2B$. In Fig. 5.7, a visual representation of the SAR geometry and target distribution used to create synthetic data can be seen. The platform moves along the y-axis at a constant velocity and elevation, simulating the imaging of the ocean with two targets. The first target exhibits vertical micro-motions along the z-axis and is presented as a vessel, as shown in the image, while the second target is defined as a static target.

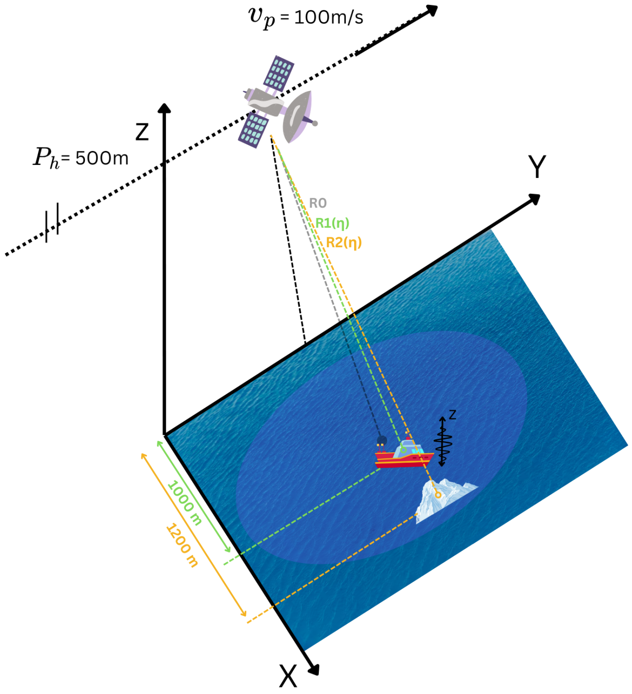


Figure 5.7: Geometry of the simulated stripmap SAR scenario. The monostatic radar platform moves along the azimuth (y) direction at constant velocity and altitude. Two point-like maritime targets are located at fixed ground-range positions.

### 5.3.2 Vibrating and static target models

To ensure robust classification performance on unseen synthetic data and real-world data, the target models incorporate realistic variability beyond just vibration presence.

Identical static targets across all realisations would lead to poor generalisation when facing diverse real objects (icebergs, rocks, buoys). Therefore, independent variations in reflectivity, phase/amplitude jitter [197], and small motion drifts are introduced to emulate the diversity of maritime targets while preserving the key discriminative feature: micro-motion signatures. This idea is supported in [198], where the authors introduce a randomisation framework to add statistical variation in target and clutter regions, to improve the generalisation of ATR models to use synthetic SAR data for training. Moreover, the limited availability of labelled SAR data to train DL models requires the generation of synthetic data for training and enabling transfer learning. In Fig. 5.8, an example of a synthetic SLC SAR image can be seen, containing the 2 classes of targets: vibrating and static, within a cluttered and noisy background. The vibrating target can be seen at around range bin 900∼1000, and can be easily recognised due to its ghost targets, caused by the phase modulation induced by the vibration. The static target is located at around range bin 1400, and as expected, appears as a bright point with no ghost targets.

The first target represents a vibrating object and undergoes sinusoidal motion along the vertical ($z$) direction according to

$$z_1(t) = z_{1,0} + A\sin(2\pi f_{\mathrm{vib}} t), \tag{5.25}$$

where the nominal height is $z_{1,0} = 0$ m, vibration amplitude $A \in [0.0002, 200]$ mm is randomly drawn uniformly, and vibration frequency $f_{\mathrm{vib}} \in [10, 100]$ Hz is also uniformly random per realisation. Its complex reflectivity magnitude is $\sigma_{\mathrm{vib}} \in [0.3, 2.5]$, with per-pulse amplitude jitter of standard deviation 20% and phase jitter of standard deviation 0.1 rad.

The second target models a quasi-static non-vessel object (e.g., iceberg) with fixed $x, y$ position and small linear range drift $v_{\mathrm{drift}} = 0.001$ m/s to emulate residual motion or platform effects:

$$R_2(t) = R_{2,0} + v_{\mathrm{drift}} t, \tag{5.26}$$

where $R_{2,0}$ is the initial slant range. Its complex reflectivity magnitude is $\sigma_{\text{static}} \in [0.5, 2.0]$, with per-pulse amplitude jitter of standard deviation 20% and phase jitter of standard deviation 0.2 rad.

It is worth noting, in the case of the vibrating object, that the range of possible vibrating frequencies is set to start at 10 Hz, which has been chosen based on the total aperture time, given by the simulator parameters in Subsection 5.3.1. The total aperture time is 0.256 s, which limits the ability to capture lower frequencies. Choosing 10 Hz as the lower bound for the simulation still represents a critical scenario, as the period of the vibration is $T = 0.1$ s, meaning that only 2 and a half vibration cycles are captured by the sensor in this case.

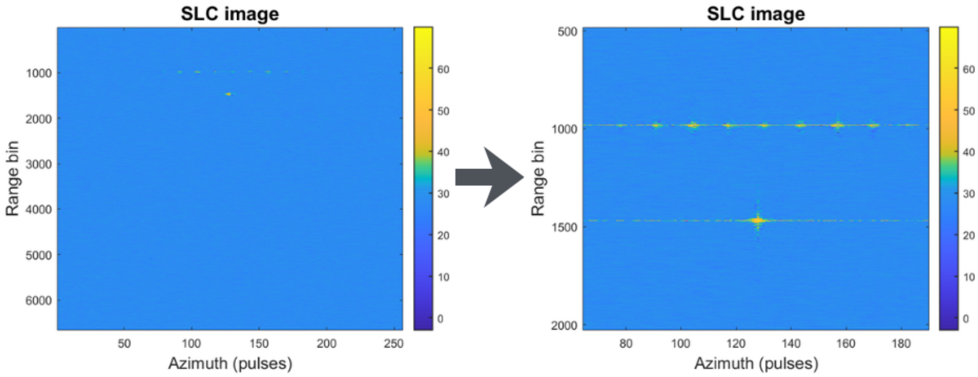


Figure 5.8: Original SLC image (left), and zoomed in image on targets (right).

### 5.3.3 Raw echo generation and noise

The raw SAR signal simulation follows the standard point-target reference model for stripmap acquisition [199, 200]. For the $n$-th pulse transmitted at slow-time $\eta_n = n/PRF$, the platform moves to position $\mathbf{r}_p(\eta_n) = [0, v_p\eta_n, 500]^T$, where $v_p = 100$ m/s defines the synthetic aperture. The instantaneous slant ranges to both targets are computed from their Euclidean distances, introducing range variation across the aperture and the corresponding Doppler phase history over the Coherent Processing Interval

(CPI) [201]. The instantaneous slant range to the vibrating target is:

$$R_1(\eta_n) = \|\mathbf{r}_p(\eta_n) - \mathbf{r}_{t1}(\eta_n)\|_2 = \sqrt{x_1^2 + (v_p\eta_n - y_1)^2 + [500 - (A\sin(2\pi f_{\text{vib}}\eta_n))]^2}, \tag{5.27}$$

where $\mathbf{r}_{t1}(\eta_n) = [1000, 0, A\sin(2\pi f_{\text{vib}}\eta_n)]^T$. And the slant range to the static target, which includes linear drift is:

$$R_2(\eta_n) = \sqrt{x_2^2 + (v_p\eta_n - y_2)^2 + 500^2} + v_{\text{drift}}\eta_n. \tag{5.28}$$

The baseband received echo from each target, after downconversion, is modelled as a time-delayed and phase-modulated replica of the transmitted chirp signal $s_{tx}(t)$, following the standard point-target SAR signal model. For the $n$-th transmitted pulse at slow time $\eta_n$, the received echoes from the vibrating and static targets are respectively expressed as

$$s_1(t,\eta_n) = \sigma_{\text{vib}} \cdot a_{\text{vib},n} \cdot e^{j\phi_{\text{vib},n}} \cdot s_{tx}\left(t - \frac{2R_1(\eta_n)}{c}\right) \cdot \text{rect}\left(\frac{t-\tau_1}{T_p}\right), \tag{5.29}$$

$$s_2(t,\eta_n) = \sigma_{\text{static}} \cdot a_{\text{static},n} \cdot e^{j\phi_{\text{static},n}} \cdot s_{tx}\left(t - \frac{2R_2(\eta_n)}{c}\right) \cdot \text{rect}\left(\frac{t-\tau_2}{T_p}\right), \tag{5.30}$$

where $\sigma_{\text{vib}}$ and $\sigma_{\text{static}}$ denote the scattering coefficients of the vibrating and static targets, respectively. The terms $a_{c,n} \sim \mathcal{N}(1, 0.2)$ model per-pulse amplitude fluctuations, while $\phi_{c,n} \sim \mathcal{N}(0, \sigma_\phi^2)$ represent random phase jitter accounting for residual phase instabilities and target-dependent scattering variability. Different phase jitter levels are used for the static and vibrating targets, with $\sigma_\phi = 0.2$ rad and $\sigma_\phi = 0.1$ rad, respectively. The two-way propagation delay is given by $\tau_c = 2R_c(\eta_n)/c$, and the rectangular window rect$(\cdot)$ enforces the finite pulse duration $T_p$, where $c \in \{\text{vib}, \text{static}\}$ denotes the target class.

The total received raw echo is obtained by coherently summing the individual target contributions,

$$r(t,\eta_n) = s_1(t,\eta_n) + s_2(t,\eta_n), \tag{5.31}$$

to which complex white Gaussian noise $w(t,\eta_n) \sim \mathcal{CN}(0, \sigma_w^2)$ is added. The noise power is adjusted to achieve a randomly selected SNR from the set $\{-10, -5, 0, 2, 5, 8\}$ dB, generating multiple independent noise realisations and enabling robustness analysis across a wide range of conditions.

### 5.3.4 Image formation and clutter model

SAR image formation processing consists of two main steps: range compression via matched filtering of the LFM chirp, and azimuth compression exploiting the Doppler phase history acquired over the synthetic aperture [200]. The simulator implements a simplified range-Doppler approach suitable for the point-target geometry, where range processing is performed independently per pulse, and azimuth focusing relies on time-domain phase compensation. A realistic sea clutter model is finally incorporated in the image domain to emulate maritime surveillance conditions.

**Range compression**: A reference chirp $s_{\mathrm{ref}}(t) = \exp(j\pi K_r t^2)$ with chirp rate $K_r = B/T_p$ is FFT'd: $S_{\mathrm{ref}}(f) = \mathrm{FFT}\{s_{\mathrm{ref}}(t)\}$. Range-compressed data is obtained via

$$R_{\mathrm{RC}}(t, \eta_n) = \mathcal{F}^{-1}\left\{\mathcal{F}\{r(t, \eta_n)\} \cdot S^{*}_{\mathrm{ref}}(f)\right\}. \tag{5.32}$$

**Azimuth focusing**: For each range bin $r$, a phase reference is computed using the static target's geometry:

$$\phi(r, \eta_n) = \exp\left(j\frac{4\pi f_c}{c} R_{\mathrm{ref}}(r, \eta_n)\right), \tag{5.33}$$

where $R_{\mathrm{ref}}(r, \eta_n) = \sqrt{r^2 + (v_p\eta_n)^2 + 500^2}$. The azimuth-compressed SLC image is

$$SLC(r, \eta_n) = \mathrm{fftshift}\left(\mathrm{ifft}\left(R_{\mathrm{RC}}(r, \eta_n) \cdot \phi(r, \eta_n)\right)\right). \tag{5.34}$$

The simulator developed is a simplified adaptation to Matlab of the simulator developed in [202].

**Clutter model**: Rayleigh-distributed speckle $A_{\mathrm{speckle}} \sim \mathrm{Rayleigh}(1)$ is spatially filtered with a 2D Gaussian kernel ($7 \times 7$, $\sigma = 2$) and multiplied by a uniform random phase to form correlated complex clutter. Clutter power is scaled to achieve a SCR of 10 dB over the target signal power and added to the SLC image. The implemented sea clutter model generates spatially correlated complex Gaussian speckle with Rayleigh-distributed amplitude, which accurately captures the fundamental statistical properties of SAR sea clutter in moderate-resolution images [203]. The Gaussian spatial filter introduces realistic texture correlation consistent with the radar resolution and sea state, while scaling to SCR = 10 dB above signal power reflects typical maritime surveillance scenarios where clutter is stronger than thermal noise but targets remain detectable [204, 205].

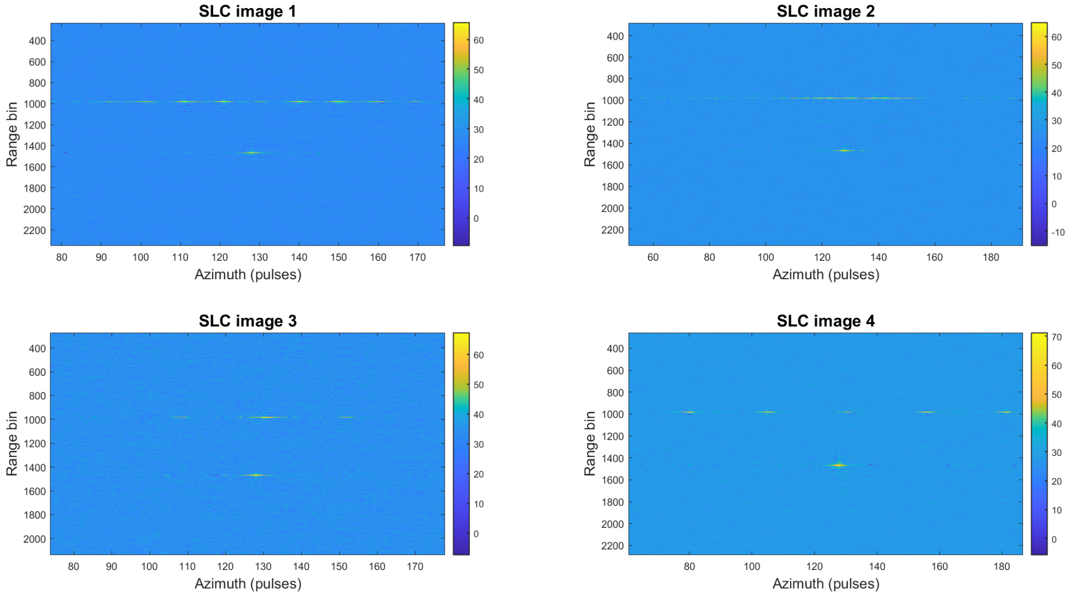
In Fig. 5.9, four synthetic SLC images are shown, each containing two point-like targets: one vibrating target and one quasi-static non-vibrating target. Although the geometric configuration remains fixed across all cases, the images exhibit noticeable differences due to variations in SNR, vibration frequency, vibration amplitude, and target reflectivity. These parameters are summarised in Table. 5.1, and commented upon in the following.

Table 5.1: Simulation parameters for the four synthetic SLC image realisations in Fig. 5.9.

| **Image** | **SNR (dB)** | $f_{\textbf{vib}}$ **(Hz)** | $A$ **(m)** | $\sigma_{\textbf{vib}}$ | $\sigma_{\textbf{static}}$ |
|---|---|---|---|---|---|
| Image 1 | 8 | 37.74 | 0.0140 | 1.4701 | 0.6949 |
| Image 2 | 5 | 15.44 | 0.0196 | 1.1222 | 0.7521 |
| Image 3 | -10 | 84.90 | 0.0023 | 0.8353 | 0.9698 |
| Image 4 | 5 | 98.69 | 0.0149 | 0.8345 | 1.5027 |

In the first image, a relatively high SNR (8 dB) combined with moderate vibration frequency (37.7 Hz) and vibration amplitude (0.014 m) produces a clear response from both targets. The vibrating target appears brighter and shows visible azimuthal phase modulation, in the form of ghost targets along the azimuth direction, indicative of micro-motion-induced Doppler modulation.

The second image is characterised by a reduced SNR (5 dB) and a lower vibration frequency (15.4 Hz) with slightly higher vibration amplitude (0.02 m). In this case, both targets remain detectable; however, the lower vibration frequency results in fewer


SLC image 1

SLC image 2

SLC image 3

SLC image 4

Figure 5.9: Four synthetic SLC images, each containing a vibrating point target and a quasi-static reference target under different imaging and motion conditions. From left to right, top to bottom: (Image 1) high SNR with moderate vibration frequency and amplitude; (Image 2) lower SNR with reduced frequency and increased amplitude; (Image 3) low SNR with high vibration frequency and very small amplitude; and (Image 4) moderate SNR with high vibration frequency and increased target reflectivity.

vibration cycles over the synthetic aperture. As a consequence, the micro-Doppler-induced effects in the vibrating target are less pronounced, leading to a more compact azimuth response compared to the first case.

The third image represents the most challenging scenario, with a low SNR of -10 dB and a small vibration amplitude (0.0023 m). Although the vibration frequency is high (84.9 Hz), the combination of strong noise and weak micro-motion amplitude significantly degrades the visibility of the vibration signature. The target responses are partially masked by noise and clutter, and the vibrating target becomes difficult to distinguish from the static target based solely on visual inspection of the SLC image.

In the fourth image, the SNR returns to a moderate level (5 dB), while the vibration frequency is high (98.7 Hz) and the vibration amplitude is comparable to that of the first image (0.012 m). Despite the vibrating target having lower reflectivity than the static target in this case, clear micro-motion-induced azimuthal modulation is again visible. The higher vibration frequency produces a more structured azimuth response for the vibrating target relative to the static one.

Overall, the observed differences across the four SLC images illustrate how SNR, vibration amplitude, vibration frequency, and target reflectivity jointly influence the appearance of vibrating and non-vibrating targets. Higher SNR and larger vibration amplitude enhance the visibility of micro-motion effects, while low SNR and weak vibrations can obscure these signatures, motivating the need for robust feature extraction and classification techniques.

## 5.4 Classification Analysis on Synthetic data

This section presents the evaluation of the proposed vibration detection and estimation framework, evaluating the feature extraction processing and different configurations. In particular, the impact of single and double autocorrelation representations in classification, and the effect of candidate refinement, is herein analyse. In this section, the goal is to identify the configuration that offers the best trade-off between detection reliability, frequency estimation accuracy, and classification performance.

### 5.4.1 Synthetic Dataset

To train and validate the classification pipeline, a synthetic dataset of SLC SAR images is generated using a Monte Carlo simulation framework based on the signal model described in Section 5.3. In this framework, each image corresponds to one independent Monte Carlo realisation obtained by randomly drawing the physical parameters.
A total of 600 stripmap SLC images were generated, each image containing:

- One vibrating target with vertical sinusoidal micro-motion along the $z$-axis, with amplitude $A$ randomly drawn from $[0.02, 200]$ mm and vibration frequency $f_{\text{vib}}$ randomly drawn from $[10, 100]$ Hz per realisation.
- One quasi-static target with small linear range drift to emulate residual motion,

with randomised reflectivity and per-pulse amplitude/phase jitter.

The generated images incorporate spatially correlated Rayleigh-distributed speckle with an SCR of 10 dB, and independent noise realisations drawn from complex Gaussian distributions to simulate varying SNR conditions. This synthetic dataset ensures a sufficient diversity of target micro-motion signatures and environmental conditions, enabling robust training and evaluation of the classifier. Furthermore, a new set of 100 images was created to be used as test data for the entire pipeline (detection-feature extraction-classification).

#### Evaluation Metrics

The trained SVM model is evaluated on a held-out test set, which is not used during training or cross-validation. Performance is quantified using overall classification accuracy as well as class-wise precision, recall, and F1-score, computed from the confusion matrix.

### 5.4.2 Training, Validation, and Testing

The dataset was split into training and testing sets using a 70/30 ratio. Specifically:

- **Training set:** 420 images used to optimise the SVM classifier parameters.
- **Testing set:** 180 images held out to evaluate classification performance on unseen data.

A multi-class linear SVM classifier is trained using the ECOC framework, and feature normalisation is applied to ensure comparable scaling. To choose the most suitable kernel for the better separation of the classes based on the extracted features, an analysis and comparison is carried out in the following.

**Cross-validation and kernel selection** To assess model stability and mitigate overfitting, a $k$-fold cross-validation strategy with $k = 5$ is performed on the training set, as seen in Fig. 5.10, for three different SVM kernels: linear, Gaussian RBF, and polynomial. For each training fold, hyperparameter tuning is performed using an internal grid search. Specifically, for the Gaussian RBF kernel, the box constraint $C$ and kernel scale $\sigma$ are optimised, while for the polynomial kernel, the box constraint $C$, kernel scale, and polynomial order are jointly tuned. The optimal parameter combination is selected based on the highest cross-validated classification accuracy within each fold, performed with the *SVMtemplate* library on Matlab. This nested procedure ensures that model evaluation is not biased by fixed hyperparameter choices and provides a more reliable estimate of generalisation performance. The cross-validated classification accuracy is computed as the complement of the average misclassification loss across folds, and it helps to ensure that the model is robust when it is fed unseen data [206]. Cross-validation is performed across the three kernels and individual performances are evaluated to make the best choice.

The cross-validated accuracy and per-class metrics were compared to select the most robust kernel. As shown in Table 5.2, the linear kernel consistently achieves the highest cross-validated accuracy for both Xcorr1 and Xcorr2 features, indicating that the extracted structure features are linearly separable.

Table 5.2: 5-fold cross-validation accuracy (%) for different kernels.

| Kernel | Xcorr1 | Xcorr2 |
|---|---|---|
| Linear | 99.44 | 99.37 |
| RBF | 98.81 | 98.10 |
| Polynomial | 99.05 | 98.57 |

Nonlinear kernels (RBF, polynomial) do not improve generalisation and slightly underperform.

The linear SVM demonstrates balanced performance across all classes, with very high precision and recall, as can be seen in Table 5.3, confirming the robustness of the selected features and their suitability for linear learners. The confusion matrices in Fig. 5.11 further illustrate the performance achieved by the linear SVM in cross-validation. Misclassifications are minimal, and vibrating samples are almost never misclassified, specially in the case of Xcorr2 features.

The combination of high cross-validated accuracy, balanced per-class performance,

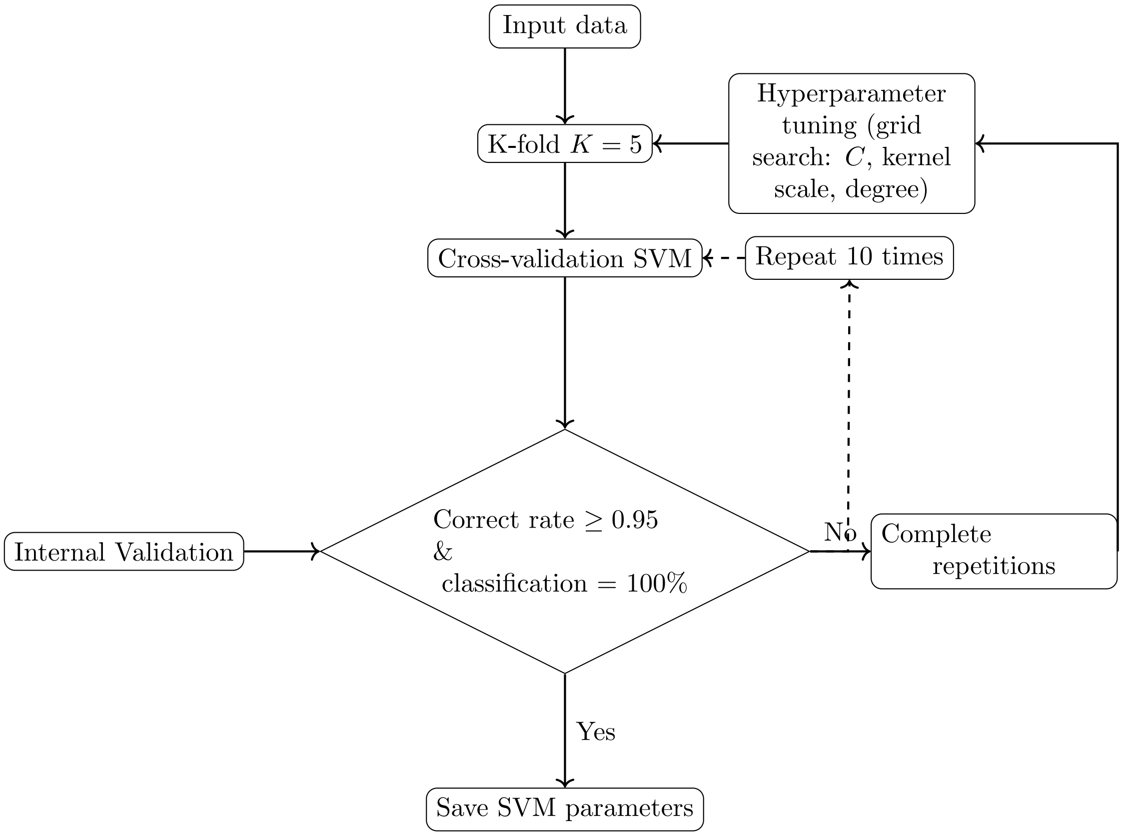
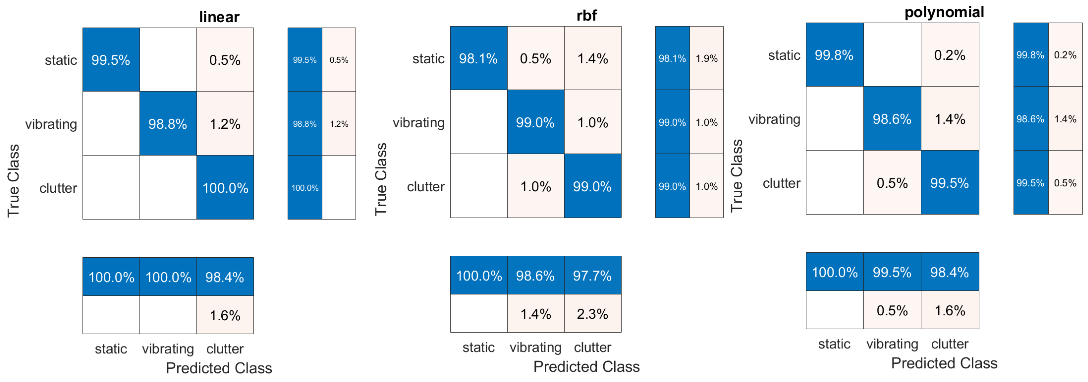


Figure 5.10: SVM training and evaluation workflow using $K$-fold cross-validation ($K = 5$) and nested hyperparameter tuning via grid search. For the Gaussian (RBF) kernel, the box constraint $C$ and kernel scale $\sigma$ are optimised; for the polynomial kernel, $C$, kernel scale, and polynomial order are tuned. Performance is evaluated across folds and the best parameter set is selected based on cross-validated accuracy.

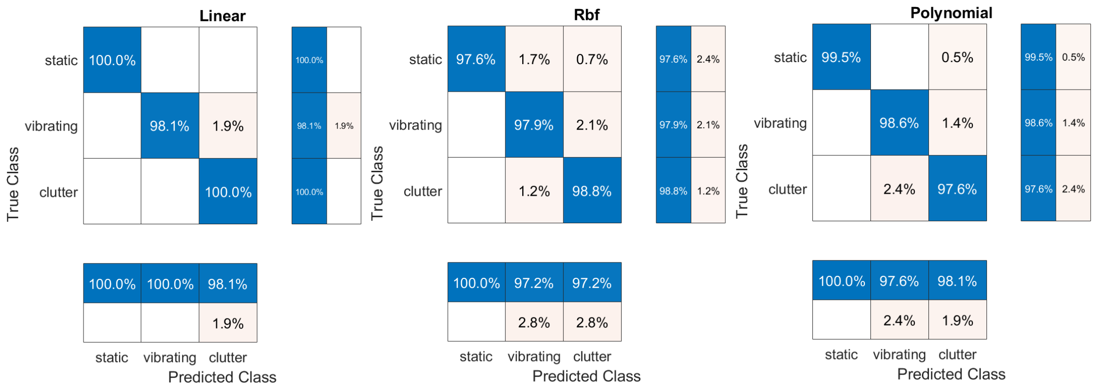
Table 5.3: Per-class precision, recall, and F1-score (%) for linear SVM on the hold-out validation set.

| | Xcorr1 | | | Xcorr2 | | |
|---|---|---|---|---|---|---|
| Class | Precision | Recall | F1 | Precision | Recall | F1 |
| Static | 100.0 | 99.44 | 99.72 | 100.0 | 100.0 | 100.0 |
| Vibrating | 100.0 | 98.33 | 99.16 | 100.0 | 99.44 | 99.72 |
| Clutter | 97.83 | 100.0 | 98.90 | 99.45 | 100.0 | 99.72 |

and minimal misclassifications demonstrates that the linear kernel is a good choice for these structure-based features. Nonlinear kernels do not provide additional benefits, since the features are already linearly separable. Furthermore, a linear SVM ensures



(a) Xcorr1 - Cross-validation



(b) Xcorr2 - Cross-validation

Figure 5.11: Confusion matrices for the cross-validation and hold-out validation of the linear SVM trained on Xcorr1 and Xcorr2. Values are row-normalised percentages.

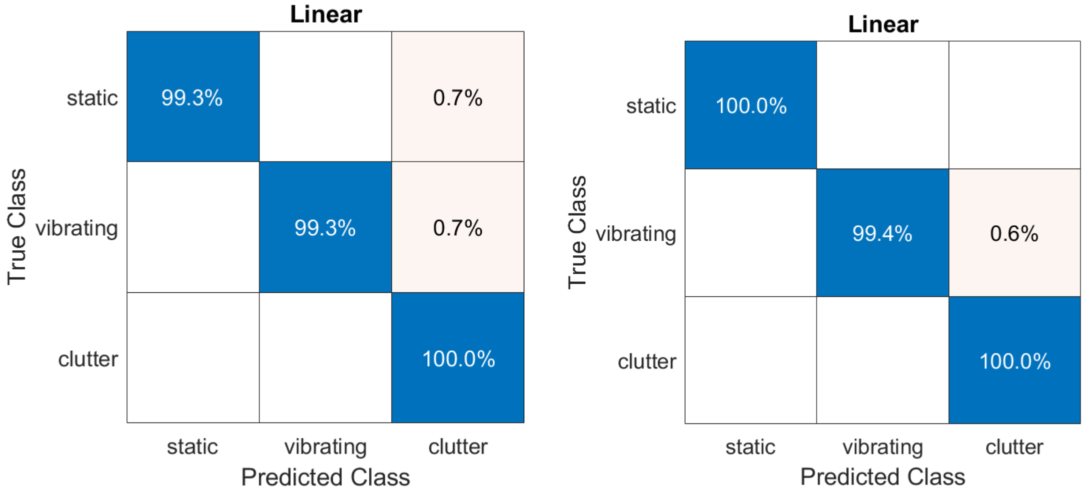
robustness, interpretability, and faster computation, making it particularly suitable for real-time or large-scale applications [207].

Following cross-validation, a final SVM model is trained and validated on the full training dataset using the following hyperparameters [208]:

- Kernel function: a linear decision boundary is selected in the feature space for the separation of the different classes;
- Feature standardisation: before training, the feature vectors are normalised to zero mean and variance 1, to ensure a fair comparison across samples;
- Multiclass strategy: an OvO strategy for multiclass classification where each pair of classes is treated as a binary problem to give the final decision;

- Class labels: class labels are selected as [0, 1, -1], corresponding to static, vibrating, and clutter, respectively.

Hold-out validation on unseen test data confirms this choice. The results are shown in Fig. 5.12, where the confusion matrices derived from the held-out validation are shown. For single autocorrelation, misclassification of vibrating targets is 0.7% of all samples in the test, corresponding to a recall of 0.993 and an F1-score of 0.997, while for double autocorrelation this number falls to 0.6%, with a recall of 0.994 and an F1-score of 0.997. For both sets of features, the classification performance improves in unseen data. The cross-validation and hold-out validation results indicate stable training behaviour and good class discrimination.



(a) Hold-out Validation Xcorr1

(b) Hold-out validation Xcorr2.

Figure 5.12: Confusion matrices for the held-out validation of the SVM trained on Xcorr1 (a), and Xcorr2 (b).

### Controlled testing using Phase-Change Regions

After testing the model with the data created with randomised parameters for generalisation, a new controlled test is proposed here. The model is evaluated using new test datasets in which the phase-change, the product between vibration amplitude and frequency ($A \cdot f_{vib}$), varies within a defined range of values.

This parameter directly influences the magnitude of phase modulation induced by the vibration of the target, and thus it determines the visibility of micro-Doppler in the image. By dividing the test into three levels: low, medium, and high phase change, the classification performance can be further analysed as a function of vibration detectability [209].

With this strategy, the goal is to assess the detection limits, evaluate the classifier performance, and the identification of phase-change thresholds for vibration identification in the cluttered environment.

**Phase-Change Regimes and Amplitude Limits** The strength of vibration-induced modulation in radar signals is given by the phase variation present on the received signal due to target micro-motion. For a monostatic radar system, the instantaneous phase of the received signal can be written as

$$\phi(t) = \frac{4\pi}{\lambda} R(t), \tag{5.35}$$

where $\lambda$ is the radar wavelength and $R(t)$ is the time-varying range between the radar and the target. For a sinusoidally vibrating target with amplitude $A$ and vibration frequency $f_{vib}$, the range variation is modeled as

$$R(t) = R_0 + A \sin(2\pi f_{vib} t), \tag{5.36}$$

which leads to a maximum phase deviation given by

$$\Delta\phi_{\max} = \frac{4\pi}{\lambda} A. \tag{5.37}$$

Although the phase deviation depends directly on the vibration amplitude, the temporal structure and detectability of the resulting modulation also depend on the vibration frequency. To capture the combined influence of amplitude and frequency on vibration observability, phase-change regimes can be characterised in terms of a phase-change modulation index $mod_{idx}$

$$mod_{idx} = \frac{4\pi A f_{vib}}{\lambda} \tag{5.38}$$

to characterise the rate at which phase modulation is introduced into the slow-time signal and is related to vibration strength [209]. Phase-change regimes can be divided into three: where $mod_{idx} \ll 1$ produces negligible micro-Doppler, $mod_{idx} \sim 1$ produces observable sidebands, and $mod_{idx} \gg 1$ produces strong nonlinear effects requiring advanced time-frequency analysis [209].

**Definition of Phase-Change Regions** Given the vibration parameter bounds used in the simulation, $A \in [A_{\min}, A_{\max}]$ and $f_{vib} \in [f_{\min}, f_{\max}]$, the phase-change parameter, $P = Af_{vib}$, spans the interval

$$P_{\min} = A_{\min} f_{\min}, \qquad P_{\max} = A_{\max} f_{\max}. \tag{5.39}$$

Within this interval, three phase-change regimes are defined (in m/s):

$$\text{Low phase change:} \quad P \in [0.0002,\ 0.2], \tag{5.40}$$

$$\text{Medium phase change:} \quad P \in [0.02,\ 2], \tag{5.41}$$

$$\text{High phase change:} \quad P \in [2,\ 20]. \tag{5.42}$$

These regimes are logarithmically separated to reflect the nonlinear relationship between phase modulation strength and vibration detectability. The lower regime corresponds to weak modulation near the noise and clutter floor, while the higher regimes represent increasingly dominant vibration-induced phase behavior.

**Amplitude Limits** For a given phase-change regime $P \in [P_{\min}, P_{\max}]$ and vibration frequency bounds $[f_{\min}, f_{\max}]$, the corresponding feasible amplitude range is given by

$$A_{\min}^{(P)} = \frac{P_{\min}}{f_{\max}}, \qquad A_{\max}^{(P)} = \frac{P_{\max}}{f_{\min}}. \tag{5.43}$$

Applying this relationship yields the following amplitude limits for each phase-change regime:

- **Low phase-change regime**

  $$A \in \left[\frac{0.0002}{100}, \frac{0.2}{10}\right] = [2 \times 10^{-6},\ 2 \times 10^{-2}]\ \text{m}, \tag{5.44}$$

  corresponding to vibration amplitudes of approximately 0.002–20 mm.

- **Medium phase-change regime**

  $$A \in \left[\frac{0.02}{100}, \frac{2}{10}\right] = [2 \times 10^{-4},\ 0.2]\ \text{m}, \tag{5.45}$$

  corresponding to amplitudes of approximately 0.2–200 mm.

- **High phase-change regime**

  $$A \in \left[\frac{2}{100}, \frac{20}{10}\right] = [2 \times 10^{-2},\ 2]\ \text{m}. \tag{5.46}$$

  In practice, this range is limited by an amplitude constraint $A \leq A_{\text{max}} = 200$ mm, resulting in an upper-bounded effective amplitude range of approximately $20 -$ 200 mm, to maintain physically meaningful results for micro-motions.

### Results of controlled test

The classifier performance was evaluated across the three phase-change regimes (low, medium, high), and the full range regime. Table 5.4 summarises the accuracy, precision, recall, and F1-score for each model-test combination. In Appendix A more detailed tables can be found where precision, recall, and F1-score are shown for every train/test combination and every class. As the main objective is vibration detection, particular attention is given to the recall of the vibrating class.

**Performance in Single Phase-Change Regimes** When trained and tested within the same phase-change regime, the classifier achieves near-perfect performance across all classes.

- **Low phase-change regime:** Both Xcorr1 and Xcorr2 features yield 100% accuracy, with perfect precision, recall, and F1-score for vibrating, static, and clutter classes. This demonstrates that the classifier can detect even weak vibration-induced micro-Doppler signatures when the test data and training data are both in the low regime of phase-change values.
- **Medium phase-change regime:** Xcorr1 achieves perfect performance, while Xcorr2 shows a slight decrease (accuracy 99.67%) due to minor misclassifications in the static and clutter classes. This suggests that Xcorr2 is slightly more sensitive to moderate phase-change variations, smoothing out an already weak vibrating signature.
- **High phase-change regime:** Both feature sets maintain 100% accuracy, reflecting the strong micro-Doppler signatures and easy detectability of higher phase-change vibrations.

**Full-Range Training Evaluated on Individual Regimes** When the classifier is trained on the full range of phase-change values and tested on individual regimes, it maintains high performance (Table 5.4):

- For low phase-change data, accuracy slightly decreases to 99.67% due to a small number of misclassifications, indicating that low-amplitude vibrations are more challenging when the classifier has also been exposed to higher-amplitude samples.
- Medium and high phase-change test sets are classified perfectly, showing robustness of the model to strong vibration-induced micro-Doppler signatures even when trained across the full range.

**Feature Concatenation Across Full Range** Finally, training and testing on concatenated features from Xcorr1 and Xcorr2 across the full range yields an overall accuracy of 99.33%. Slight reductions in precision and recall occur for the vibrating and static classes, reflecting the increased complexity of the combined feature space. Precision and recall for the static class are 99.33% and 98.67%, respectively, and 98.68% and

99.33% for the vibrating class. These results are summarised in Table A.6 in Appendix A. However, the model remains highly effective, demonstrating that feature fusion preserves performance while enabling a single classifier to handle the entire phase-change spectrum.

Table 5.4: Classification accuracy (%) across phase-change regimes for different training and testing configurations.

| **Training** | **Testing** | **Low** | **Medium** | **High** |
|---|---|---|---|---|
| Xcorr1 | Xcorr1 | 100 | 100 | 100 |
| Xcorr2 | Xcorr2 | 100 | 99.67 | 100 |
| Full-range Xcorr1 | Xcorr1 | 99.67 | 100 | 100 |
| Full-range Xcorr2 | Xcorr2 | 99.67 | 100 | 100 |
| [Xcorr1, Xcorr2] | [Xcorr1, Xcorr2] | 99.33 | 99.33 | 99.33 |

**Controlled Test Findings** These results point out the following:

1. The classifier can reliably detect vibrations across all phase-change regimes when trained and tested consistently.
2. Low phase-change signals are the limiting factor for detection; minor misclassifications appear when training spans the full range.
3. Feature concatenation enables a unified model for the entire phase-change spectrum with minimal power consumption increase and minimal loss in performance.

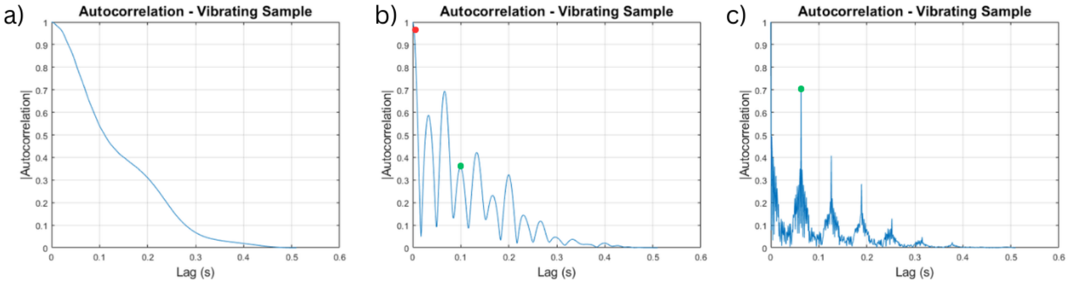
Overall, the controlled testing confirms that the model effectively captures the relationship between phase-change strength and micro-Doppler visibility, establishing vibration detectability under different amplitude and frequency combinations. Furthermore, when misclassifications across the different configurations are analysed, a pattern can be detected where they occur primarily in low-amplitude or low phase-change samples. For example, in the medium phase-change regime (Xcorr2 $\rightarrow$ Xcorr2), the following can be observed:

- **Vibrating target misclassified as static:** Sample index 57, estimated frequency = NaN Hz, true frequency = 44.962 Hz, amplitude = 0.002 m, SNR = 5 dB, phase-change = 0.093 m/s.
- **Correctly classified vibrating targets:** 99 samples, RMSE = 68.862 Hz, bias = 10.532 Hz.
- **Highest error among correctly classified:** Sample index 40, estimated frequency = 500 Hz, true frequency = 10.587 Hz, amplitude = 0.123 m, SNR = 0 dB, phase-change = 1.302 m/s.
- **Lowest error among correctly classified:** Sample index 22, estimated frequency = 15.873 Hz, true frequency = 15.886 Hz, amplitude = 0.116 m, SNR = 5 dB, phase-change = 1.841 m/s.

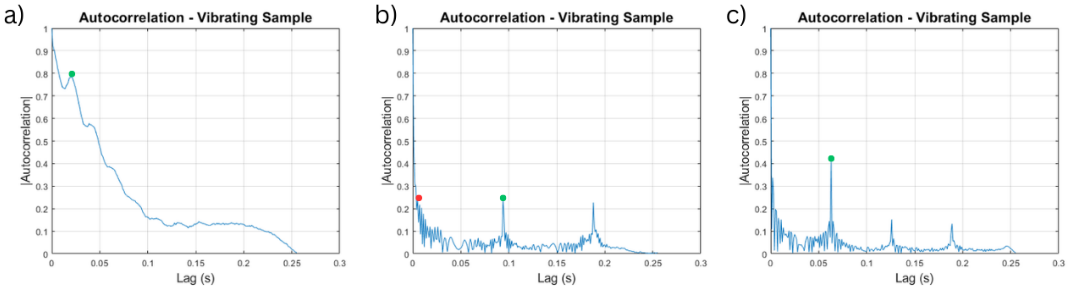
In Fig. 5.13, the examples mentioned above are illustrated, and the difference between applying single and double autocorrelation can be seen.

In Fig. 5.13.1, the autocorrelation signature for the misclassified vibrating sample (a), the correctly classified sample with wrong estimated vibration frequency (b), and the correctly classified and estimated frequency of the vibration (c), are shown. It can be noticed that a) resembles a static signature, and so the misclassification is justified. In b), the signature exhibits a more vibrating-like pattern, but the algorithm identifies the wrong peak for the frequency estimation, represented by a red dot in the image. The identified autocorrelation lag in this case is 0.002 s, which, in turn, returns a vibration frequency of 500 Hz, far from the correct frequency of 10.587 Hz, represented by the green dot at $\sim$ 0.1 s lag. Finally, a correctly classified sample with correct frequency estimation is shown in Fig. 5.13.1 c), with the correct lag pointed out by the green dot at $\sim$ 0.094 s.

Comparing the double autocorrelation results with single autocorrelations, a few key differences can be spotted. Fig. 5.13.2 correspond to the same samples as in Fig. 5.13.1, with the only difference that they have undergone single autocorrelation instead of double. In this case, the three samples are correctly classified, and a) and c) have a good frequency estimation. In a) the estimated frequency is ∼ 47 Hz, and in c) the estimation is again very close to the true number, 15.886 Hz. Sample b) is also correctly classified, but the estimated frequency is still far from the real number, with the resulting estimation being 200 Hz, instead of the true 10.587 Hz. Again, the correct and incorrect frequency estimations are shown by a red and a green dot, respectively.



(1) Double autocorrelation representation.



(2) Single autocorrelation representation.

Figure 5.13: Autocorrelation images of three representative samples in the medium phase-change regime: a) misclassified vibrating target, b) correctly classified vibrating signal with wrong estimated vibration frequency, c) correctly classified target with lowest frequency estimation error.

These results highlight the observation that the double autocorrelation needs to be applied with caution, as it can flatten the signal loosing its vibrational signature. The misclassified sample's vibrating signature is too weak to overcome the double autocorrelation processing. Although the purpose of this section does not focus on the accurate estimation of the vibration frequency, it is worth noticing that some samples are correctly classified, but the estimation of the frequency is not correct. Low SNR can make the micro-Doppler signature noisy, and although the classifier can detect the presence of vibration in general, accurately measuring its properties becomes more difficult. In summary, the errors in frequency estimation reveal that:

- High errors are associated with low SNR, low amplitude, or small phase-change, even when the target is correctly classified.
- Low errors correspond to samples with higher phase-change and sufficient amplitude/SNR.
- No vibrating target is misclassified as clutter in any tested regime.

The controlled experiments are useful to highlight how double and single autocorrelation impact classification and frequency estimation, highlighting the algorithm's sensitivity to signal characteristics and noise. Fig. 5.14 summarises the results of the controlled tests for the different feature configurations across phase-change regimes.

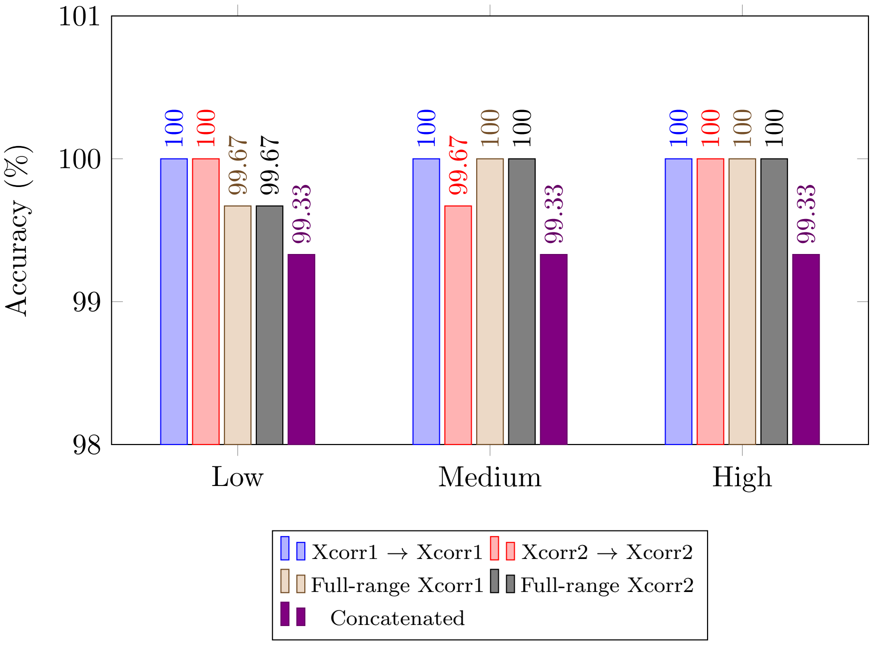


Figure 5.14: Classification accuracy across phase-change regimes for different feature sets and training configurations.

## 5.5 Full Pipeline Configurations analysis

To assess the influence of the autocorrelation representation and the generalisation capability of the classifier, two training-testing configurations were evaluated:

- Case A: Single autocorrelation used for both training and testing;
- Case B: Double autocorrelation used for both training and testing.

And each configuration was evaluated under two range bin candidate selection strategies:

1. Refined detection, where CFAR detection is followed by STFT-based refinement before the feature extraction step, as described in 5.2.2;
2. A second selection strategy is also analysed, to study if the CFAR step could be replaced by an STFT-only approach, where vibration candidates are selected directly based on the STFT analysis. When the STFT detection method is not paired with the CFAR detection step, it becomes very power and time-consuming if the area where the vibrating target is expected is not known beforehand. Nevertheless, results will be shown where the RD SLC image was cropped around the expected vibration to speed up this process. Otherwise, suppressing the CFAR method is not advisable.

In Fig. 5.15, both pipelines are shown, with the full pipeline CFAR+STFT and the STFT-only (skipping CFAR and contiguous range bin averaging) represented by a solid arrow and a dashed arrow, respectively.

The evaluation is performed on the full range synthetic data, and for each case, performance is evaluated in terms of range detection accuracy, frequency estimation accuracy, mean frequency error, frequency accuracy over true vibrating range bins only, and standard classification metrics (precision, recall, and F1-score).

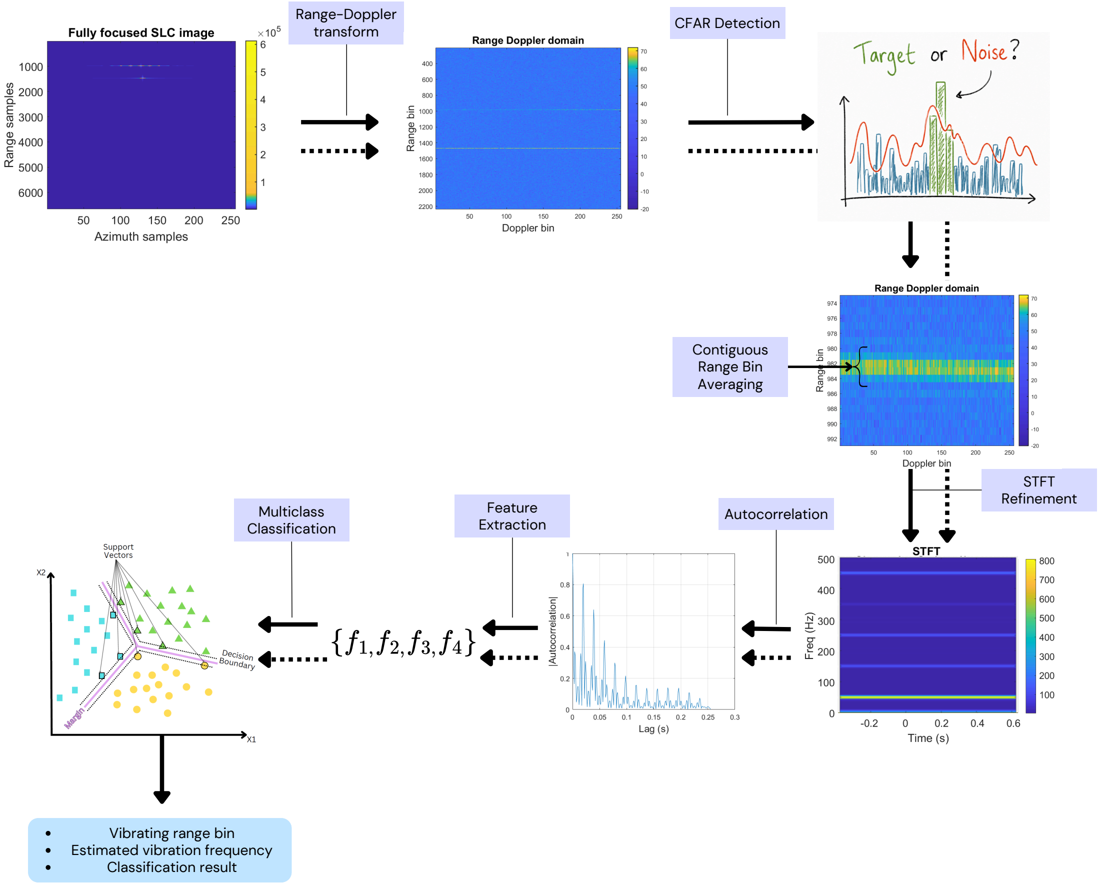


Figure 5.15: Block diagram illustrating the two processing pipelines considered in this work: the full CFAR+STFT pipeline (solid arrow) and the STFT-only pipeline (dashed arrow).

### 5.5.1 Experimental results on synthetic data

The results obtained from all evaluated configurations on full range synthetic data built from randomised parameters, as described in Sections 5.3, and 5.4.1, are presented herein. Table 5.5 summarises the performance metrics across the two cases (A and B) for both the refined pipeline (CFAR+STFT) and the STFT-only approach, respectively. The reported metrics include range detection accuracy, estimated frequency accuracy, mean frequency error, frequency accuracy over true bins, mean frequency error for true bins, autocorrelation configuration, precision, recall, and F1-score.

Table 5.5 provides a direct comparison of detection and estimation performance across different candidate selection strategies (CFAR+STFT and STFT-only) and fea-

ture representations (Case A and case B). It highlights the outcomes for each configuration, enabling the assessment of the impact of candidate refinement and autocorrelation order (single or double). When CFAR is combined with STFT, a clear improvement is observed from case A to case B, where case B significantly improves range detection accuracy (from 73.25% to 92.27%). In this case the mean frequency estimation error is also reduced by more than 60 Hz, which indicates that double autocorrelation enhances the robustness of both detection and frequency estimation in the presence of noise.
When STFT is applied alone in the detection pipeline, both case A and case B improve their range bin detection accuracy, exciding 97%, which suggests that STFT can be used alone for a reliable range location. However, as mentioned before, this can be a very time comsuming process when a smaller area in the SAR image cannot be identified before processing (problem that CFAR solves in the proposed pipeline). The double autocorrelation (case B) again reduces the mean frequency error for the true bins, but the frequency accuracy remains lower than case A. This holds for both CFAR+STFT and STFT alone, and has been analysed in Fig. 5.13, where the same targets are shown after single and double autocorrelation.
Across all configurations, the classification metrics remain consistently high, with precision, recall, and F1-scores above 0.93. Case B with CFAR+STFT achieves the highest overall classification performance, reaching perfect precision and an F1-score of 0.99. Overall, these results show that incorporating double autocorrelation improves frequency estimation accuracy and classification robustness, particularly when combined with CFAR+STFT detection preprocessing.

**Impact of Candidate Refinement** Across all cases, the refined pipeline consistently improves classification performance and frequency estimation accuracy, compared to the STFT-only approach. While STFT-only detection achieves very high range detection accuracy (up to 98.67%), it is observed to result in significantly larger mean frequency errors. As mentioned before, although the STFT-only approach has a good range bin detection accuracy, it requires a significant previous reduction of the range dimensionality, otherwise the process becomes too time-consuming, and the

Table 5.5: Results table for case A (Xcorr1) and B (Xcorr2) for different combinations of the pipeline.

| **Metric** | **Case A CFAR+STFT** | **Case B CFAR+STFT** | **Case A STFT** | **Case B STFT** |
|---|---|---|---|---|
| **Range detection accuracy (%)** | 73.25 | **92.27** | 97.53 | 98.67 |
| **Estimated frequency accuracy (%)** | 64.58 | **75.90** | 79.98 | 77.40 |
| **Mean frequency error (Hz)** | 74.46 | **9.74** | 36.19 | 21.41 |
| **Frequency accuracy - true bins (%)** | 85.33 | **78.00** | 80.00 | 73.00 |
| **Mean frequency error - true bins (Hz)** | 16.09 | **2.35** | 19.44 | 11.62 |
| **Autocorrelation configuration** | Xcorr1/Xcorr1 | **Xcorr2/Xcorr2** | Xcorr1/Xcorr1 | Xcorr2/Xcorr2 |
| **Precision** | 0.98 | **1** | 0.93 | 0.93 |
| **Recall** | 0.97 | **0.97** | 0.93 | 0.93 |
| **F1-score** | 0.98 | **0.99** | 0.93 | 0.93 |

possibility of an on-board deployment becomes unfeasible.

In contrast, the refined approach reduces the number of candidate bins by focusing on the energy peaks in the RD image, detected by the CFAR algorithm. Although this refinement slightly reduces the overall range detection accuracy in some cases, it leads to evident lower mean frequency errors and higher precision and recall values, indicating improved detection confidence and estimation robustness.

**Effect of Autocorrelation Order** The choice of autocorrelation representation has a noticeable impact on both frequency estimation and classification performance. Both single and double autocorrelation result in high F1-scores, indicating effective discrimination between vibrating and static targets.

However, the double autocorrelation representation consistently achieves lower mean frequency errors, enhanced when combined with the refined candidate selection strategy. This suggests that higher-order autocorrelation can result in a more stable feature representation for frequency estimation.

**Frequency Estimation Accuracy and Detection** The results demonstrate that estimation accuracy is strongly influenced by the preceding detection and refinement stages. The refined pipeline selectively operates on range bins with higher signal-to-noise ratios and stronger spectral coherence, hence improving the quality of the input signals used for frequency estimation. As a result, significantly lower frequency errors are achieved.

This observation highlights the connection between detection and estimation stages in practical radar signal processing systems and emphasises the importance of candidate selection for reliable vibration characterisation.

**Selection of the Optimal Configuration** Based on the combined evaluation of detection accuracy, frequency estimation performance, and classification reliability, the refined pipeline using double autocorrelation for both training and testing (Case B refined) is identified as the optimal configuration for the simulated data. This setup achieves the highest aggregated F1-score (0.985), perfect precision, and the lowest mean frequency error over true vibrating range bins (2.35 Hz), while maintaining high range detection accuracy.

## 5.6 Experimental Validation on Real SAR Data

Synthetic SAR datasets provide a controlled environment to develop and validate the proposed vibration-based detection and classification methodology. The classifier was trained exclusively using simulated SAR data, where vibration characteristics and scene parameters are fully controlled. However, real SAR acquisitions introduce additional challenges, including speckle variability, clutter heterogeneity, imperfect calibration, and unknown target characteristics. Therefore, validating the proposed pipeline on real SAR imagery presents a cross-domain validation scenario and is a crucial step toward assessing its practical applicability.

Vessels and icebergs can exhibit similar backscatter intensity and spatial extent in focused SAR images, making discrimination based only on amplitude or shape unreliable. This ambiguity is especially critical in regions such as Greenland, where icebergs are abundant. Being able to distinguish between an iceberg and an operating vessel may be crucial in areas where maritime traffic is highly restricted, and early identification of vessel activity enables timely responses to potentially unauthorised operations, contributing to maritime safety and security.

The objective of this section is to evaluate the performance of the proposed detection-location-classification pipeline when applied to real SAR data. Using real SAR images, including TerraSAR-X, Umbra, Capella, and ICEYE acquisitions, the analysis aims to:

- assess the robustness of the CFAR-based detection under real clutter conditions;
- examine whether vibration-induced micro-motion signatures can be reliably extracted from range-focused SAR data extracted from SLC complex data;
- evaluate the ability of the trained classifier to discriminate between vibrating and static objects.

Known targets within the studied scenes are used as reference cases for the evaluation of the methodology, including controlled acquisitions of a corner reflector mounted on a vibrating shaker, and active vessel measurements. These acquisitions were scheduled by the Sensor Signal Processing & Security (SSP&S) lab at Strathclyde University. Moreover, SAR images around Greenland are used to evaluate iceberg detectability. This analysis focuses on demonstrating the feasibility, consistency, and physical interpretability of the proposed approach when applied to real-world SAR data and realistic operational constraints.

### 5.6.1 Dataset description

The real data experiments are conducted on a set of different sensors and different acquisition modes. The selected images include:

- Known vibrating targets, extracted from controlled acquisitions where the ground truth is available;

- Static targets, including icebergs located off the coast of Nuuk, Greenland, as representative non-vibrating targets.

All scenes are processed in SLC format, and the acquisition parameters such as range and azimuth resolution, pixel spacing, bandwidth, and PRF, among others, vary across sensors and imaging modes. Due to the large size and high data volume of SLC SAR images, the only preprocessing step applied prior to the proposed processing pipeline is a spatial cropping around the area of interest. Table 5.6 summarises the real SAR acquisitions used for validation, including information on the sensor platform, radar mode, spatial resolution, the target in the scene, and the source of the ground truth used to locate the targets and validate the results after processing.

Table 5.6: Overview of the real SAR acquisitions used to evaluate the proposed detection and classification pipeline, showing sensor characteristics, acquisition modes, spatial resolution, known targets, and ground-truth sources.

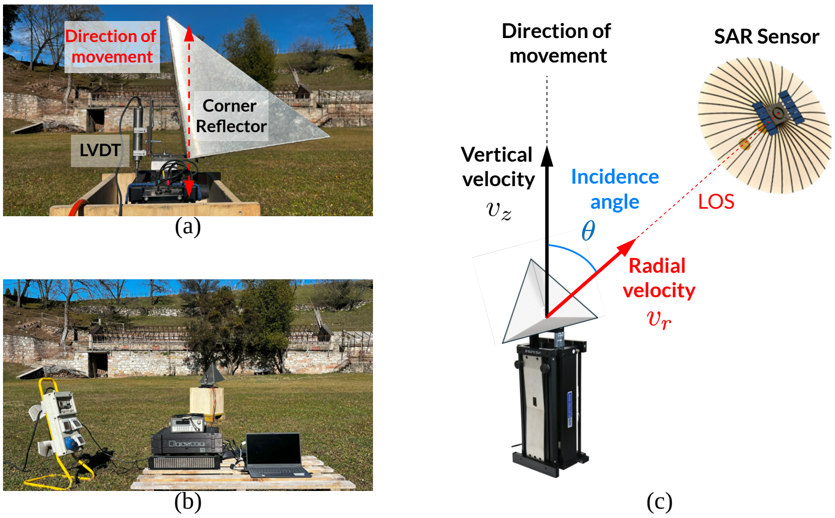
| **Scene ID** | **R1** | **R2** | **R3** | **R4** | **R5** | **R6** |
|---|---|---|---|---|---|---|
| **Sensor** | TerraSAR-X | Umbra-08 | Umbra-04 | Umbra-05 | Capella Space | Capella Space |
| **Location** | Trento (Italy) | Galway (Ireland) | Glasgow (UK) | Glasgow (UK) | Nuuk (Greenland) | Greenland |
| **Acquisition Mode** | StripMap | Spotlight | Spotlight | Spotlight | Spotlight | Spotlight |
| **Polarisation** | V | H | V | V | H | H |
| **Resolution Range x Azimuth (m)** | 0.58 x 0.23 | 0.94 x 0.32 | 0.44 x 0.21 | 0.24 x 0.19 | 0.27 x 0.05 | 0.32 x 0.05 |
| **Known Targets** | Shaker | Vessel Galway Girl | Shaker | Shaker Fence | Possible Iceberg | Melting Ice |
| **Vibration Frequency (Hz)** | 2 | 36 | 20 | 2 | - | - |
| **Vibration Amplitude (mm)** | 30 | - | 2.21 | 2.15 | - | - |
| **Ground Truth Source** | Accelerometer | GPS<br>Accelerometer | GPS<br>Accelerometer | Accelerometer | Visual Inspection | Visual Inspection |

### 5.6.2 Performance on real data

After introducing the real SAR acquisitions used for validation, the proposed detection and classification pipeline is applied to the real SAR images. The objective is to evaluate how well the algorithm identifies vibrating targets, such as vessels and vibrating corner reflectors, from static targets, such as icebergs and sea clutter, under real conditions. Each scenario is presented individually, with detection and classification results.

#### Scenario R1: Shaker Test in Trento, Italy

This scenario is given by a controlled TerraSAR-X acquisition where the target was a corner reflector mounted on a mechanical shaker. The setup of the experiment can be seen in Fig. 5.16, where a) and b) depict the corner reflector on the shaker with annotations on the direction of the movement, and c) represents the acquisition geometry of the experiment [210]. The satellite is tasked to pass over the designated area at a specific date and time, which allows to perform different experiments, such as vibrating structures, walking and running people, and even rotating blades. In Fig. 5.17 the area in Trento imaged by the SAR is shown, with the exact location of the shaker shown up close for better visualisation. Already, the vibrating target can be seen, along with its



(a) (b) (c)

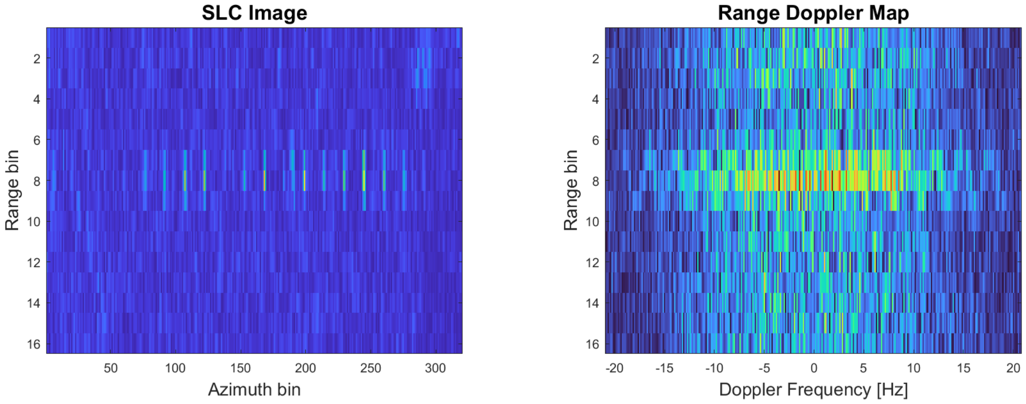
Figure 5.16: Images showing the experimental setup of the shaker (a and b), and the SAR acquisition geometry (c) in relation to the target [210].

ghost targets.

Figure 5.17: Area imaged by the TerraSAR-X in Trento, Italy. The location of the shaker is highlighted by a green point in the image and the zoomed in area.

In this scenario, the shaker was vibrating at 2 Hz with an amplitude of 30 mm. The processing chain includes cropping the original SAR image to a 16 range bins x 320 azimuth bins, recalculating the sampling frequency, CFAR-based detection in the RD domain, aggregation of contiguous detections, STFT-based refinement, autocorrelation feature extraction, and final classification using the trained multi-class SVM.

**SLC image and RD Representation** Fig. 5.18 presents the SLC image and the corresponding RD representation for scenario R1. The RD image (Fig. 5.18b) exhibits a localised region of higher reflectivity around range bins 7-8, indicating the presence of a potential target. The SLC image of the same scene, shown in Fig. 5.18a, reveals several bright scatterers in the vicinity of range bins 7-8. In addition to the main dominant response, secondary bright returns, known as ghost targets, are also visible, which indicates the likely presence of a target that exhibits micro-Doppler.



(a) SLC image

(b) RD image

Figure 5.18: SLC image and RD images corresponding to scenario R1.

**CFAR Detection and Range-Bin Aggregation** A one-dimensional CFAR detector is applied along the range dimension using a sliding window with guard 2, to avoid energy leakage contamination, and training cells equal to 3, to estimate background noise. The $P_{fa}$ is set to 0.1 for candidate generation. The CFAR threshold T is therefore influenced by these variables and it is different for every evaluated range bin, according to the estimated noise statistics.

Figure 5.19 illustrates the range domain energy profile across range bins together with the adaptive CFAR used for candidate selection. Range bins 7 and 8 are identified as potential target locations and, since they are contiguous, they are merged into a single cluster to for one candidate detection region for subsequent refinement. This aggregation step avoids bias toward wider detections while preserving the vibration signature, so detection does not depend on target spatial size.

The individual cross-range profiles associated with range bins 7 and 8 are shown in Fig. 5.20, along with the resulting aggregated candidate profile after CFAR clustering, to be used for the refinement step.

According to the ground truth, the true vibrating target is located at range bin 8, which agrees with the representative detected bins.

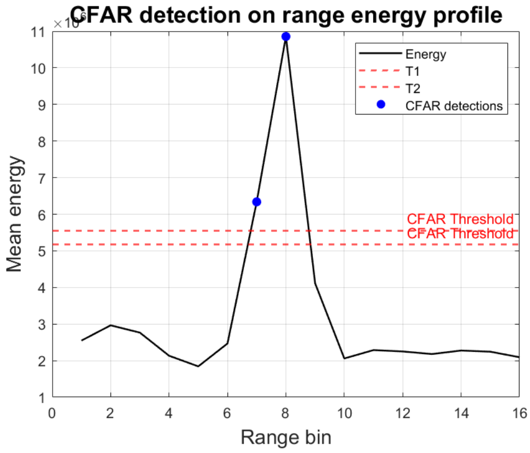
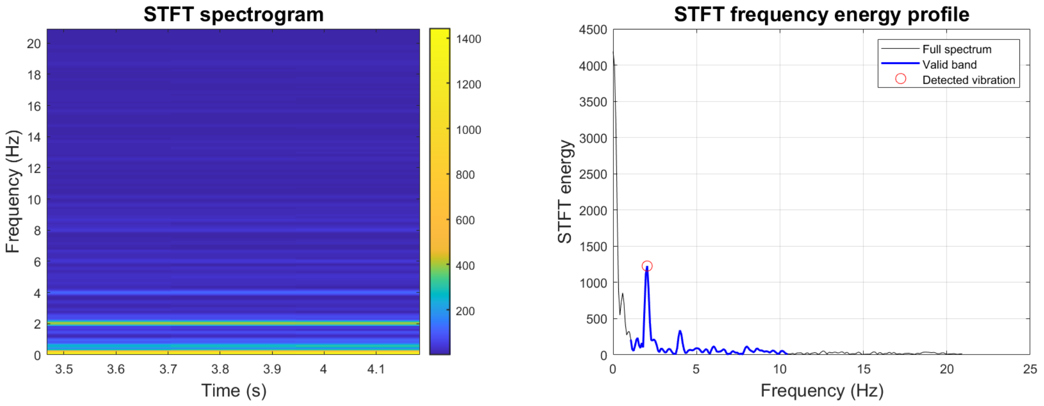


Figure 5.19: CFAR detection of range bins 7 and 8 with their respective threshold, where the lower one is the threshold corresponding to range bin 7.

**Time-Frequency Analysis and STFT Refinement** To further validate the presence of vibration, an STFT is computed on the aggregated signal. The resulting spectrogram, shown in Fig. 5.21, reveals a clear frequency component within the expected physical vibration band. In this case a vibration band has been set (1-10 Hz) but this step can be ignored when the ground truth is not available.



(a) STFT Spectrogram.

(b) STFT Frequency energy profile.

Figure 5.21: Time-frequency analysis through STFT.

This stage is designed to suppress weaker detections or anomalies and retains only the cluster with the strongest and most consistent spectral content. In this case, only one cluster is detected, and its representative range bin is therefore passed on to the

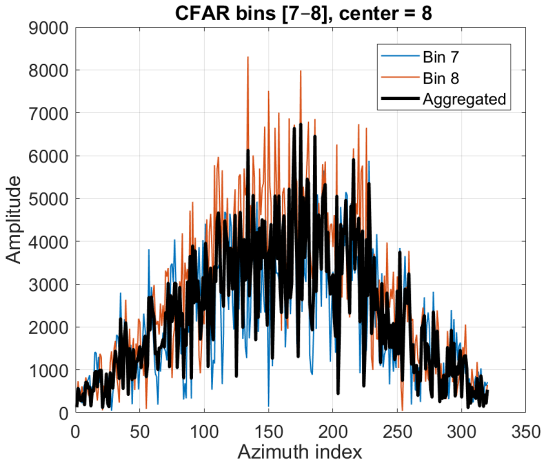
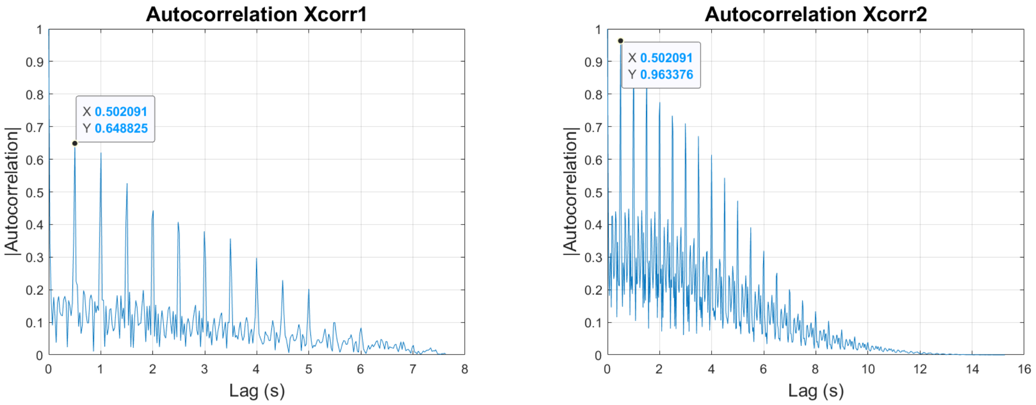


Figure 5.20: Individual cross-range profiles for range bin 7 and 8, and CFAR aggregation of candidate cross-range profiles.

feature extraction and final classification steps.

**Autocorrelation Analysis and Feature Extraction** Fig. 5.22 shows the normalised autocorrelation function computed from the final selected range bin where (a) is the single autocorrelation and (b) is the double. The periodic structure is clearly visible, indicating the presence of a vibrating target. Furthermore, both signals peak at around lag = 0.5 s, which corresponds to the true vibrating frequency of the target (2 Hz). From this stage, the frequency is already estimated very precisely. From the



(a) Xcorr1.

(b) Xcorr2.

Figure 5.22: Autocorrelation results from single autocorrelation (a), and double autocorrelation (b).

autocorrelation response, a set of structural features is extracted and used as input to the trained multi-class SVM classifier.

**Classification** Finally, both SVM models (the one trained on Xcorr1 and the one trained on Xcorr2) label the detected target correctly. Fig. 5.23 presents the final detection map, where the sample can be seen as classified as vibrating, represented by blue colour. A static target or clutter would be seen as a magnitude orange or red, respectively. Both single and double autocorrelations result in the same estimated frequency, 1.9917 Hz, which is close to a perfect estimation of the vibration.

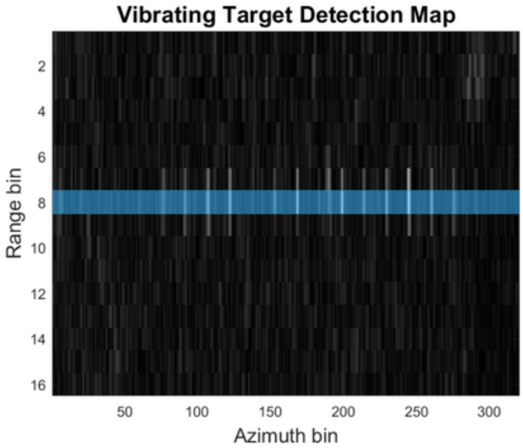


Figure 5.23: Final detection map where 3 (red) corresponds to the detection of clutter, 2 (orange) corresponds to the detection of a static target, and 1 (blue) represents the detection of a vibrating object.

The algorithm successfully detects the correct range location, aggregates neighbouring detections appropriately, identifies vibration through time-frequency and autocorrelation analysis, and assigns the correct class label. No false vibrating detections are observed in this scenario. These results reassure the validity of the method presented in the chapter to be used in the real world.

### Scenario R2: Vessel “Galway Girl”, off the Coast of Ireland

This scenario is given by a controlled Umbra-08 acquisition carried out by the team at the SSP&S lab, where the cooperative maritime target was a vessel off the coast of Ireland. The *Galway Girl* boat is a passenger ship with a length overall of 24 m, a beam of 6.2 m , and a draught of 1.8 m, and it can be seen in Fig. 5.24, image taken by the team the day of the acquisition.

Figure 5.24: Photo of the Galway Girl boat.

Ground truth measurements were collected using onboard Global Positioning System (GPS) and accelerometer instrumentation. Analysis of these measurements indicated that, at the time of acquisition, the vessel exhibited a dominant vibration frequency of approximately 36 Hz. The imaged scene is shown in Fig. 5.25, where the vessel’s location at the moment of the satellite pass is highlighted and enlarged for clarity.

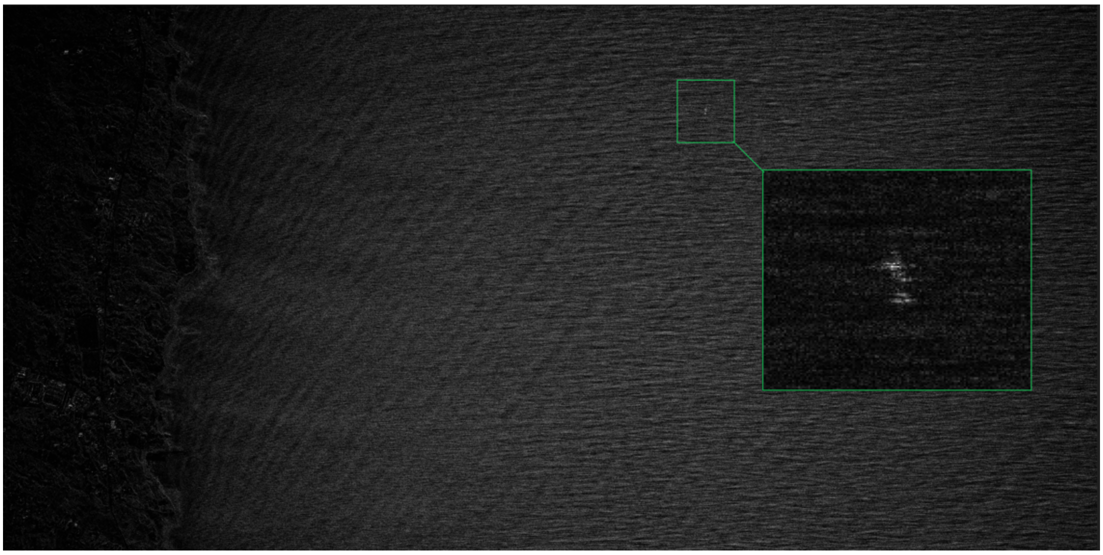


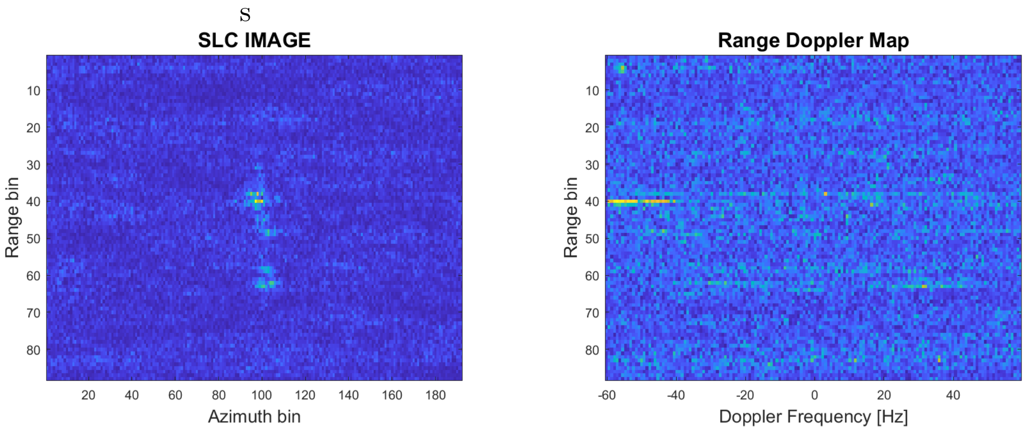
Figure 5.25: Area imaged by the Umbra-08 in Ireland. The location of the vessel is highlighted by a green square in the image and the zoomed in area.

For processing, the original SAR image is cropped around the area of interest, the new sampling frequency is calculated, and CFAR-based detection in the RD domain, aggregation of contiguous detections, STFT-based refinement, autocorrelation feature extraction, and final classification are performed.

**SLC image and RD Representation** First, the original SAR image is cropped around the area of interest, as seen in Fig. 5.26a. SLC and RD representations of the scene are shown in Fig. 5.26, which after cropping covers 88 range bins and 192 azimuth bins. The cropping in azimuth needs to include sufficient bins to be able to extract valuable vibrational information. In Fig. 5.26b, it can already be seen that the possible vibrating target is around range bin 40.

**CFAR Detection and Range-Bin Aggregation** The CFAR detector is applied along the range dimension using a sliding window with guard 2, and training cells 3. The $P_{fa}$ is set to 0.1 for candidate generation.

Fig. 5.27 shows the energy profile across range bins and the adaptive CFAR thresh-



(a) SLC image

(b) RD image

Figure 5.26: SLC image and RD images corresponding to scenario R2.

old corresponding to the candidate range bin (40). In this case, the detector is only flagging one range bin, so the aggregation of contiguous range bins step is not performed.

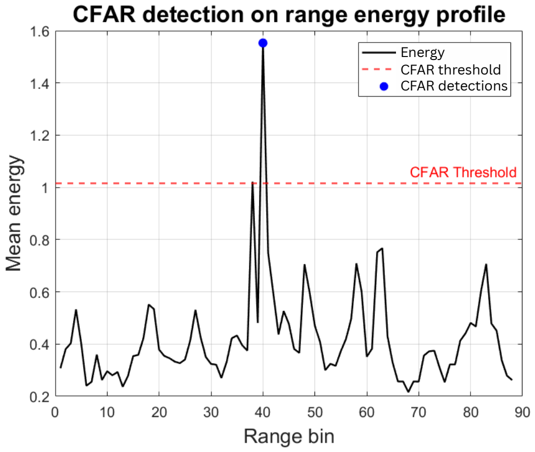
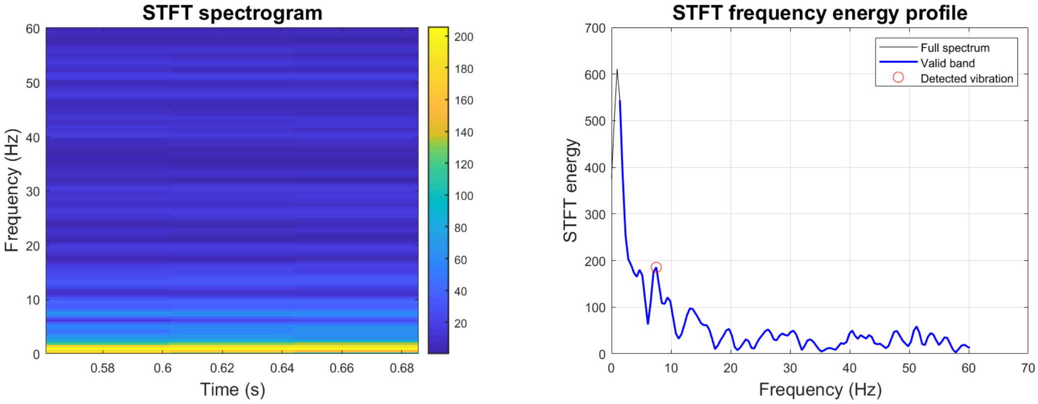


Figure 5.27: CFAR detection of range bin 40 with its threshold.

**Time-Frequency Analysis and STFT Refinement** For the refinement of the detected range bins, STFT is computed on the signal. The resulting spectrogram, shown in Fig. 5.28a shows a peak at around 7 Hz, which does not match the ground truth in this case. In Fig. 5.28b, this peak can be seen more clearly. No peaks around the true vibrating frequency seem to arise from the STFT detection step, however the



(a) STFT Spectrogram.

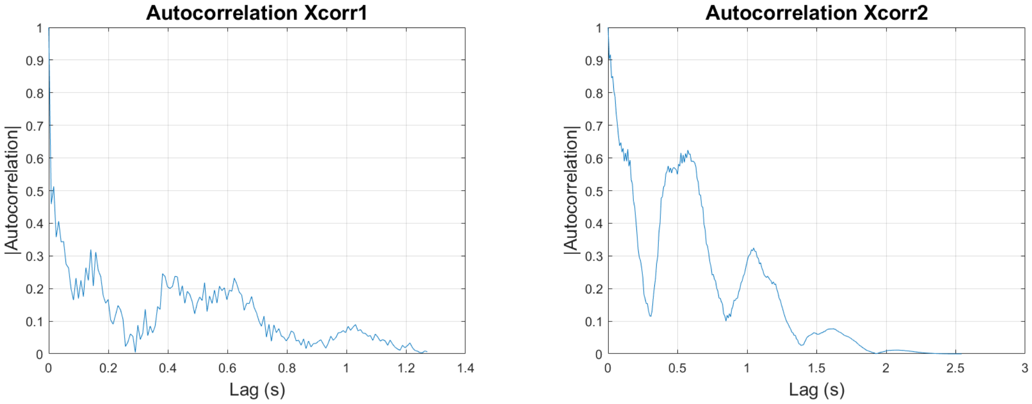
(b) STFT Frequency energy profile.

Figure 5.28: Time-frequency analysis through STFT.

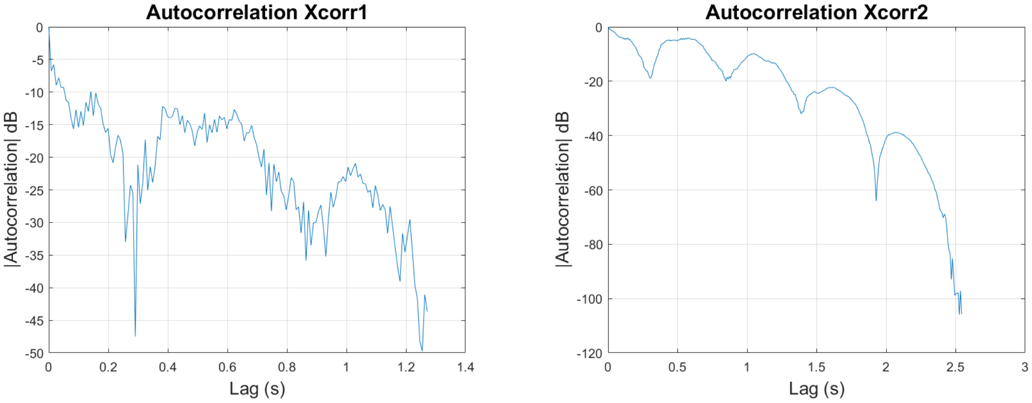
candidate range bin is allowed to be passed on to the next step of the pipeline.

**Autocorrelation Analysis and Feature Extraction** Fig. 5.29 shows the normalised autocorrelation function computed from the final selected range bin where (a) is the single autocorrelation and (b) is the double. Furthermore, both signals are also presented in dB scale, in Fig. 5.29 (c) and (d). The periodic structure is somewhat visible in Xcorr1, but a clearer vibrating pattern arises from double autocorrelation, indicating the presence of a vibrating target. Looking to the dB scale representation, Xcorr2 clearly exhibit a periodic pattern. Nevertheless, although the pattern is picked up by the autocorrelation processing, the estimation of the frequency at this step is not correct, being 25.94 Hz and 16.51 Hz in the single and double autocorrelation cases, respectively. This is to be expected in real scenarios, where the SLC images are influenced by many factors that degrade the signal and false peaks might be picked up by the system. From the autocorrelation response, a set of structural features is extracted and used as input to the trained multi-class SVM classifier.

**Classification** Finally, both SVM models (Xcorr1 and Xcorr2), label the detected target correctly. Fig. 5.30 presents the final detection result, where the sample can be seen as classified as vibrating, represented by the colour blue in the detection map. However, as mentioned above, the estimated frequency in this case is not accurate.



(a) Xcorr1. (b) Xcorr2.



(c) Xcorr1. (d) Xcorr2.

Figure 5.29: Autocorrelation results from single autocorrelation (a) in linear scale and (c) in db scale, and double autocorrelation (b) and (d), in linear and dB scales respectively.

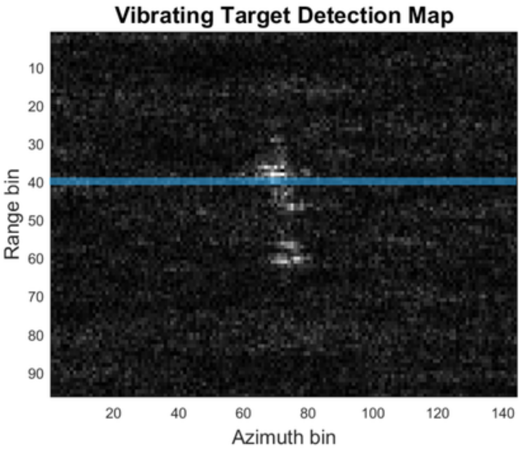


Figure 5.30: Final detection map for the Galway Girl target where the blue colour for detection corresponds to a vibrating target.

The algorithm successfully detects the correct range location, identifies vibration through time-frequency and autocorrelation analysis, and assigns the correct class label. No false vibrating detections are observed in this scenario, but the estimated frequency does not match the ground truth data.

### Scenario R3: Shaker in Glasgow Green Park, UK

This scenario is given by a controlled Umbra-04 acquisition carried out by the team at the SSP&S lab, where the controlled target was a corner reflector mounted on a mechanical shaker in Glasgow Green park. The ground truth was collected with an accelerometer, and its frequency was 20 Hz with an amplitude of 2.21 mm, at the time of the satellite pass. In Fig. 5.31, the images area can be seen, with a special focus on the area where the testing was being done. The shaker is located in the centre of the zoomed-in area, although its visibility is somewhat poor.

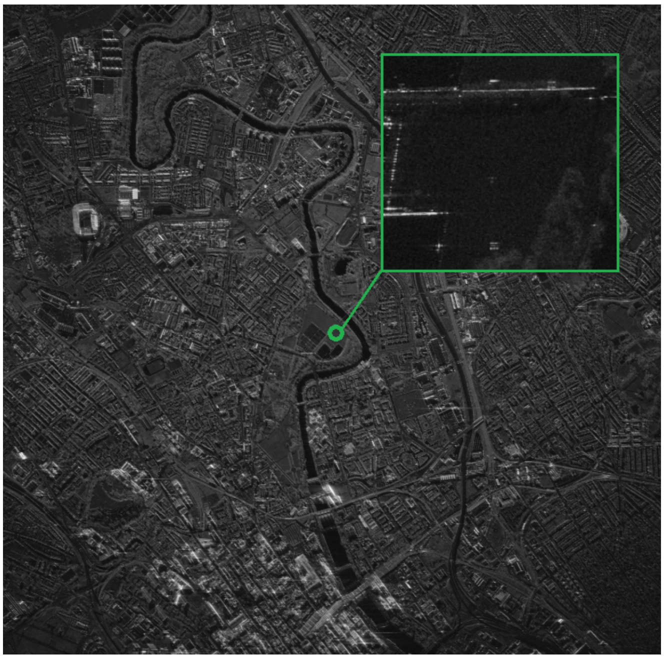
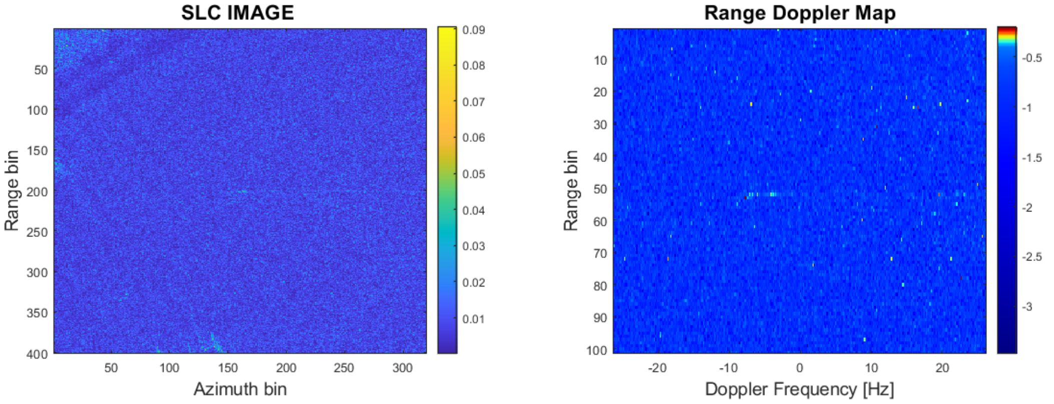


Figure 5.31: Area imaged by the Umbra-04 sensor in Glasgow. The location of the shaker is highlighted by a green square in the image and the zoomed in area, located in Glasgow Green.

For processing, the original SAR image is cropped around the area of interest (as seen in Fig. 5.32), the new sampling frequency is calculated, and CFAR-based detection in the RD domain, aggregation of contiguous detections, STFT-based refinement, autocorrelation feature extraction, and final classification are performed.

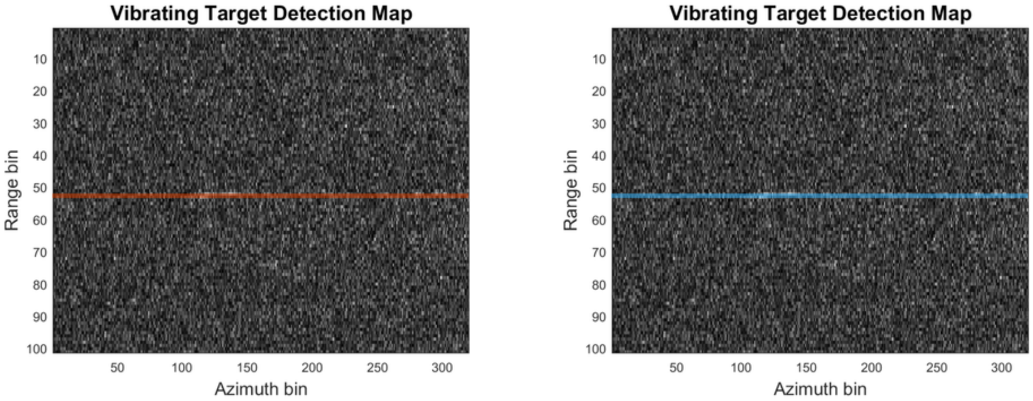
The CFAR detection step produces range bins 52 and 53 as detected bin candidates and, since these detections are contiguous, they are grouped into a single cluster and aggregated by averaging. STFT is computed on the signal for refinement and, although no peaks around the true vibrating frequency arise, vibrational content is detected.



(a) SLC image (b) RD image

Figure 5.32: SLC image and RD images corresponding to scenario R2.

Fig. 5.33 presents the final detection map for both Xcorr1 (a), and Xcorr2, (b), where the sample is only classified as vibrating in the case of the double autocorrelation. The algorithm successfully detects the correct range location, identifies vibration through double autocorrelation analysis, assigns the correct class label, and estimates



(a) Xcorr1. (b) Xcorr2.

Figure 5.33: Final detection map of the classification, for Xcorr1 and Xcorr2, where blue means vibration and red means clutter.

the correct vibration frequency (20 Hz).

The results are coherent with the behaviour exhibited by the synthetic data previously. When the amplitude of the vibration is very low, the detection becomes more challenging and the autocorrelation representations struggle to show a periodic pattern. However, in this case, the vibration and the true frequency were correctly identified, which shows the potential of the technique in the real world.

### Scenario R4: Shaker and Fence in Glasgow Green, UK

This scenario is given by a controlled Umbra-05 acquisition carried out by the team at the SSP&S lab, where the controlled target was a corner reflector mounted on a shaker in Glasgow Green park, and a fence for a static reference.

Both targets under study can be seen in Fig. 5.34. The ground truth for the shaker was collected with an accelerometer, and its frequency was 2 Hz with an amplitude of 2.15 mm. In Fig. 5.35, the imaged area can be seen, with a special focus on the area where the targets under study are located.


(a) Fence.

(b) Shaker setup.

Figure 5.34: Fence (a) and mechanical shaker (b) in Glasgow Green park.

First, the large SAR image is cropped around the two areas of interest, as seen in Fig. 5.35, and each area is processed independently with the aim of testing the pipeline with a static and a vibrating target. In Fig. 5.36, both SLC and RD representations of the targets are shown, where (a) and (b) correspond to the corner reflector attached to the shaker, and (c) and (d) correspond to the static fence.

The detection step outputs range bin 10 and 11 as candidates for the shaker, and

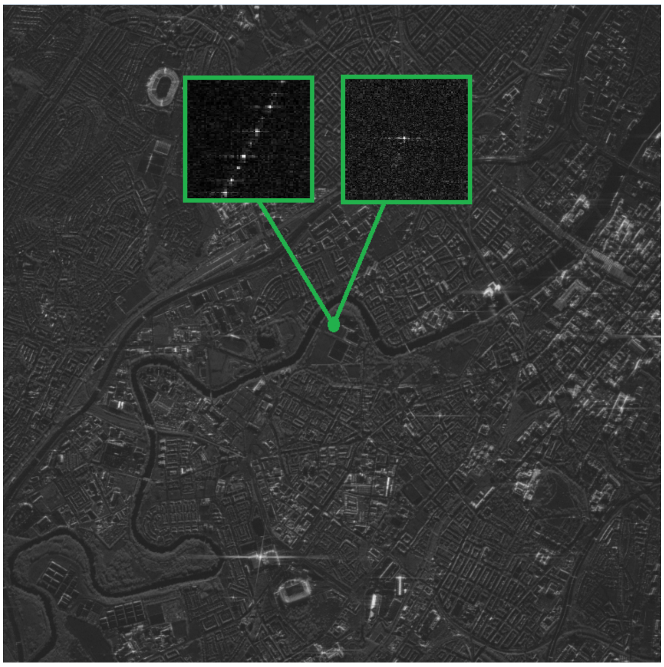
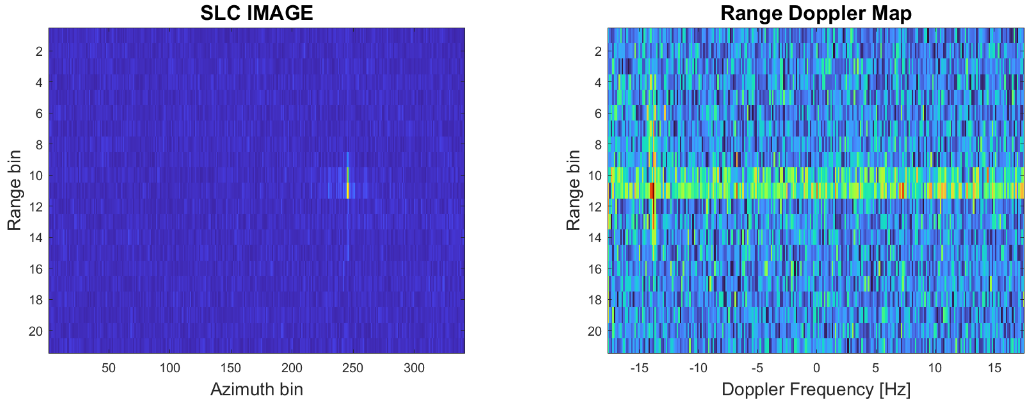


Figure 5.35: Area imaged by the Umbra-05 sensor in Glasgow. The location of the shaker is highlighted by a green square in the image and the zoomed in area, located in Glasgow Green.

13 range bins in total for the fence, of which 8 clusters are formed through aggregation of the contiguous bins.

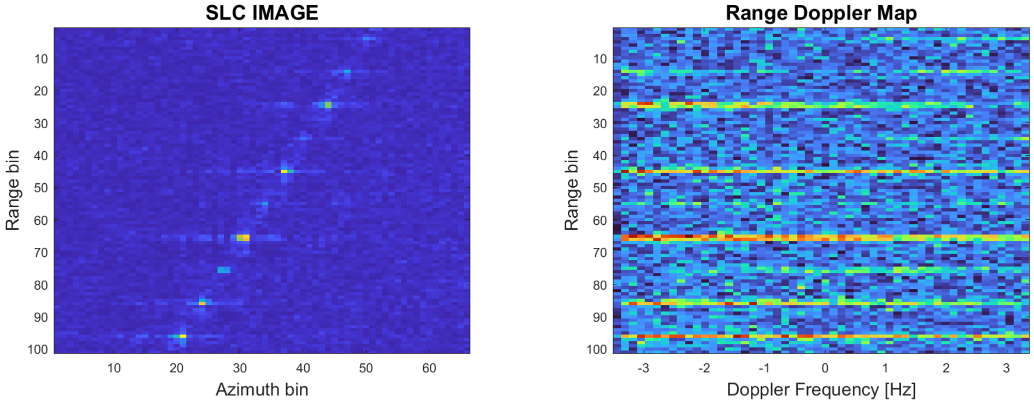
In the case of the shaker, the Xcorr1-based classifier identifies a vibrating target at range bin 10, consistent with a periodic structure in the single-autocorrelation response. In contrast, when Xcorr2 features are used, the same range bin is classified as a static target.

In the fence experiment, four range bins are ultimately detected. Using Xcorr1 features, range bins 25 and 45 are classified as static targets, while range bins 55 and 76 are labelled as clutter. However, when Xcorr2 features are employed, all detected range bins are classified as static, reflecting the loss of structure and the dominance of smooth, low-variation patterns in the Xcorr2 domain. The final detection map can be seen in Fig. 5.37. The algorithm successfully detects the correct range locations for the shaker and the fence scatterers, identifies vibration through autocorrelation analysis, and assigns the correct class label only when Xcorr1 is applied for the shaker, and when



(a) SLC image shaker.

(b) RD image shaker.

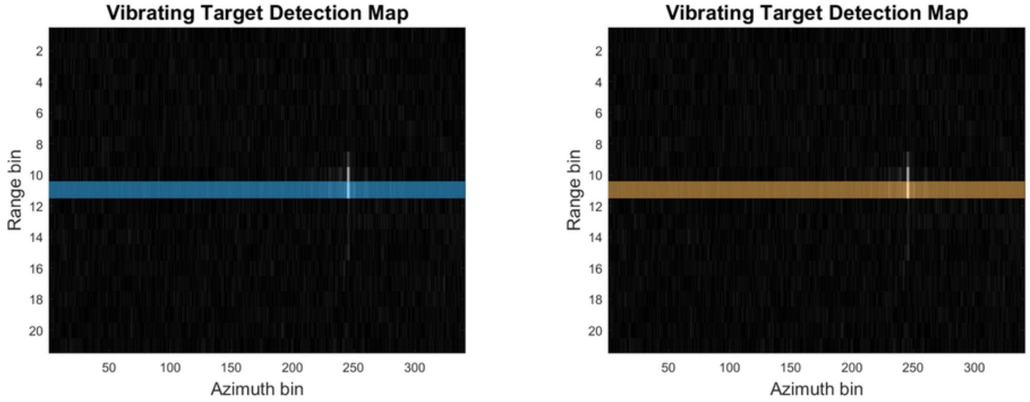


(c) SLC image fence.

(d) RD image fence.

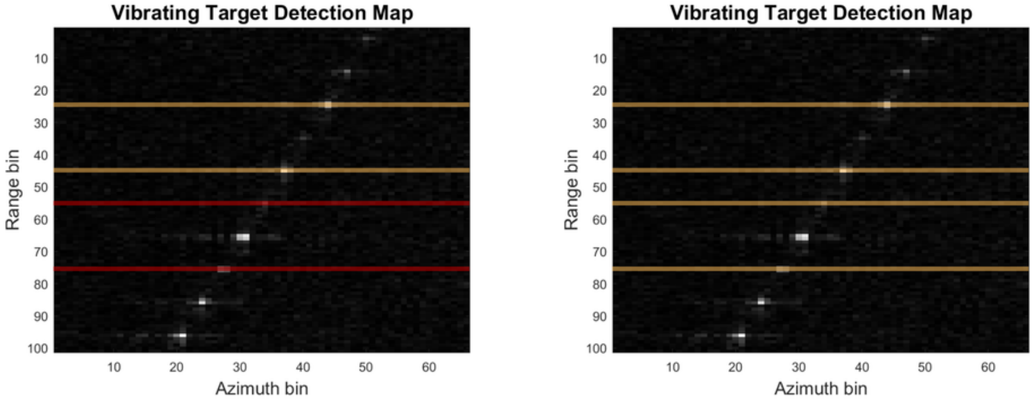
Figure 5.36: SLC image and RD images corresponding to scenario the shaker (a) and (b), and the fence (c) and (d).

Xcorr2 is applied for the fence. It also estimates an approximation of the vibration frequency of the shaker (1.69 Hz).



(a) Xcorr1.

(b) Xcorr2.



(c) Xcorr1.

(d) Xcorr2.

Figure 5.37: Final detection map of the classification, for Xcorr1 and Xcorr2, where blue means vibration, orange menas static target, and red means clutter.

### Scenario R5: Nuuk, Greenland

This Spotlight Ultra Capella Space acquisition is analysed in order to find an iceberg. At a glance, no apparent icebergs or melting ice pieces can be seen in the image, but looking closely, a possible target is identified in it. In Fig. 5.38, the imaged area is shown highlighting the potential target of interest in green, an unknown object close to the coast. To properly identify what it is, the proposed methodology needs to be applied.

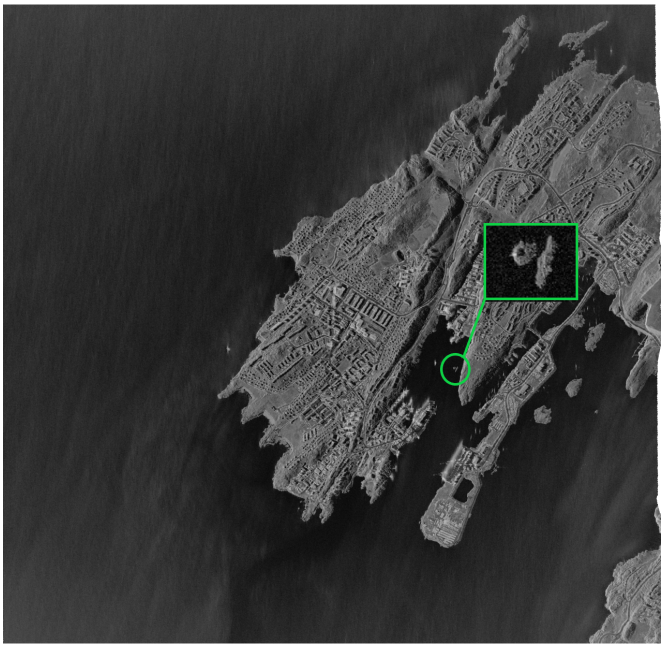
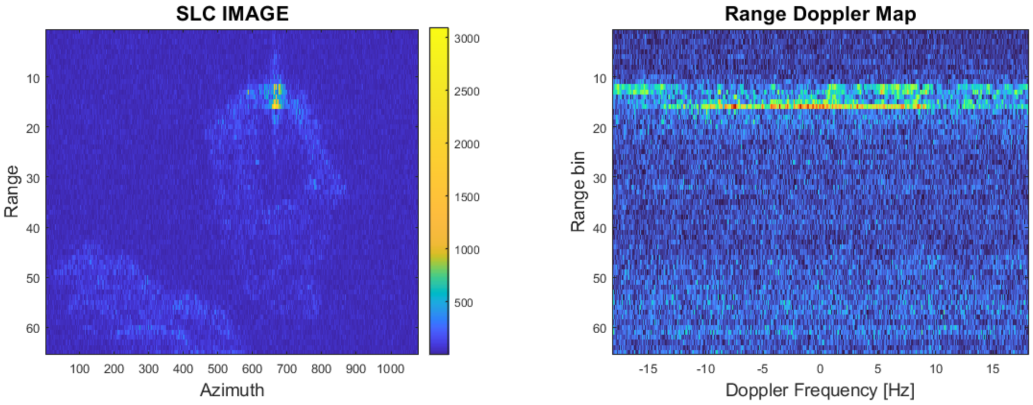


Figure 5.38: Area imaged by the Capella-C15 sensor in Greenland, off the coast of Nuuk. The location of the two possible icebergs highlighted in green.

The sub-image is cropped from the original SAR scene, and the new sampling frequency is recalculated to preserve temporal and azimuthal fidelity. The cropped image undergoes the full detection and classification chain: CFAR in the RD domain, aggregation of contiguous detections, STFT refinement, autocorrelation feature extraction, and finally, SVM-based classification.

After cropping around the region of interest, azimuth defocusing is performed to generate the RD representation. Fig. 5.39 shows both SLC and RD images, where the suspected target can be seen, exhibiting a small area with noticeable brighter scatterers.



(a) SLC image (b) RD image

Figure 5.39: SLC and RD images corresponding to the target 1 - possible iceberg.

Through the detection step, three consecutive range bins are detected, 14, 15, and 16. Fig. 5.40 presents the final detection map for both Xcorr1, and Xcorr2, where the sample is classified as vibrating in both cases. The result was an estimated frequency of 0.50 Hz in the case of the Xcorr1, and 0.37 Hz in the case of the Xcorr2. However, the target was selected as a possible iceberg, which does not match the results obtained.

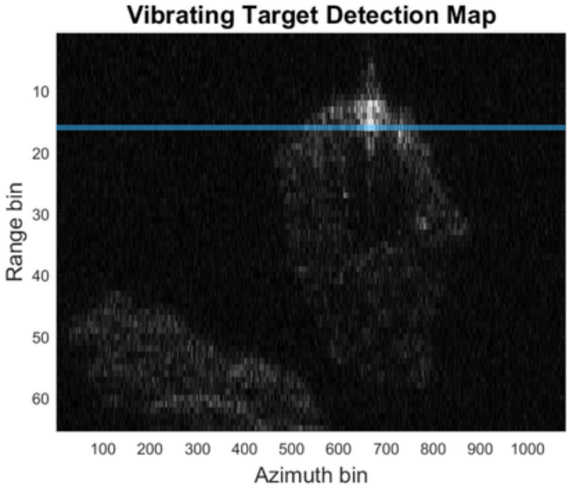


Figure 5.40: Final detection map of the classification, for Xcorr1 and Xcorr2, where blue means vibration.

After the final classification indicates that the target is vibrating, the area is examined more closely through visual inspection. This reveals that the object is not an iceberg but a nautical marker located near the harbour of Nuuk. As shown in Fig. 5.41, a metallic structure is mounted on top of a rock formation in the water, indicating that it is an aid to navigation for warning vessels of hazardous rocks and assist in safe sailing.

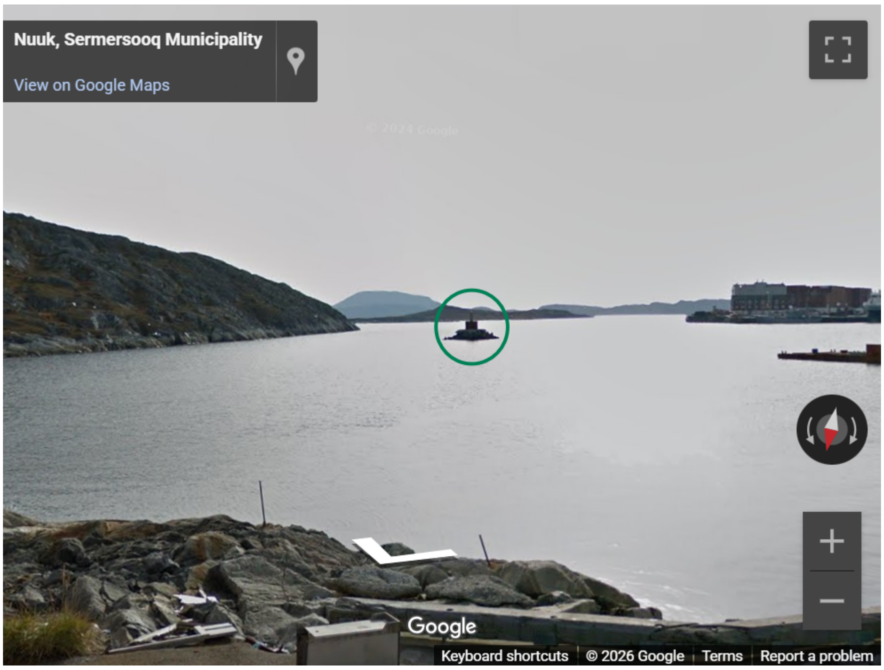


Figure 5.41: Nautical marker location highlighted in green. Source: Google Street View, Nuuk, Sermersooq Municipality. © Google.

The detected vibration can be explained by the physical characteristics of the object identified as a nautical marker. Such markers are commonly constructed from metal and mounted on rock formations above the water, making them susceptible to environmental forces such as wind, waves, and water currents. These external forces can induce mechanical oscillations in the structure, which can be captured by the sensing system. Given the limited visual resolution and perspective of the image, the exact type of marker cannot be determined, however, it is likely to be a fixed navigation beacon or a daymark [211]. The object was likely under certain weather conditions that made it oscillate, although not in a perfectly periodic pattern, but enough for the classifier to determine that it exhibited some vibration.

### Scenario R6: Melting ice, Greenland

This scenario investigates an area in Greenland where partially melting ice formations are visually identified. A Capella-03 Spotlight Mode acquisition is examined to identify possible residual iceberg structures. The area is shown in Fig. 5.42, where fragmented returns are visible, likely corresponding to melting ice formations.

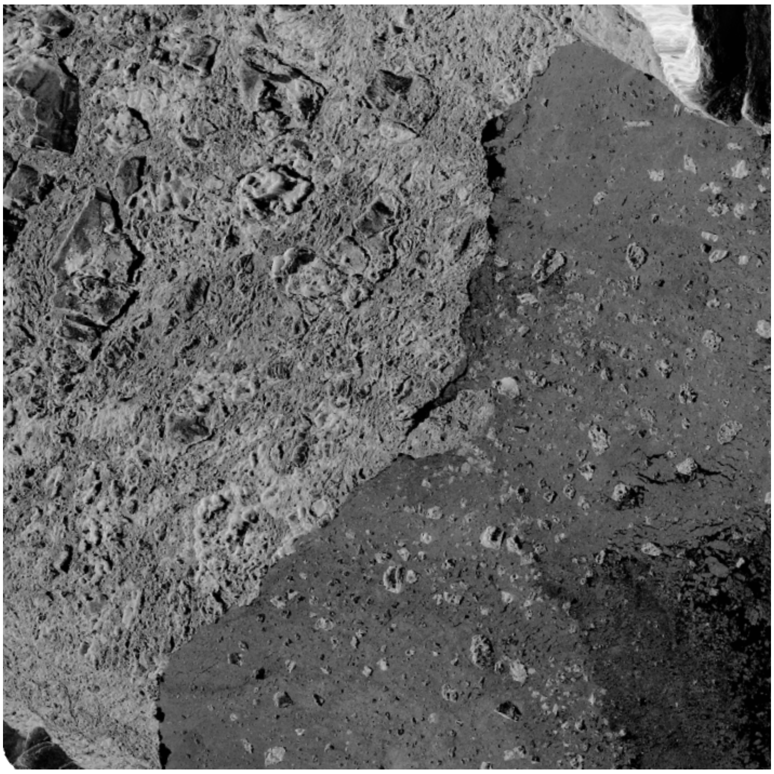
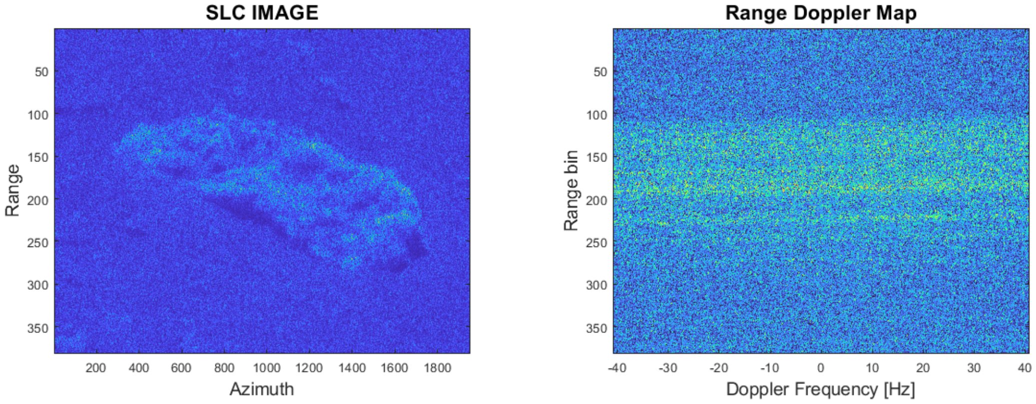


Figure 5.42: Area imaged by the Capella-C03 sensor in Greenland.

The selected region of interest is cropped from the large SAR scene, and azimuth compression inversion is applied to obtain the corresponding RD representation. The outputs are presented in Fig. 5.43. The dielectric contrast of melting surfaces results in weak, spatially diffuse scatterers, as can be appreciated from the image.



(a) SLC image

(b) RD image

Figure 5.43: SLC and RD images corresponding to Target 2 - possible iceberg.

The candidate range bins 188 and 368 are highlighted by the detection step and forwarded to the following steps of the pipeline. The multi-class SVM classifier assigns the clutter label for both Xcorr1 and Xcorr2 analysis, as seen in the final detection map in Fig. 5.44.

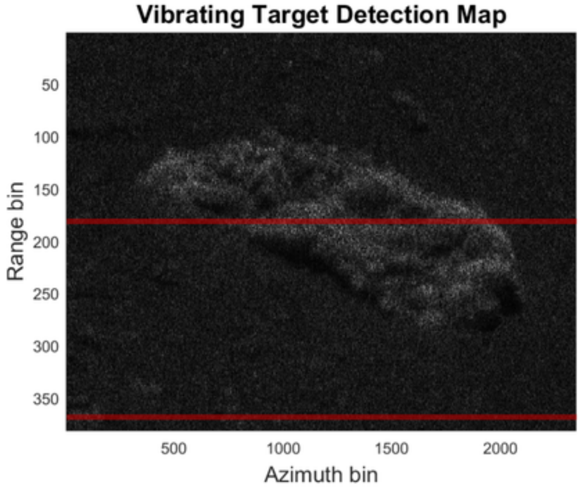


Figure 5.44: Final detection map of the classification, for Xcorr1 and Xcorr2, where red means clutter.

Examining the area under study in this case and the autocorrelation results, the classification outcome aligns with expectations, as the object appears poorly visible to the sensor or obscured by significant background clutter.

Icebergs may be classified as clutter because their surface rugosity and morphological features make them less distinguishable to the radar. The lack of a consistent oscillatory signature and the weak backscatter contrast indicate that the observed features are not active vibration sources but melting ice with spatially varying reflectivity. This misclassification as clutter does not indicate a fundamental limitation of vibration-based discrimination, as the method remains effective in distinguishing vibrating from non-vibrating targets. When surfaces undergo more challenging characteristics, like rugosity changes in this case, the pipeline can still classify the object as non-vibrating.

## 5.7 Discussion and Conclusions

The conducted experiments across controlled and real-world scenarios demonstrate the robustness and adaptability of the proposed SAR-based vibration detection and classification framework. Using a consistent data processing chain comprising CFAR detection, range-bin aggregation, time-frequency analysis, autocorrelation-based feature extraction, and multi-class SVM classification, the methodology effectively identified and discriminated targets exhibiting distinct micro-motion characteristics from static targets and clutter in different scenarios and across different SAR imagery providers.

**Detection and Classification Performance** The controlled experiment in Scenario R1 established the reliability of the full processing pipeline under known conditions. The shaker, vibrating at 2 Hz with a 30 mm amplitude, was detected at the correct range bin and accurately classified as a vibrating target. Both single and double autocorrelation paths estimated the vibration frequency with a good relative error below 0.5%, validating the time-frequency and autocorrelation modules as precise frequency estimators.

Subsequent scenarios involving natural and semi-controlled environments presented increasing degrees of signal complexity. In scenario R2 (Galway Girl boat), micro-motions were detected under increased clutter and reduced SNR, confirming the method's capacity to extract subtle motion signatures from distributed scatterers.

In contrast, scenarios R5 and R6 (Greenland) highlight the challenges associated with classifying targets under Arctic maritime conditions. In R5, the classifier successfully identified a vibrating object near Nuuk, later verified as a fixed nautical beacon rather than an iceberg. The misinterpretation of this non-ice scatterer illustrates the method's sensitivity to structures that exhibit low-frequency oscillations induced by environmental effects (such as waves). Meanwhile, R6 demonstrated that melting ice masses with diffuse scattering are classified as clutter due to their non-periodic dynamics. These results collectively underscore both the strengths and limitations of vibration-based discrimination, particularly when applied to heterogeneous natural targets.

Across all scenarios, CFAR detection proved efficient in identifying potential range bins of interest while maintaining manageable false-alarm rates. The adaptive aggregation of neighbouring detections effectively preserved the micro-motion signature, ensuring that the subsequent STFT step operated on clean, representative signals. This hierarchical approach consistently reduced misclassification errors.

**Analysis of Autocorrelation-Based Features** The introduction of dual autocorrelation levels (single and double) proved to be useful in distinguishing periodic vibrations from random clutter. In controlled and semi-controlled settings, both autocorrelation metrics manifested distinct periodic structures, enabling reliable frequency estimation and target classification. In natural scenarios (e.g., maritime or melting ice), the resulting correlations were noise-like, reflecting either low coherence or non-rigid surface motion.

The performance of the SVM classifiers trained on these features confirms that autocorrelation encapsulates critical temporal stability information beyond the spectral domain. The autocorrelation-based features demonstrated robustness to moderate noise levels in the considered datasets and provided stable descriptors for SVM classification, complementing the information from time-frequency analysis.

**Limitations and Environmental Sensitivity** Despite the strong performance in controlled environments, the experiments reveal several challenges when extending the approach to real scenes. The main limitations include:

- Environmental excitation ambiguity: Vibration signatures in metallic and infrastructural targets (e.g., beacons, bridge decks) may originate from external forces such as wind or waves, complicating the physical interpretation of detected motion in the backscatteres that form the SAR image.
- Decorrelation in melting and icy surfaces: Rapid dielectric changes, sub-pixel motion, and multi-path scattering degrade phase stability and bury the patterns detected by the autocorrelation method.
- Resolution and SNR constraints: Even in high-resolution Capella Spotlight acquisitions, weak coherent returns limit detectability of potential vibrations.
- Generalisation of classifiers: The SVM model, although robust to moderate variation, relies on representative training data. Applying the same classifier to very different environments (maritime vs. urban) leads to reduced accuracy.

These findings highlight the importance of context-aware interpretation, especially when transferring models trained in controlled settings to natural environments.

### 5.7.1 Conclusion

The experimental validation presented herein confirms that the proposed vibration detection framework effectively identifies, characterises, and classifies dynamic scatterers in high-resolution SAR imagery. From the precisely controlled shaker test to complex Arctic scenes, the approach consistently captured underlying motion signatures, adapting across different sensors and environmental conditions.

While the system's performance decreases when applied to incoherent natural targets such as melting ice or rough iceberg surfaces, it still offers valuable diagnostic information about surface dynamics and environmental consequences. The integration of CFAR-based detection, STFT refinement, and autocorrelation feature learning represents a solid foundation for future autonomous target monitoring systems.

The exploitation of micro-motion characterisation for maritime target classification remains largely unexplored due to several challenges, including the effects of the platform motion and the complexity of isolating subtle phase variations in distributed targets, such ships. In this context, the approach proposed in this work aims to complement existing methods by exploiting frequency-domain signatures that capture dynamic target behaviour not accessible through conventional image-based techniques. The results not only advance vibration detection in SAR data but also contribute to broader research efforts in remote sensing of structural health, maritime surveillance, and monitoring, providing a scalable methodology for recognising micro-motion signatures in complex SAR observation scenarios.

# Chapter 6

# Conclusion and Future Work

This thesis contributes to the advancement of the field of SAR image processing and its applications in maritime surveillance, with a particular focus on multitemporal image coregistration, and detection, classification, and characterisation of dark vessels. The research addresses critical challenges in multitemporal SAR processing and target recognition, providing innovative solutions to enhance maritime surveillance applications.

A contribution of this thesis is multitemporal SAR coregistration. A novel joint registration strategy, JCLS, based on cross-cross-correlations between image patches was developed to improve alignment accuracy in low-SNR and low-coherence conditions. Unlike conventional approaches that independently register each secondary image to a master, the proposed framework exploits reciprocal misregistration information between all image pairs within the stack, enhancing the CLS optimisation procedure. Experimental validation on multitemporal airborne SAR data demonstrated improved robustness, particularly in scenarios involving relative rotations. The coregistration framework addresses a complementary challenge relevant to maritime surveillance, where accurate multitemporal alignment is essential in applications such as vessel change detection, interferometric analysis, and long-term monitoring of maritime infrastructures. In dynamic ocean environments, where decorrelation and low SNR are common, improved registration stability can enhances the reliability of operational maritime monitoring.

Furthermore, the ATR methodologies developed contribute to enhancing the real-time detection and tracking of dark vessels. By introducing spectral profiles and Krawtchouk moments, this work provides a signature representation of SAR data, advancing the accuracy and reliability of target recognition.

Moreover, this thesis introduces and validates a vibration-based detection and classification framework for high-resolution SAR imagery, built on CFAR detection, slow-time signal aggregation, time–frequency analysis, and autocorrelation-based feature extraction combined with multi-class SVMs. Through controlled experiments and real-world scenarios (such as infrastructure monitoring, Arctic acquisitions, and melting ice areas in Greenland), the approach has been shown to detect and characterise micro-motions, differentiate vibrating from non-vibrating scatterers, and highlight the limitations of vibration-based discrimination in low-coherence natural targets such as melting ice or rough iceberg surfaces. This demonstrates that SAR can be exploited not only for shape and intensity-based target recognition, but also for dynamic behaviour analysis.

Despite these advancements, several challenges remain. The limited availability of large-scale labelled SAR datasets remains a constrain for DL-based approaches, and while techniques such as transfer learning offer some solutions, further refinement is necessary. Computational complexity also presents a problem, particularly for methodologies involving multitemporal coregistration analysis, as they require significant processing resources and multiple acquisitions of the same area. Moreover, the robustness of the proposed techniques to extreme environmental conditions, such as high sea states or adverse weather, still requires further investigation. For the vibration-detection framework, the experiments in Greenland highlight the difficulty of distinguishing between man-made targets (e.g., beacons) and natural objects.

**Future Work** The results of this thesis show that robust multitemporal processing, spectral feature extraction, and micro-motion analysis from SAR imagery are feasible in maritime surveillance scenarios. However, additional developments can be explored to further improve robustness, adaptability to complex environments, and scalability for operational use. A few interesting avenues to pursue as future research are:

1. **Multitemporal Coregistration and Geometric Refinement.** Future work

on the cross-cross-correlation coregistration framework should be done under realistic maritime conditions, and adapting the optimisation to low-coherence regions and keypoints sparsity. Addressing computational cost will also be necessary to enable near-real-time maritime applications.

**2. Polarimetric and Multi-Temporal Feature Integration.** The integration of polarimetric information with repeat-pass coherence or interferometric phase stability could enhance the separation between natural and man-made targets. In particular, this approach may help differentiate surface rugosity of natural objects from real mechanical motion in Arctic and coastal environments, improving iceberg–vessel discrimination and infrastructure analysis.

**3. Multi-Sensor Feature Fusion.** Adaptive feature fusion can be explored to improve discrimination between vessels and natural maritime targets. The combination of SAR features with optical, AIS, thermal, or environmental data could help resolve ambiguities and provide a more physically consistent interpretation of observed signatures, particularly in challenging scenarios.

**4. Real-Data for Training and Generalisation.** While the presented methodologies demonstrate strong performance under controlled and limited real-data conditions, further validation on larger annotated maritime SAR datasets is required. Training on more diverse real-world data would improve robustness to environmental variability and enhance transferability across sensors and imaging configurations.

**5. Advanced Learning Architectures and Transferability.** Extending the current machine learning frameworks toward convolutional or hybrid architectures trained on multi-sensor datasets could enhance robustness and generalisation. Techniques such as self-supervised learning or domain adaptation could help address limited training data.

**6. Extension of Micro-Motion Analysis.** Future work on the vibration-based framework may focus on combining autocorrelation features with coherence or interferometric indicators to better distinguish mechanical motion from environmental variability. Validation across a wider range of maritime structures and operational environments would help to further assess robustness and practical applicability.

**7. Operational Deployment and Onboard Optimisation.** Optimising algorithmic efficiency for resource-constrained platforms, including CubeSats and small satellite constellations, represents an important step toward scalable maritime surveillance. Efficient implementations of the coregistration, feature extraction, and classification pipelines would enable near-real-time onboard processing, supporting faster response in security-critical scenarios.

**8. Introduction of a Synthetic Iceberg-like model.** Experimental results showed that melting ice was classified as clutter, highlighting the need for more accurate signal simulation methods that better capture the characteristics of these types of objects and improve their separability from clutter.

Overall, future work should aim to consolidate advances in coregistration for maritime environments, feature fusion, adaptive learning, and real-data validation. Reinforcing both low-level SAR processing and higher-level dynamic analysis will further improve the robustness of SAR-based maritime monitoring systems.

# Appendix A

# Detailed Classification Results for Controlled Tests

Table A.1: Classification results for the low phase-change regime. Accuracy, precision, recall, and F1-score are reported for the vibrating (vib), static, and clutter classes.

| **Metric** | **Xcorr1 - Xcorr1** | | | **Xcorr2 - Xcorr2** | | |
|---|---|---|---|---|---|---|
| | Vib. | Static | Clutter | Vib. | Static | Clutter |
| Precision | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Recall | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| F1-score | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Accuracy | 100.00 | | | 100.00 | | |

Table A.2: Classification results for the middle phase-change regime. Accuracy, precision, recall, and F1-score are reported for the vibrating (vib), static, and clutter classes.

| **Metric** | **Xcorr1 → Xcorr1** | | | **Xcorr2 → Xcorr2** | | |
|---|---|---|---|---|---|---|
| | Vib. | Static | Clutter | Vib. | Static | Clutter |
| Precision | 1.0000 | 1.0000 | 1.0000 | 0.9901 | 1.0000 | 0.9950 |
| Recall | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9900 | 0.9950 |
| F1-score | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Accuracy | 100.00 | | | 99.67 | | |

Table A.3: Classification results for the high phase-change regime. Accuracy, precision, recall, and F1-score are reported for the vibrating (vib), static, and clutter classes.

| **Metric** | **Xcorr1 → Xcorr1** | | | **Xcorr2 → Xcorr2** | | |
|---|---|---|---|---|---|---|
| | Vib. | Static | Clutter | Vib. | Static | Clutter |
| Precision | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Recall | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| F1-score | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Accuracy | 100.00 | | | 100.00 | | |

Table A.4: Classification results for the Full range and tested on low, medium, and high phase-change regime. Accuracy, precision, recall, and F1-score are reported for the vibrating (vib), static, and clutter classes.

| **Train** | **Full Range Xcorr1** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Test** | **Low Xcorr1** | | | **Middle Xcorr1** | | | **High Xcorr1** | | |
| **Metric** | **Vib.** | **Static** | **Clutter** | **Vib.** | **Static** | **Clutter** | **Vib.** | **Static** | **Clutter** |
| Precision | 1.0000 | 0.9901 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Recall | 0.9900 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| F1-score | 0.9950 | 0.9950 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Accuracy | 99.67 | | | 100.00 | | | 100.00 | | |

Table A.5: Classification results for Full range and tested on low, medium, and high phase-change regime. Accuracy, precision, recall, and F1-score are reported for the vibrating (vib), static, and clutter classes.

| **Train** | **Full Range Xcorr2** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Test** | **Low Xcorr2** | | | **Middle Xcorr2** | | | **High Xcorr2** | | |
| **Metric** | **Vib.** | **Static** | **Clutter** | **Vib.** | **Static** | **Clutter** | **Vib.** | **Static** | **Clutter** |
| Precision | 1.0000 | 0.9901 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Recall | 0.9900 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| F1-score | 0.9950 | 0.9950 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Accuracy | 99.67 | | | 100.00 | | | 100.00 | | |

Table A.6: Classification results for the concatenation of features from Xcorr1 and Xcorr2 on the full range of phase differences. Accuracy, precision, recall, and F1-score are reported for the vibrating (vib), static, and clutter classes.

| **Metric** | **[Xcorr1,Xcorr2]** → **[Xcorr1,Xcorr2]** | | |
|---|---|---|---|
| | Vib. | Static | Clutter |
| Precision | 0.9933 | 0.9868 | 1.0000 |
| Recall | 0.9867 | 0.9933 | 1.0000 |
| F1-score | 0.9900 | 0.9900 | 1.0000 |
| Accuracy | 99.33 | | |

# Appendix B

# List of Publications

- McCleary, J., García, L. P., Ilioudis, C., & Clemente, C. (2021, May). Sign language recognition using micro-Doppler and explainable deep learning. In 2021 IEEE Radar Conference (RadarConf21) (pp. 1-6). IEEE.
- Garcia, L. P., Pallotta, L., Clemente, C., Giunta, G., & Soraghan, J. J. (2024). A cross-cross-correlation based method for joint coregistration of rotated multitemporal synthetic aperture radar images. IET Radar, Sonar & Navigation, 18(1), 198-209.
- Garcia, L. P., Addabbo, P., Orlando, D., Biondi, F., Furano, G., Imbembo, E., ... & Clemente, C. (2024, April). Maritime target classification from SLC SAR data Spectral Profiles. In EUSAR 2024; 15th European Conference on Synthetic Aperture Radar (pp. 421-426). VDE.
- Garcia, L. P., Furano, G., Ghiglione, M., Zancan, V., Imbembo, E., Ilioudis, C., ... & Trucco, P. (2024). Advancements in on-board processing of synthetic aperture radar (SAR) data: enhancing efficiency and real-time capabilities. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.
- Zavagli, M., Nasso, I., Santi, F., Pastina, D., Vecchioli, F., Minati, F., Constantini, M., Parra Garcia, L., Clemente, C., & Corvino, M. (2024, July). Advanced ISAR processing applied to VHR SAR data for security applications. In IGARSS

2024-2024 IEEE International Geoscience and Remote Sensing Symposium (pp. 1173-1177). IEEE.

- Parra Garcia, L., Macdonald, M., & Clemente, C (2026). Automatic location, characterisation, and classifiaction of vibrating maritime targets using high-order features. Pending publication in The European Radar Conference (EuRAD) 2026, London, UK.

# Bibliography

[1] L. Pallotta, C. Clemente, T. Borreca, G. Giunta, and J. J. Soraghan. A joint coregistration of rotated multitemporal sar images based on the cross-cross-correlation. In *International Conference on Radar Systems (RADAR 2022)*, volume 2022, pages 32–37, 2022.

[2] Laura Parra Garcia, Luca Pallotta, Carmine Clemente, Gaetano Giunta, and John J Soraghan. A cross-cross-correlation based method for joint coregistration of rotated multitemporal synthetic aperture radar images. *IET Radar, Sonar & Navigation*, 18(1):198–209, 2024.

[3] Laura Parra Garcia, Pia Addabbo, Danilo Orlando, Filippo Biondi, Gianluca Furano, Ernesto Imbembo, Christos Ilioudis, and Carmine Clemente. Maritime target classification from slc sar data spectral profiles. In *EUSAR 2024; 15th European Conference on Synthetic Aperture Radar*, pages 421–426. VDE, 2024.

[4] M. M. Woolfson. *The fundamentals of imaging: from particles to galaxies*. World Scientific, London, 2012.

[5] ESA. Radar course 3 - synthetic aperture radar - ers radar course 3 - esa operational eo missions - earth online - esa, 2020.

[6] Scheer and Melvin. *Principles of Modern Radar: Volume 3: Radar Applications*. 2013.

[7] Yan Lu, Siwei Kou, and Xiaopeng Wang. Micro-doppler effect and sparse representation analysis of underwater targets. *Sensors*, 23(19):8066, 2023.

[8] V. C. Chen. *The Micro-Doppler effect in radar.* Artech House, 2011.

[9] Floyd M Henderson and Anthony J Lewis. Principles and applications of imaging radar. manual of remote sensing, volume 2. 1998.

[10] Robert N McDonough. *Synthetic aperture radar: systems and signal processing.* Wiley, 1991.

[11] Seyed Alireza Khoshnevis and Seyed Ghorshi. A tutorial on tomographic synthetic aperture radar methods. *SN Applied Sciences*, 2(9):1504, 2020.

[12] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine*, 1(1):6–43, 2013.

[13] Baolong Wu, Chengjin Liu, and Jianlai Chen. A review of spaceborne high-resolution spotlight/sliding spotlight mode sar imaging. *Remote Sensing*, 17(1), 2025.

[14] Paul A Rosen, Scott Hensley, Ian R Joughin, Fuk K Li, Soren N Madsen, Ernesto Rodriguez, and Richard M Goldstein. Synthetic aperture radar interferometry. *Proceedings of the IEEE*, 88(3):333–382, 2000.

[15] Didier Massonnet and Kurt L Feigl. Radar interferometry and its application to changes in the earth's surface. *Reviews of geophysics*, 36(4):441–500, 1998.

[16] R Keith Raney. Hybrid-polarity sar architecture. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11):3397–3404, 2007.

[17] Capella Space. What are capella's collect modes?, 2024.

[18] Umbra Lab Inc. Remote sensing, 2024. Accessed: 2024-11-14.

[19] ICEYE. Types of sar collection - dwell imaging. `https://sar.iceye.com/5.2.0/productguide/typesofsarcollection/#dwell-imaging`, 2024. Accessed: 2024-11-14.

[20] Marwan Younis. 6th esa advanced training course on land remote sensing- synthetic aperture radar (sar): Principles and applications. `https://eo4society.esa.int/wp-content/uploads/2021/02/D1T2a_LTC2015_Younis.pdf`, 2015.

[21] Anup Parashar. *A Study on Range Cell Migration Correction in SAR Imagery and MATLAB Implementation of Algorithms.* PhD thesis, Department of Electronics and Communication Engineering National Institute of Technology, Rourkela Rourkela, Odisha, India, 2015.

[22] Ian G Cumming and Frank H Wong. Digital processing of synthetic aperture radar data. *Artech house*, 1(3):108–110, 2005.

[23] L. Pallotta, G. Giunta, and C. Clemente. Sar image registration in the presence of rotation and translation: A constrained least squares approach. *IEEE Geoscience and Remote Sensing Letters*, 18(9):1595–1599, 2021.

[24] Gonzalo Simarro, Daniel Calvete, Francesca Ribas, Yeray Castillo, and Càrol Puig-Polo. Uortos: Methodology for co-registration and subpixel georeferencing of satellite imagery for coastal monitoring. *Remote Sensing*, 17(7), 2025.

[25] Luca Pallotta, Carmine Clemente, Gaetano Giunta, and John J Soraghan. Coregistration method for rotated/shifted fopen sar images. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 2454–2457. IEEE, 2022.

[26] Oktay Karakuş, Igor Rizaev, and Alin Achim. Ship wake detection in sar images via sparse regularization. *arXiv preprint arXiv:1904.03309*, 2019.

[27] Yoon-Kyung Lee, Hahn Chul Jung, Keunyong Kim, Yeongjae Jang, Joo-Hyung Ryu, and Sang-Wan Kim. Assessment of maritime vessel detection and tracking using integrated sar imagery and ais/v-pass data. *Ocean Science Journal*, 59(2):27, 2024.

[28] Shakila Kahar, Fengming Hu, and Feng Xu. Ship detection in complex environment using sar time series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:3552–3563, 2022.

[29] Ms P. S. Tondewad and Ms M. P. Dale. Remote sensing image registration methodology: Review and discussion. *Procedia Computer Science*, 171:2390–2399, 2020.

[30] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004.

[31] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin. Sar-sift: A sift-like algorithm for sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1):453–466, 2015.

[32] S. Paul and U. C. Pati. A block-based multifeature extraction scheme for sar image registration. *IEEE Geoscience and Remote Sensing Letters*, 15(9):1387–1391, September 2018.

[33] Carlos Santamaria, Marlene Alvarez, Harm Greidanus, Vasileios Syrris, Pierre Soille, and Pietro Argentieri. Mass processing of sentinel-1 images for maritime surveillance. *Remote Sensing*, 9(7), 2017.

[34] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359, Jun. 2008.

[35] H. Goncalves, L. Corte-Real, and J. A. Goncalves. Automatic image registration through image segmentation and sift. *IEEE Transactions on Geoscience and Remote Sensing*, 49(7):2589–2600, July 2011.

[36] Guili Tang, Zhonghao Wei, and Long Zhuang. Sar image registration: The combination of nonlinear diffusion filtering, hessian features and edge points. *Sensors*, 24(14):4568, 2024.

[37] Wannan Zhang. Robust registration of sar and optical images based on deep learning and improved harris algorithm. *Scientific Reports*, 12(1):5901, 2022.

[38] Mohammad Saadatseresht and Amin Ghannadi. Efficient method for outlier removal in sar image matching based on epipolar geometry. *IET Radar, Sonar & Navigation*, 12, 08 2018.

[39] Abdelhameed Eltanany, Ahmed Amein, and M. Safy. Sar images co-registration based on gradient descent optimization. 9, 12 2019.

[40] E. Sansosti, P. Berardino, M. Manunta, F. Serafino, and G. Fornaro. Geometrical sar image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 44(10):2861–2870, October 2006.

[41] D. Li and Y. Zhang. A fast offset estimation approach for insar image subpixel registration. *IEEE Geoscience and Remote Sensing Letters*, 9(2):267–271, 2011.

[42] Jiaxing Chen, Hongtu Xie, Lin Zhang, Jun Hu, Hejun Jiang, and Guoqian Wang. Sar and optical image registration based on deep learning with co-attention matching module. *Remote Sensing*, 15(15), 2023.

[43] L. Pallotta, G. Giunta, and C. Clemente. Subpixel sar image registration through parabolic interpolation of the 2-d cross correlation. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6):4132–4144, 2020.

[44] Andrey Pleskachevsky, Björn Tings, Stefan Wiehle, James Imber, and Sven Jacobsen. Multiparametric sea state fields from synthetic aperture radar for maritime situational awareness. *Remote Sensing of Environment*, 280:113200, 2022.

[45] E.H. Peterson, G. Fotopoulos, A. Schmitt, R.E. Zee, and A. Roth. Registration of multi-frequency sar imagery using phase correlation methods. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 3708–3711, 2011.

[46] Yifan Zhang, Zhiwei Li, Wen Wang, Minzheng Mu, and Bangwei Zuo. A robust registration method for multi-view sar images based on best buddy similarity. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48:881–886, 2024.

[47] Gang Hong and Yun Zhang. Combination of feature-based and area-based image registration technique for high resolution remote sensing image. In *2007 IEEE international geoscience and remote sensing symposium*, pages 377–380. IEEE, 2007.

[48] Y. Li, X. Zhang, Y. Zhang, and H. Luo. A survey on sar and optical satellite image registration. *Remote Sensing*, 15(3):850, 2023.

[49] Wenfei Zhang, Ruipeng Zhao, Yongxiang Yao, Yi Wan, Peihao Wu, Jiayuan Li, Yansheng Li, and Yongjun Zhang. Multi-resolution sar and optical remote sensing image registration methods: A review, datasets, and future perspectives. *arXiv preprint arXiv:2502.01002*, 2025.

[50] Leilei Jia, Jian Dong, Siyuan Huang, Limin Liu, and Junning Zhang. Optical and sar image registration based on multi-scale orientated map of phase congruency. *Electronics*, 12(7):1635, 2023.

[51] Keke Zhang, Anxi Yu, Wenhao Tong, and Zhen Dong. A robust sar-optical heterologous image registration method based on region-adaptive keypoint selection. *Remote Sensing*, 16(17):3289, 2024.

[52] A. Lopes, R. Touzi, and E. Nezry. Adaptive speckle filters and scene heterogeneity. *IEEE Transactions on Geoscience and Remote Sensing*, 28(6):992–1000, 1990.

[53] Qingqing Li, Di Zhang, Zhikai Yu, Panhu Li, and Jiaqi Li. A matching method for low snr sar images. In *2021 2nd China International SAR Symposium (CISS)*, pages 1–4, 2021.

[54] H.A. Zebker and J. Villasenor. Decorrelation in interferometric radar echoes. *IEEE Transactions on Geoscience and Remote Sensing*, 30(5):950–959, 1992.

[55] Howard A Zebker, John Villasenor, et al. Decorrelation in interferometric radar echoes. *IEEE Transactions on geoscience and remote sensing*, 30(5):950–959, 1992.

[56] Andrew Hooper, David Bekaert, Karsten Spaans, and Mahmut Arıkan. Recent advances in sar interferometry time series analysis for measuring crustal deformation. *Tectonophysics*, 514:1–13, 2012.

[57] Thomas P Ager. An introduction to synthetic aperture radar imaging. *Oceanography*, 26(2):20–33, 2013.

[58] Yuanxin Ye, Chao Yang, Bai Zhu, Liang Zhou, Youquan He, and Huarong Jia. Improving co-registration for sentinel-1 sar and sentinel-2 optical images. *Remote Sensing*, 13(5), 2021.

[59] Weihao Jiang, Anxi Yu, Zhen Dong, and Qingsong Wang. Comparison and analysis of geometric correction models of spaceborne sar. *Sensors*, 16(7):973, 2016.

[60] Carmine Clemente, Alessio Balleri, Karl Woodbridge, and John J Soraghan. Developments in target micro-doppler signatures analysis: radar imaging, ultrasound and through-the-wall radar. *EURASIP Journal on Advances in Signal Processing*, 2013(1):1–18, 2013.

[61] Finlay Rollo, Christos Ilioudis, and Carmine Clemente. Advances in micro-doppler processing in synthetic aperture radar: A review of techniques, results, and future trends. *IEEE Geoscience and Remote Sensing Magazine*, 2025.

[62] European Space Agency. Satellite frequency bands.

[63] M. Ruegg, E. Meier, and D. Nuesch. Vibration and rotation in millimeter-wave sar. *IEEE Transactions on Geoscience and Remote Sensing*, 45(2):293–304, 2007.

[64] Filippo Biondi, Pia Addabbo, Silvia Liberata Ullo, Carmine Clemente, and Danilo Orlando. Perspectives on the structural health monitoring of bridges by synthetic aperture radar. *Remote Sensing*, 12(23), November 2020.

[65] Alessandro Lotti, Aleksanteri B Vattulainen, Sebastian Diaz Riofrio, Chiara Suppi, Enrico Tubaldi, Daniele Zonta, Pietro Milillo, and Carmine Clemente.

Monitoring bridge vibrations via spaceborne sar micro-doppler. *Structural Control and Health Monitoring*, 2026(1):3858095, 2026.

[66] Aleksanteri B Vattulainen, Alessandro Lotti, Sebastián Díaz Riofrío, Chiara Suppi, Enrico Tubaldi, Daniele Zonta, Pietro Milillo, and Carmine Clemente. Assessment of spaceborne sar micro-motion measurement for vibration-based shm. *IEEE Access*, 14:6043–6064, 2026.

[67] Vitali Kozlov, Sergey Kosulnikov, Dmitry Filonov, Andrey Schmidt, and Pavel Ginzburg. Coupled micro-doppler signatures of closely located targets. *Physical Review B*, 100(21):214308, 2019.

[68] Finlay Rollo, Aleksanteri B Vattulainen, Christos Ilioudis, Pietro Milillo, and Carmine Clemente. Scale invariant coherent change detection to locate micro-motion in single pass sar images. In *2024 IEEE International Workshop on Technologies for Defense and Security (TechDefense)*, pages 58–62. IEEE, 2024.

[69] Francisco German Perez Venegas. Detection and classification of vibrating objects in sar images. 2019.

[70] R. K. Raney. Synthetic aperture imaging radar and moving targets. *IEEE Transactions on Aerospace and Electronic Systems*, AES-7(3):499–505, 1971.

[71] Mouloud Ayad, Mohamed Rezki, Kamel Saoudi, Mourad Benziane, Abderrazak Arabi, and Djamel Chikouche. Wavelet transforms coefficients and autocorrelation of gear system for early damage detection. In *2015 7th International Conference on Modelling, Identification and Control (ICMIC)*, pages 1–6. IEEE, 2015.

[72] Xining Zhang, Rongtong Zhou, and Wenwen Zhang. Improved local cepstrum and its applications for gearbox and rolling bearing fault detection. *Measurement Science and Technology*, 30(7):075007, 2019.

[73] Omri Matania, Lior Bachar, Eric Bechhoefer, and Jacob Bortman. Signal processing for the condition-based maintenance of rotating machines via vibration analysis: A tutorial. *Sensors*, 24(2):454, 2024.

[74] Hongyu Yuan, Rui Zhao, Yu Wang, Qing Bai, Hongjuan Zhang, Yan Gao, and Baoquan Jin. Long-distance detection for periodic vibration signal in $\phi$-otdr system using global phase demodulation method. *IEEE Sensors Journal*, 21(23):26799–26804, 2021.

[75] Jin Yang, Yumei Wen, and Ping Li. Information processing for leak detection on underground water supply pipelines. In *Third International Workshop on Advanced Computational Intelligence*, pages 623–629. IEEE, 2010.

[76] Guocheng Yang and Huimin Yu. Vital signs detection based on uwb radar using trajectory capture and peak capture. In *2020 3rd International Conference on Electron Device and Mechanical Engineering (ICEDME)*, pages 657–661. IEEE, 2020.

[77] Weibo Huo, Yulin Huang, Jifang Pei, Qian Zhang, Qin Gu, and Jianyu Yang. Ship detection from ocean sar image based on local contrast variance weighted information entropy. *Sensors*, 18(4):1196, 2018.

[78] A. K. Jain, R. P. W. Duin, and J. C. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, January 2000.

[79] H. Chiang, R. Moses, and L. Potter. Model-based bayesian feature matching with application to synthetic aperture radar target recognition. *Pattern Recognition*, 34(8):1539–1553, 2001.

[80] C. Nilubol, Q. Pham, R. Mersereau, M. Smith, and M. Clements. Hidden markov modelling for sar automatic target recognition. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 1061–1064, May 1998.

[81] R. English, S. Rawlinson, and N. M. Sandirasegaram. Development of an atr workbench for sar imagery. Technical Report DRDC-OTTAWA-TR-2002-155, Defence Research and Development Canada Ottawa, Ottawa, ON, Canada, 2002.

[82] Khalid El-Darymli, Eric W Gill, Peter Mcguire, Desmond Power, and Cecilia Moloney. Automatic target recognition in synthetic aperture radar imagery: A state-of-the-art review. *IEEE access*, 4:6014–6058, 2016.

[83] Günter Saur and Michael Teutsch. Sar signature analysis for terrasar-x-based ship monitoring. In *Image and Signal Processing for Remote Sensing XVI*, volume 7830, pages 518–525. SPIE, 2010.

[84] Carlos Santamaria, Harm Greidanus, Mélanie Fournier, Torkild Eriksen, Michele Vespe, Marlene Alvarez, Virginia Fernandez Arguedas, Conor Delaney, and Pietro Argentieri. Sentinel-1 contribution to monitoring maritime activity in the arctic. In *Proceedings of the ESA Living Planet Symposium, Prague, Czech Republic*, pages 9–13, 2016.

[85] Henning Heiselberg. Ship-iceberg detection & classification in sentinel-1 sar images. In *13th International Conference on Marine Navigation and Safety of Sea Transportation*, 2019.

[86] Capella Space. Sar imagery products. https://www.capellaspace.com/.

[87] Adugna G. Mullissa, C. Persello, and A. Stein. Polsarnet: A deep fully convolutional network for polarimetric sar image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12:5300–5309, 2019.

[88] Laura Parra Garcia, Gianluca Furano, Max Ghiglione, Valentina Zancan, Ernesto Imbembo, Christos Ilioudis, Carmine Clemente, and Paolo Trucco. Advancements in on-board processing of synthetic aperture radar (sar) data: enhancing efficiency and real-time capabilities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[89] Pan Xu, Qingyang Li, Bo Zhang, Fan Wu, Ke Zhao, Xin Du, Cankun Yang, and Ruofei Zhong. On-board real-time ship detection in hisea-1 sar images based on cfar and lightweight deep learning. *Remote Sensing*, 13(10), 2021.

[90] Muhammed Burak Alver, Sara Atito, and M. Çetin. Sar atr in the phase history domain using deep convolutional neural networks. In *Remote Sensing*, 2018.

[91] Yinjie Xie, Wenxin Dai, Zhenxin Hu, Yijing Liu, Chuan Li, and Xuemei Pu. A novel convolutional neural network architecture for sar target recognition. *Journal of Sensors*, 2019:1–9, 2019.

[92] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

[93] Zezhou Yang and Shaodi Jing. Sar image classification method based on improved capsule network. In *Journal of Physics: Conference Series*, volume 1693, page 012181. IOP Publishing, 2020.

[94] Yikui Zhai, Hui Ma, Jian Liu, Wenbo Deng, Lijuan Shang, Bing Sun, Ziyi Jiang, Huixin Guan, Yihang Zhi, Xi Wu, et al. Sar atr with full-angle data augmentation and feature polymerisation. *The Journal of Engineering*, 2019(19):6226–6230, 2019.

[95] Jiayuan Kong and Fangzheng Zhang. Sar target recognition with generative adversarial network (gan)-based data augmentation. In *2021 13th International Conference on Advanced Infocomm Technology (ICAIT)*, pages 215–218. IEEE, 2021.

[96] Maha Al Mufti, Esra Al Hadhrami, Bilal Taha, and Naoufel Werghi. Sar automatic target recognition using transfer learning approach. In *2018 International Conference on Intelligent Autonomous Systems (ICoIAS)*, pages 1–4. IEEE, 2018.

[97] Jianwei Li, Zhentao Yu, Lu Yu, Pu Cheng, Jie Chen, and Cheng Chi. A comprehensive survey on sar atr in deep-learning era. *Remote Sensing*, 15(5), 2023.

[98] Mohammad Rostami, Soheil Kolouri, Eric Eaton, and Kyungnam Kim. Deep transfer learning for few-shot sar image classification. *Remote Sensing*, 11(11):1374, 2019.

[99] Haitao Lang, Siwen Wu, and Yongjie Xu. Ship classification in sar images improved by ais knowledge transfer. *IEEE Geoscience and Remote Sensing Letters*, 15(3):439–443, 2018.

[100] Carmine Clemente, Luca Pallotta, Domenico Gaglione, Antonio De Maio, and John J. Soraghan. Automatic target recognition of military vehicles with krawtchouk moments. *IEEE Transactions on Aerospace and Electronic Systems*, 53(1):493–500, 2017.

[101] Air Force Research Laboratory. Moving and stationary target acquisition and recognition (mstar) dataset. `https://www.sdms.afrl.af.mil/index.php?collection=mstar`, 1995. Accessed: 2024-12-04.

[102] Carmine Clemente, Luca Pallotta, Domenico Gaglione, Antonio De Maio, and John J Soraghan. On the use of image moments for atr from sar images. *NATO SET on Radar Imaging and Target Identification*, 18, 2015.

[103] Zhixu Wang, Zhihui Xin, Xiaoqiao Huang, Yu Sun, and Jiayu Xuan. Overview of sar image feature extraction and target recognition. In *3D Imaging Technologies—Multi-dimensional Signal Processing and Deep Learning: Mathematical Approaches and Applications, Volume 1*, pages 69–75. Springer, 2021.

[104] Pan Xu, Qingyang Li, Bo Zhang, Fan Wu, Ke Zhao, Xin Du, Cankun Yang, and Ruofei Zhong. On-board real-time ship detection in hisea-1 sar images based on cfar and lightweight deep learning. *Remote Sensing*, 13(10):1995, 2021.

[105] Ruochen Wu. Two-parameter cfar ship detection algorithm based on rayleigh distribution in sar images. 2021.

[106] Kefeng Ji, Xiangwei Xing, Huanxin Zou, and Jixiang Sun. A novel variable index and excision cfar based ship detection method on sar imagery. *Journal of Sensors*, 2015(1):437083, 2015.

[107] Hicham Madjidi, Toufik Laroussi, and Nedjma Detouche. Cfar detection of ships based on median absolute deviation for sar imagery in a heterogeneous weibull sea clutter. pages 1–6, 2024.

[108] Jiaqiu Ai, Xuezhi Yang, Fang Zhou, Zhangyu Dong, Lu Jia, and He Yan. A correlation-based joint cfar detector using adaptively-truncated statistics in sar imagery. *Sensors*, 17(4):686, 2017.

[109] Odysseas A Pappas, Alin M Achim, and David R Bull. Superpixel-guided cfar detection of ships at sea in sar imagery. pages 1647–1651, 2017.

[110] Ming Liu, Shichao Chen, Fugang Lu, Mengdao Xing, and Jingbiao Wei. Realizing target detection in sar images based on multiscale superpixel fusion. *Sensors*, 21(5):1643, 2021.

[111] Xue Wen, Shaoming Zhang, Jianmei Wang, Tangjun Yao, and Yan Tang. A cfar-enhanced ship detector for sar images based on yolov5s. *Remote Sensing*, 16(5):733, 2024.

[112] Jian Li, Yu Liu, Xueqian Wang, Zhizhuo Jiang, and Yaowen Li. A robust cfar algorithm based on superpixel merging operation for sar ship detection. In *Proceedings of the 2024 7th International Conference on Image and Graphics Processing*, pages 388–393, 2024.

[113] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IRE transactions on information theory*, 8(2):179–187, 1962.

[114] Pew-Thian Yap, Raveendran Paramesran, and Seng-Huat Ong. Image analysis by krawtchouk moments. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 12:1367–77, 02 2003.

[115] Cecilia Di Ruberto, Andrea Loddo, and Lorenzo Putzu. On the potential of image moments for medical diagnosis. *Journal of Imaging*, 9(3), 2023.

[116] Amy Chiang and Simon X Liao. Image analysis with legendre moment descriptors. *J. Comput. Sci.*, 11(1):127–136, 2015.

[117] Maofu Liu, Yanxiang He, and Bin Ye. Image zernike moments shape feature evaluation based on image reconstruction. *Geo-spatial Information Science*, 10(3):191–195, 2007.

[118] Satish Kumar. Comparative results of zernike moments and pseudo-zernike moments. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 1254–1259, 2016.

[119] Parminder Kaur, Husanbir Singh Pannu, and Avleen Kaur Malhi. Comprehensive study of continuous orthogonal moments—a systematic review. 52(4), 2019.

[120] Ramakrishnan Mukundan, SH Ong, and Poh Aun Lee. Image analysis by tchebichef moments. *IEEE Transactions on image Processing*, 10(9):1357–1364, 2001.

[121] Ramakrishnan Mukundan and KR Ramakrishnan. *Moment functions in image analysis: theory and applications.* World scientific, 1998.

[122] Pew Thian Yap. *Image analysis by discrete orthogonal moments.* PhD thesis, Jabatan Kejuruteraan Elektrik, Fakulti Kejuruteraan, Universiti Malaya, 2003.

[123] also LC Biedenharn, J Mi Blatt, and ME Rose. Some properties of the racah and associated coefficients. *Reviews of Modern Physics*, 24(4):249, 1952.

[124] Cecilia Di Ruberto, Andrea Loddo, and Lorenzo Putzu. On the potential of image moments for medical diagnosis. *Journal of Imaging*, 9(3):70, 2023.

[125] Ramakrishnan Mukundan. Some computational aspects of discrete orthonormal moments. *IEEE Transactions on image processing*, 13(8):1055–1059, 2004.

[126] Amir Tahmasbi, Fatemeh Saki, and Shahriar B Shokouhi. Classification of benign and malignant masses based on zernike moments. *Computers in biology and medicine*, 41(8):726–735, 2011.

[127] Luo Zhu, Jiaping Liao, Xiaoqin Tong, Li Luo, Bo Fu, and Guojun Zhang. Image analysis by modified krawtchouk moments. In *Advances in Neural Networks–*

*ISNN 2009: 6th International Symposium on Neural Networks, ISNN 2009 Wuhan, China, May 26-29, 2009 Proceedings, Part III 6*, pages 310–317. Springer, 2009.

[128] P.T. Yap, P. Raveendran, and S.H. Ong. Krawtchouk moments as a new set of discrete orthogonal moments for image reconstruction. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, volume 1, pages 908–912 vol.1, 2002.

[129] Jing Tang Xing. Chapter 2 - cartesian tensor and matrix calculus. In Jing Tang Xing, editor, *Fluid-Solid Interaction Dynamics*, pages 43–55. Academic Press, 2019.

[130] Barmak Honarvar Shakibaei Asli and Maryam Horri Rezaei. Four-term recurrence for fast krawtchouk moments using clenshaw algorithm. *Electronics*, 12(8), 2023.

[131] Christoph Bandt. Small order patterns in big time series: A practical guide. *Entropy*, 21(6), 2019.

[132] Guojun Zhou and John H. L. Hansen. Nonlinear speech analysis and acoustic model adaptation with applications to stress classification and speech recognition. 1999.

[133] Mouloud Ayad, Mohamed Rezki, Kamel Saoudi, Mourad Benziane, Abderrazak Arabi, and Djamel Chikouche. Wavelet transforms coefficients and autocorrelation of gear system for early damage detection. In *2015 7th International Conference on Modelling, Identification and Control (ICMIC)*, pages 1–6, 2015.

[134] Yinwei Li, Qi Wu, Jiawei Jiang, Xia Ding, Qibin Zheng, and Yiming Zhu. A high-frequency vibration error compensation method for terahertz sar imaging based on short-time fourier transform. *Applied Sciences*, 11(22):10862, 2021.

[135] Christopher M Bishop. *Neural networks for pattern recognition.* Oxford university press, 1995.

[136] Zhimian Zhang, Haipeng Wang, Feng Xu, and Ya-Qiu Jin. Complex-valued convolutional neural network and its application in polarimetric sar image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12):7177–7188, 2017.

[137] Stephen Grossberg. Recurrent neural networks. *Scholarpedia*, 8(2):1888, 2013.

[138] Xiao Xiang Zhu, Sina Montazeri, Mohsin Ali, Yuansheng Hua, Yuanyuan Wang, Lichao Mou, Yilei Shi, Feng Xu, and Richard Bamler. Deep learning meets sar, 2020.

[139] S. Kazemi, B. Yonel, and B. Yazıcı. Deep learning for direct automatic target recognition from sar data. *2019 IEEE Radar Conference (RadarConf)*, pages 1–6, 2019.

[140] Todd G Nick and Kathleen M Campbell. Logistic regression. *Topics in biostatistics*, pages 273–301, 2007.

[141] Haiyan Zhao. Target recognition of sar images based on svm and ksrc. *Computational Intelligence and Neuroscience*, 2021(1):4322678, 2021.

[142] Marc Wieland, Wen Liu, and Fumio Yamazaki. Learning change from synthetic aperture radar images: Performance evaluation of a support vector machine to detect earthquake and tsunami-induced changes. *Remote Sensing*, 8(10):792, 2016.

[143] Mateo Gašparović and Dino Dobrinić. Green infrastructure mapping in urban areas using sentinel-1 imagery. *Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering*, 42(2):337–353, 2021.

[144] Mohammadreza Sheykhmousa, Masoud Mahdianpari, Hamid Ghanbari, Fariba Mohammadimanesh, Pedram Ghamisi, and Saeid Homayouni. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:6308–6325, 2020.

[145] Mahesh Pal. Support vector machines/relevance vector machine for remote sensing classification: A review, 2011.

[146] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, 1(1):6–43, 2013.

[147] Won-Kyung Baek, Euihyun Kim, Ho-Kun Jeon, Kwang-Jae Lee, Sang-Wan Kim, Yoon-Kyung Lee, and Joo-Hyung Ryu. Monitoring maritime ship characteristics using satellite remote sensing data from different sensors. *Ocean Science Journal*, 59(1):8, 2024.

[148] Maximilian Rodger and Raffaella Guida. Mapping dark shipping zones using multi-temporal sar and ais data for maritime domain awareness. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 3672–3675. IEEE, 2022.

[149] Nestor Yague-Martinez, Francesco De Zan, and Pau Prats-Iraola. Coregistration of interferometric stacks of sentinel-1 tops data. *IEEE geoscience and remote sensing letters*, 14(7):1002–1006, 2017.

[150] Zhang Yunjun, Heresh Fattahi, Xiaoqing Pi, Paul Rosen, Mark Simons, Piyush Agram, and Yosuke Aoki. Range geolocation accuracy of c-/l-band sar and its implications for operational stack coregistration. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–19, 2022.

[151] Gaetano Giunta and Umberto Mascia. Estimation of global motion parameters by complex linear regression. *IEEE transactions on image processing*, 8(11):1652–1657, 1999.

[152] L. Pallotta, G. Giunta, C. Clemente, and J. J. Soraghan. Sar coregistration by robust selection of extended targets and iterative outlier cancellation. *IEEE Geoscience and Remote Sensing Letters*, 19(4501405):1–5, 2022.

[153] Luca Pallotta and Giuseppe Giunta. Accurate delay estimation for multisensor passive locating systems exploiting the cross-correlation between signals cross-correlations. *IEEE Transactions on Aerospace and Electronic Systems*, 58(3):2568–2576, June 2022.

[154] Khalid El-Darymli, Peter McGuire, Desmond Power, and Ciara R. Moloney. Target detection in synthetic aperture radar imagery: A state-of-the-art survey. *Journal of Applied Remote Sensing*, 7(1):071598, 2013.

[155] Luca Pallotta, Carmine Clemente, Giuseppe Giunta, and J. J. Soraghan. Coregistration method for rotated/shifted fopen sar images. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 2454–2457, 2022.

[156] Christophe Leys, Christophe Ley, Olivier Klein, Patrick Bernard, and Laurent Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766, 2013.

[157] H. Fitriyah and A. S. Budi. Outlier detection in object counting based on hue and distance transform using median absolute deviation (mad). In *Proceedings of the International Conference on Sustainable Information Engineering and Technology (SIET)*, pages 217–222, September 2019.

[158] C Casteel and L Gorham. Gotcha volumetric sar data set overview.

[159] Basheera M Mahmmod, Alaa M Abdul-Hadi, Sadiq H Abdulhussain, and Aseel Hussien. On computational aspects of krawtchouk polynomials for high orders. *Journal of Imaging*, 6(8):81, 2020.

[160] Shanghai Key Laboratory of Intelligent Sensing and Recognition. *Instructions of OpenSARShip*. Shanghai Key Laboratory of Intelligent Sensing and Recognition, 2017.

[161] Calibration operator — snap 9.0.0. `https://step.esa.int/main/wp-content/help/versions/9.0.0/snap-toolboxes/org.esa.s1tbx.s1tbx.op.calibration.ui/operators/CalibrationOp.html`, 2021.

[162] Ridha Touzi, Jeff Hurley, and Paris W Vachon. Optimization of the degree of polarization for enhanced ship detection using polarimetric radarsat-2. *IEEE Transactions on Geoscience and Remote Sensing*, 53(10):5403–5424, 2015.

[163] Motofumi Arii. Ship detection from full polarimetric sar data at different incidence angles. In *2011 3rd International Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, pages 1–4. IEEE, 2011.

[164] The MathWorks Inc. What is meant by a recurrent neural network?

[165] Robert DiPietro and Gregory D Hager. Deep learning: Rnns and lstm. In *Handbook of medical image computing and computer assisted intervention*, pages 503–519. Elsevier, 2020.

[166] MathWorks. *LSTM Layer Documentation*. MathWorks, 2024.

[167] MathWorks. *Fully Connected Layer Documentation*. MathWorks, 2024.

[168] MathWorks. *Softmax Layer Documentation*. MathWorks, 2024.

[169] MathWorks. *Classification Layer Documentation*. MathWorks, 2024.

[170] Meysam Vakili, Mohammad Ghamsari, and Masoumeh Rezaei. Performance analysis and comparison of machine and deep learning algorithms for iot data classification. *arXiv preprint arXiv:2001.09636*, 2020.

[171] Newsom. Multinomial logistic regression models. `https://web.pdx.edu/~newsomj/cdaclass/ho_multinomial.pdf`, 2021.

[172] Jinglu He, Wenlong Chang, Fuping Wang, Qian Wang, Yinghua Li, and Yuquan Gan. Polarization matters: On bilinear convolutional neural networks for ship classification from synthetic aperture radar images. In *2022 4th International*

*Conference on Natural Language Processing (ICNLP)*, pages 315–319. IEEE, 2022.

[173] MathWorks. *Constant False Alarm Rate (CFAR) Detection*. MathWorks, 2025. Online documentation.

[174] Michael F Rimbert. *Constant false alarm rate detection techniques based on empirical distribution function statistics*. Purdue University, 2005.

[175] Giorgio Guerzoni, Elahe Faghand, Loris Vincenzi, Elisa Bassoli, and Giorgio Matteo Vitetta. A novel doppler-based phase unwrapping algorithm for mmwave mimo radars and its application to displacement estimation in structural health monitoring. *Mechanical Systems and Signal Processing*, 235:112777, 2025.

[176] Francesco De Zan and Biondo Biondi. Phase unwrapping of angle-domain common image gathers. Sep report 131, Stanford Exploration Project, Oct 2007. Available online: `https://sep.stanford.edu/data/media/public/docs/sep131/book.pdf`.

[177] Eugene A. Morelli and Jared A. Grauer. Practical aspects of the frequency domain approach for aircraft system identification. In *Proceedings of the AIAA Aviation Forum*, Atlanta, GA, USA, 2018. American Institute of Aeronautics and Astronautics. NASA Technical Report Server Document ID: 20190001862.

[178] Petre Stoica, Randolph L Moses, et al. *Spectral analysis of signals*, volume 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[179] Piet MT Broersen. Length and quality of lagged-product autocorrelation estimates. In *2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007*, pages 1–6. IEEE, 2007.

[180] Do-Hyun Park, Min-Wook Jeon, Da-Min Shin, and Hyoung-Nam Kim. Lpi radar detection based on deep learning approach with periodic autocorrelation function. *Sensors*, 23(20), 2023.

[181] MathWorks, Inc. Find periodicity using autocorrelation. `https://www.mathworks.com/help/signal/ug/find-periodicity-using-autocorrelation.html`, 2025. Accessed: 2025-01-XX.

[182] Spectral Dynamics. Understanding phase and coherence in vibration control testing. `https://www.spectraldynamics.com/support/technical-library/understanding-phase-and-coherence-in-vibration-control-testing`, 2025. Accessed: 2025-01-XX.

[183] U.S. Department of Commerce, National Oceanic and Atmospheric Administration. Doppler radar meteorological observations: Part b – doppler radar theory and meteorology. Technical Report FMH-11B, Office of the Federal Coordinator for Meteorological Services and Supporting Research, Washington, DC, USA, 2005. Autocorrelation technique described on p.2-11.

[184] Juan Pablo Navarro Castillo, Rolf Scheiber, Marc Jäger, and Alberto Moreira. Sub-aperture motion-adaptive reconstruction techniques for digital beamforming airborne sar. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[185] Hamid Karimi-Rouzbahani, Mozhgan Shahmohammadi, Ehsan Vahab, Saeed Setayeshi, and Thomas Carlson. Temporal variabilities provide additional category-related information in object category decoding: A systematic comparison of informative eeg features. *Neural Computation*, 33(11):3027–3072, 10 2021.

[186] Linsheng Bu, Defeng Chen, Tuo Fu, Huawei Cao, and Wanyu Chang. Transformer architecture for micromotion target detection based on multi-scale subaperture coherent integration. *Remote Sensing*, 17(3), 2025.

[187] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics.* John Wiley & Sons, Hoboken, NJ, USA, 2012.

[188] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[189] Himanshu Misra, Syed Ikbal, Hervé Bourlard, and Hynek Hermansky. Spectral entropy based feature for robust ASR. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 193–196, 2004.

[190] Alain De Cheveigné and Hideki Kawahara. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.

[191] Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014.

[192] MathWorks. ClassificationECOC — multiclass model for support vector machines (svms) and other classifiers. `https://uk.mathworks.com/help/stats/classificationecoc.html`, 2026. Accessed: 2026-01-11.

[193] John Shawe-Taylor and Shiliang Sun. A review of optimization methodologies in support vector machines. *Neurocomputing*, 74(17):3609–3618, 2011.

[194] Hao Gao and Xu Dong Zhang. Automatic radar waveform recognition using svm. In *Mechanical and Electrical Technology IV*, volume 229 of *Applied Mechanics and Materials*, pages 2348–2351. Trans Tech Publications Ltd, 11 2012.

[195] Author Unknown. Classification of polarimetric sar image based on multiple-component scattering model and texture features using svm. *EURASIP Journal on Advances in Signal Processing*, 2010, 2009.

[196] Kaiyin Yu, Yuanyuan Qi, Lai Shen, Xiaofeng Wang, Daying Quan, and Dongping Zhang. Radar signal recognition based on bagging svm. *Electronics*, 12(24), 2023.

[197] Wenqin Wang. Analytical modeling and simulation of phase noise in bistatic synthetic aperture radar systems. *Fluctuation and noise letters*, 6(03):L297–L303, 2006.

[198] Minjun Kim, Ohtae Jang, Haekang Song, Heesub Shin, Jaewoo Ok, Minyoung Back, Jaehyuk Youn, and Sungho Kim. Soft segmented randomization: Enhancing domain generalization in sar atr for synthetic-to-measured. *IEEE Access*, 2024.

[199] Guillermo Garza. Mathematics of synthetic aperture radar imaging. 2011.

[200] John C.. Curlander and Robert N McDonough. *Synthetic aperture radar: systems and signal processing*. Wiley., 1991.

[201] Ian G Cumming and Frank H Wong. Digital processing of synthetic aperture radar data. *Artech house*, 1(3):108–110, 2005.

[202] Jach. radar_sar_rma: A python implementation of the range migration algorithm (rma) to generate synthetic aperture radar (sar) images. GitHub repository, 2014. `https://github.com/Jach/radar_sar_rma`.

[203] Luke Rosenberg, Simon Watts, and Stephen Bocquet. Application of the k+rayleigh distribution to high grazing angle sea-clutter. In *2014 International Radar Conference*, pages 1–6, 2014.

[204] Michael J Collins and Jeremy M Allan. Modeling and simulation of sar image texture. *IEEE Transactions on Geoscience and Remote Sensing*, 47(10):3530–3546, 2009.

[205] Jaime Martín-de Nicolás, María-Pilar Jarabo-Amores, David Mata-Moya, Nerea del Rey-Maestre, and José-Luis Bárcena-Humanes. Statistical analysis of sar sea clutter for classification purposes. *Remote Sensing*, 6(10):9379–9411, 2014.

[206] Afonso Martini Spezia and Mariana Recamonde-Mendoza. Comparing cluster-based cross-validation strategies for machine learning model evaluation. *arXiv preprint arXiv:2507.22299*, 2025. License: CC BY 4.0.

[207] Sri Yarsasi and Angga Iskoko. A text classification approach for detecting cyberbullying risk on twitter using support vector machine with naive bayes and

random forest comparison. *International Journal of Informatics and Information Systems*, 8(4):175–187, 2025.

[208] MathWorks. *What Is a Support Vector Machine?* MathWorks, 2026. Accessed: 2026-01-10.

[209] Victor C Chen, Fayin Li, S-S Ho, and Harry Wechsler. Micro-doppler effect in radar: phenomenon, model, and simulation study. *IEEE Transactions on Aerospace and electronic systems*, 42(1):2–21, 2006.

[210] Aleksanteri B. Vattulainen, Alessandro Lotti, Sebastián Díaz Riofrío, Chiara Suppi, Enrico Tubaldi, Daniele Zonta, Pietro Milillo, and Carmine Clemente. Assessment of spaceborne sar micro-motion measurement for vibration-based shm. *IEEE Access*, 14:6043–6064, 2026.

[211] International Hydrographic Organization. S-57 appendix a: Object catalogue, edition 3.1. Technical Report 31ApAch1, International Hydrographic Organization, November 2000. IHO Transfer Standard for Digital Hydrographic Data — Object Catalogue.