HOUDA HARKAT

INTELLIGENT SUPPORT SYSTEM FOR GROUND PENETRATING RADAR DATA INVERSION

UNIVERSITY SIDI MOHAMMED BEN ABDELLAH/UNIVERSIDADE DO ALGARVE
Faculty of Sciences and Tehnology, Fez/Faculdade de Cîencias e Tecnologia, Faro
2018

HOUDA HARKAT

INTELLIGENT SUPPORT SYSTEM FOR GROUND PENETRATING RADAR DATA INVERSION

PhD in Applied Physics
(Specialty in Telecommunication)
Work carried out under the supervision of:
Saad Dosse Bennani, António Eduardo de Barros Ruano and Maria da Graça Ruano

UNIVERSITY SIDI MOHAMMED BEN ABDELLAH /UNIVERSIDADE DO ALGARVE
of Sciences and Tehnology, Fez/Faculdade de Cîencias e Tecnologia, Faro
2018

INTELLIGENT SUPPORT SYSTEM FOR GROUND PENETRATING RADAR DATA INVERSION

Declaration of authorship of work

I declare to be the author of this work, which is original and unpublished. Authors and works consulted are duly cited in the text and are included in the list of references included.

_____

Houda Harkat

To my Mother, Aicha
To my brother and sister, Ayoub and Soukaina
To my friends and my family

Acknowledgement

Completing a PhD thesis is a challenge that no one can afford to endure alone. Luckily, on the way to the end of the road, I got invaluable help and encouragement from my family, friends, advisors and many other inspirational individuals.

Primarily, foremost I would like to express my deepest thanks and appreciation to my supervisors, Prof. António Ruano and Prof. Graça Ruano, for having provided me with their expertise, valuable advice, direction and suggestions during the past two years in Portugal. I would like to thank my supervisor from Morocco, Prof. Saad Dosse Bennani, for his time, guidance, encouragements, and support during the years of PhD. Without their inspiration and support, this dissertation would have been impossible.

I would particularly like to acknowledge the University of Algarve for accepting me as a visiting investigator, for giving me the opportunity to do a big part of research in 'Centro de Sistemas Inteligentes' (CSI) of the faculty of sciences and technologies, University of Algarve. I also would like to express my gratitude to the staff of the mobility office of the University of Algarve for all their kind supports.

I wish to acknowledge Author Christian Maas, from the Institute of Geophysics, University of Münster, Germany, for providing a part of data and for helping in the labelling process of patches. I would like to thank authors [1] for making the data of the survey mission public for other researchers. Without their generosity, several tests would not have been executed.

I am also grateful to Mr. Sergio Silva, in CSI Laboratory, for his precious help in solving implementation issues. I would like to thank my friend, Brianna Blackburn, Austin College Alum, for here precious time that she had spent helping me to correct my English.

My sincere appreciation goes to all my professor, for their continuous encouragement and support during the past 8 years of my academic life. They introduced me to the world of academic research and supervised several of my projects during Ing. years.

Finally, yet most importantly, my special thanks goes to my beloved mother who taught me how to think and how to love. Thank you to my friends and family.

Abstract

Ground penetrating radar (GPR) is a high-resolution geophysical sensing technology that does not require prior devices installation, and that is commonly used in a range of mineral exploration applications, and in the military field. It is a non-invasive technique relying on the electromagnetic wave propagation theory, which makes the data inversion process extremely hard since the numerical solution of the wave equations is computationally demanding.

From a pattern recognition view, this task could be solved by the standard fitting algorithms, as the Generalized Hough transform adapted for conic curves. Nevertheless, this yields bad results within the higher resolution radargrams. In literature, this was overcome by a two-stage approach that consists of a discrimination of regions containing hyperbolas (target signatures), before Hough application over these selected patches. The proposed classifiers, for discrimination purpose, are typically very complex with multiple combined stages, which hinders to meet the desired performance for real-time exploitation.

This thesis presents an intelligent support system for ground penetrating radar data inversion, based on a low complexity discrimination system and a 1-D Hough schema, to determine the hyperbolas parameters of GPR targets of B-scan images.

The research hereby reported includes as well a proposal of a 1D Hough schema.

The 1D Hough transform rely on a primordial pre-processing phase that directly affects his performance. A velocity correction step was proposed as well.

For discrimination of GPR hyperbolas, a radial basis functions neural network classifier was designed using a Multi Objective Genetic Algorithm (MOGA). In fact, the classifier structure, parameters, and features were chosen based on the use of different (often conflicting) objectives ensuring maximization of the classification precision and a good generalization of its performance over unseen data.

Several scenarios for choosing best MOGA's architecture were conducted. The best results were obtained from the scenario where a prior Mutual Information feature selection (MIFS) procedure was conducted before supplying the data to MOGA.

Two non-public databases were constructed to test the automatic proposed system. For the first dataset, an accuracy of 88.9% with an False Negative Rate (FN) of 20.69% was obtained over a validation set of 628 samples. For the second database, within a validation set of 592 samples, an accuracy of 95.26%, with an FN of 9.33% was achieved.

Present results show that the MOGA designed RBF classifier achieved better classification results than Support Vector Machines (SVM) and Convolutional Neural Network (CNN), despite the huge difference in complexity of the two classifiers. The proposed approach compares also favourably with other similar published solutions, obtaining the same level of performance or slightly better, with a large decrease of model complexity.

Keywords: Ground Penetrating Radar, Data Inversion, Hough Transform, RBF Neural Networks; Multi-Objective Genetic Algorithm; Mutual Information

Résumé

Le géo-radar (GPR) est une technologie de détection géophysique à haute résolution qui n'exige pas l'installation de dispositifs antérieurs, et qui est généralement utilisée dans les applications d'exploration minérale et dans le domaine militaire. C'est une technique non invasive repose sur la théorie de propagation des ondes électromagnétiques, ce qui rend le processus d'inversion de données extrêmement difficile puisque la solution numérique des équations d'onde demande beaucoup de temps de calculs.

De point de vue de la reconnaissance de formes, cette tâche pourrait être résolue par les algorithmes d'ajustement standard, comme la transformée de Hough généralisée adaptée aux courbes coniques. Néanmoins, cela donne de mauvais résultats dans les radargrammes (images radar GPR ou B-scans) à plus haute résolution. Dans la littérature, ce problème a été surmonté par une approche en deux étapes qui consiste en une discrimination des régions contenant des hyperboles (signatures cibles), avant d'appliquer la transformée de Hough sur les patches sélectionnés. Les classificateurs proposés dans la phase de discrimination, sont généralement très complexes avec plusieurs étapes combinées, ce qui les empêche d'atteindre la performance souhaitée pour une utilisation en temps réel (sur le terrain).

Cette thèse présente un système intelligent d'inversion des données radar pénétrant au sol, basé sur un système de discrimination de faible complexité combiné avec un schéma de Hough unidimensionnel (‘1D Hough’), pour déterminer les paramètres des hyperboles des cibles GPR dans les images B-scan.

La recherche rapportée ici inclut également une proposition d'un schéma Hough unidimensionnel.

La transformation de ‘1D Hough’ repose sur une phase de prétraitement primordiale qui affecte directement ses performances. Une étape de correction de la vitesse a également été proposée.

Pour la discrimination des hyperboles GPR, un classificateur de réseau de neurones à fonction de base radiale (RBF) a été conçu en utilisant un algorithme génétique multi-objectif (MOGA). En effet, la structure, les paramètres et les caractéristiques du classificateur ont été choisis en fonction de l'utilisation d’objectifs multiple (souvent conflictuels) assurant une maximisation de la précision de la classification et une bonne généralisation de ses performances par rapport aux données invisibles (partition de test).

Pour choisir la meilleure architecture, plusieurs scénarios ont été menés en utilisant le Framework MOGA. Les meilleurs résultats ont été obtenus à partir du scénario où une procédure antérieure de sélection d'entités d'informations mutuelles (MIFS) a été effectuée avant de fournir les données à MOGA.

Deux bases de données non publiques ont été construites pour tester le système automatique proposé. Pour la première base de données, une précision de 88,9% avec un taux de faux négatif (FN) de 20,69% a été obtenue sur un ensemble de validation de 628 échantillons. Pour la deuxième base de données avec un ensemble de validation de 592 échantillons, une précision de 95,26%, avec un FN de 9,33% a été atteinte.

Les résultats actuels montrent que le classificateur RBF conçu par MOGA est meilleur en termes de performances de classification que les machines à vecteur de support (SVM) et le réseau de neurones convolutionnels (CNN), malgré l'énorme différence de complexité des deux classificateurs. Les résultats obtenus par l'approche proposée se comparent également favorablement avec d'autres solutions publiées similaires, obtenant le même niveau de performance ou légèrement meilleur, avec une forte diminution de la complexité du modèle.

Mots-clés: Radar pénétrant au sol, inversion de données, transformée de Hough, réseaux de neurones RBF; Algorithme génétique multi-objectif; Informations mutuelles…

ملخص

رادار الاختراق الأرضي (GPR) هو تقنية استشعار جيوفيزيائية عالية الدقة غير مسبقة التثبيت، هذا الأخير يستخدم بشكل شائع في مجموعة تطبيقات استكشاف المعادن، وكذلك في المجال العسكري .

إنها تقنية ليس قابلة للتنقيب حيث تعتمد على نظرية انتشار الموجات الكهرومغناطيسية، مما يجعل عملية عكس البيانات صعبة للغاية.

اعتمادا على طريقة التعرف على الأشكال، يمكن حل هذه الإشكالية من خلال خوارزميات التركيب القياسية، باعتماد تحويل هاف (Hough) في صيغته التي تتكيف مع المنحنيات المخروطية. ولكن في هذه الحالة (شكل هذلول، حيث أن الفضاء الممعلم ثلاثي الأبعاد) فإن هذا يؤدي إلى نتائج سيئة بالنسبة لصور الرادارغرام العالية الدقة .

فمن خلال الدراسات السابقة، تم التغلب على هذه الإشكالية المطروحة بتتبع نهج جديد يتكون مرحلتين تتمحور حول تمييز المناطق التي تحتوي على الهذاليل (التوقيعات المستهدفة، والتي تثبت وجود هدف أو شيء تحت الأرض)، قبل تطبيق هاف على هذه المساحة المحددة .

عادة ما تكون المصنفات المقترحة، لغرض التمييز، معقدة للغاية مع مراحل متعددة مشتركة، والتي تستخدم للاستغلال في الوقت الحقيقي.

تقدم هذه الأطروحة نظاماً ذكياً لعكس بيانات رادار الاختراق الأرضي، استناداً إلى نظام تمييز منخفض التعقيد ومخطط هاف أحادي البعد (هاف1D, 1D Hough)، لتحديد معلمات الهذاليل لأهداف GPR لصور.B-scan

البحث هنا هو اقتراح لمخطط هاف 1D. إن تحويل هاف 1D يرتكز خاصة على مرحلة ما قبل المعالجة الأولية التي تؤثر بشكل مباشر على أدائه. وقد اقترحت خطوة لتصحيح السرعة أيضا.

لتمييز الهذاليل لأهداف GPR ، تم تطوير شبكة دالـة القاعدة الشعاعية (RBF) باستخدام خوارزمية جينية متعددة الأهداف (MOGA).في الواقع، يتم اختيار بنية المصنف ومعلماته وخصائصه بناءً على هدف تحقيق مستوى عالٍ من الدقة. وبالفعل، فقد تم اختيار البنية والمعلمات وخصائص المصنف وفقًا لاستخدام أهداف متعددة (غالباً ما تكون متعارضة) مما يضمن تحقيق مستوى عالٍ من دقة التصنيف والتعميم الجيد لأدائه مع البيانات غير المرئية (قسم الاختبار).

لاختيار أفضل بنية، أجريت عدة سيناريوهات باستخدام إطار MOGA. تم الحصول على أفضل النتائج من السيناريو حيث تم تنفيذ إجراء سابق لانتقاء المعلومات بناءا على مفهوم المعلومات المتبادلة (MIFS) قبل تقديم البيانات إلى MOGA.

أجريت عدة سيناريوهات لاختيار أفضل بنية، تم إنشاء قاعدتي بيانات غير عامتين لاختبار النظام المقترح التلقائي. بالنسبة لمجموعة البيانات الأولى، تم الحصول على دقة قدرها 88.9٪ مع معدل سلبي كاذب (FN) بنسبة 20.69٪، عبر مجموعة تحقق تتكون من 628 عينة. بالنسبة لقاعدة البيانات الثانية، تم الحصول على دقة قدرها 95.26 ٪، مع معدل سلبي كاذب (FN) بنسبة 9.33 ٪، عبر مجموعة تحقق تتكون من 592 عينة.

تظهر النتائج الحالية أن RBF المصمم باستعمال MOGA حقق نتائج تصنيف أفضل من آلات دعم المتجهات (SVM) والشبكة العصبية التحويلية(CNN) ، على الرغم من الاختلاف الكبير في تعقيد المصنفين. كما يقارن النهج المقترح مع الحلول المنشورة المشابهة الأخرى، مع انخفاض كبير في تعقيد النموذج.

كلمات البحث: رادار الإخترق الأرضي، عكس البيانات، تحويل هاف، الشبكات العصبية. خوارزمية جينية متعددة الأهداف، المعلومات المتبادلة

List of Tables

List of Figures

Contents

List of Acronyms

AMM Adaptive Mixtures Method

ANN Artificial Neural Network

AproxHull Approximate convex hull

BP Back Propagation

CNN Convolutional neural Network

CWD Choi-Williams Distribution

EMD Emperical Mode Decomposition

FCM Fuzzy C-Means

FN False Negative

FP False Positive

GA Genetic Algorithm

GPR Ground Penetrating Radar

GLCM Gray Level Co-occurrence Matrix

GM Gray Matter

HHT Hilbert Huang Transform

IMF Intrinsic Model Functions

k-NN k-Nearest Neighbors

LBP Local Binary Pattern

LM Levenberg-Marquardt

MLP Multi-Layer Perceptron

MOGA Multi Objective Genetic Algorithm

MUSIC MUltiple SIgnal Classification

MIFS Mutual Information Feature Selection

RBF Radial Basis Functions

SVM Support Vector Machine

SVD Singular Values Decomposition

SPWVD Smoothed Pseudo Wigner Ville Decomposition

SPAWVD Smoothed Pseudo Affine Wigner Ville Decomposition

WVD Wigner Ville Decomposition

1. Introduction

1. Overview

Geophysics is a non-destructive and non-invasive earth science that aim to measure spatial and temporal variations in the physical properties of the subsoil. Within this discipline, the very latest science and technology innovation in instrumentation, data acquisition, and advanced computer modelling and interpretation are used in subsurface exploration.

Indeed, there are four physical properties to be measured, which involves four types of geophysical methods, but each method has several measurement techniques. The four main physical parameters of a medium are: electrical resistivity, density, wave propagation velocity and magnetic susceptibility. Each of these parameters determines a fundamental geophysical prospecting method: electrical, gravimetric, seismic and magnetic. In addition, one can have hybrid methods such as the electromagnetic methods.

Ground Penetrating Radar (GPR), a very promising electromagnetic technique, requires a deep understanding of the underlying science for successful application. The ground penetration radar is a geophysical method based on the propagation of electromagnetic waves of varying frequencies in the band [1MHz-3GHz]. Electromagnetic waves are reflected or diffracted in the borders of objects with a contrast in electrical and magnetic properties. The dielectric permittivity, electric conductivity, and magnetic permeability are the three parameters that determine the reflectivity of layers and limits the depth of penetration. For the surface-surface prospection, the most common operation mode of GPR is the reflection mode. The traces of returned waves are collected either continuously or in stations along a line, thus creating a time cross-section or an image of the subsurface. These images are called B-scans profiles where reflections of waves are encoded in the form of hyperbolas signatures. These hyperbolic structures contain all the information about depth, position of target and material type.

In general, geophysics survey techniques provide valuable insight to geologists studying groundwater flow, contaminant, transport, as well as petroleum engineers dealing with complicated geologic conditions in reservoir rocks at greater depths.

To support the production of oil and gas, Morocco operated multiple survey operations, between 2000 and 2011, where more than 20000km2 of 3D and more than 70000km of 2D

seismic data was acquired offshore, and more than 1300km2 of 3D and more than 7000km of 2D seismic data acquired onshore.

It is interesting to mention that GPR technique relies on the principle of propagation of electromagnetic waves, while seismic tomography relies on the propagation of seismic waves. However, there are no significant differences in treatment and interpretation of the data.

GPR technique is a very attractive technology which supports a wide range of applications, depending on the used frequency band. It provides accurate topographic maps which are important for both civilian and military applications.

GPR is widely deployed for cavities and voids detections. In fact, the detection of cavities is important to avoid potential collapses and extend the safety and operational lifetime of constructions or industrial facilities, hence, preventing disasters and human casualties.

Another important application of GPR, in the military domain, is landmine detection. According to official figures released by the Moroccan authorities, the victims of landmines detonation in Morocco amount to 2536 for the period ranging from 1975 to 2012. From this number, 831 passed away after the detonation of the explosive devices, while the remaining 1705 survivors suffer from permanent disabilities, in addition to the already serious social, psychological, and economical implications.

In fact, one of the worst problems that humanity faces is the buried landmines and unexploded ordnance. Millions of landmines have already been scattered over many post-war countries. Complete clearance of a minefield is required in order to restore public confidence. Therefore, mine-detection techniques require extremely high detection rates and precision.

2. Problem Description

GPR data inversion is not an easy problem to solve. For example, for resistivity electrical tomography, a commercialized software (*res2dmod* & *res3mod* [2]), realizes the inversion process of the data, which then gives a global vision of the profile model.

However, for GPR inversion, other facts appear like electromagnetic dispersion, attenuation, and clutter signal created by little diffracting objects in the subsurface layers which could cause a misinterpretation of the results. In reality, the challenging step when interpreting GPR data is the process of localization and parameter reconstruction of hyperbola signatures with high accuracy.

Hough transform [3], a pattern recognition tool used for GPR data inversion, relying on a voting process over an accumulator entries that represent the possible parameters of the shape (further explanations are given in chapter 3).

Traditionally, the Hough transform is used to reconstruct the hyperbolas caused by targets reflections, but its implementation is difficult because we are talking of three-dimensional parametric space. In fact, for hyperbolic arcs, a 3D array is required which increases the computing time dramatically. However, the discrete nature of the voting process causes peak generation problems in the Hough transform space. It might split one peak into multiple peaks closely located. It also spreads the peak to multiple cells around the real one, blurring the real positions of the peaks and, hence, limiting the efficiency of the Hough transform especially with the noise present within the image space. A special case is when the resolution (i.e. the size of the accumulator) is set so high that the peak will be too flat to be accurately detected. Additionally, the trade-off between the execution time for the curve identification and spatial resolution are exponentially related.

3. Objectives

This PhD aimed to construct a prototype of an automatic support system for GPR data inversion by:

1. Providing a complete tool for pre-processing GPR images, to facilitate the inversion procedure since, in very cluttered environments, the corrections done on the field are not sufficient.
2. Experimentation of several signal processing tools for GPR data inversion (Hough transform, time frequency approaches, please see Section …).
3. Adaptation of the classical Hough transform [3] to be adequate to use for large and noisy images.
4. Using the available multi-objective evolutionary methodology for designing RBF neural classifiers to discriminate between regions containing hyperbolas and regions that are part of the background of the radargrams, before Hough transform application. The referred system allows, besides designing a less complex classifier topology and determining its parameters, to perform feature selection in two stages (genetic algorithm technique [4, 5] and mutual information approach [6]), according to different objectives and priorities.

In contrast with the existing approaches found in the literature, where the classifier topology is very complex, our system performs well with a minimum set of features and a less number of neurons in the hidden layer.

5. Performing a validation of the classification system and a comparison with the developed 1D Hough schema and the contributions of other researchers.

4. Major Contributions

A less complex algorithm '1D Hough' which overcomes the major drawbacks presented by the classical Hough transform was proposed.

Regarding the strategy adopted by researchers in the field, designing a high performance classifier to narrow down the position of hyperbolas to certain areas before applying Hough transform. The designed classifier can be incorporated in a platform to be used in onsite survey operations in the field since it has a very low complexity and provides its outputs almost immediately.

A thorough review has been done on the classifiers used in other works, as well as feature extraction techniques (please refer to section 2.5 and section 2.4). The used features in this study are HOS cumulants. These features are different from moments but they rely on central moment calculations. To our knowledge, none of the classifiers proposed in the literature has employed these features, which:

a) Are robust against Gaussian noise.

b) Have the ability to capture some affine details on the samples.

c) Could give higher detection accuracy with a small number of features.

A completed bibliographic study was realized about feature selection (please refer to section 2.5.3). None of the proposed approaches had combined a filter and a wrapper method. This was done in this work take advantage of the higher capabilities of the wrapper methods while avoiding the problem of the higher computational time requirement.

Several experiments were conducted in MOGA, with two databases (Maas database and Dauphin Island database), and the corresponding obtained models were evaluated using a validation set (that usually corresponds to 20% of the database). To construct the dataset of the conducted experiments, AproxHull [7, 8] is used to incorporate convex points in the training set. This allowed MOGA to use the whole range of the data where the classifier is going to be used.

The best results are obtained from an ensemble of preferable models of the experiment whose training set contained all convex points with some random sample to complete the set. Values of validation accuracy of 88.99% with a false negative rate over validation set (FNv) of 20.69% were obtained with Maas dataset (validation set of 628 samples), while a validation accuracy of 95.26% with an FNv rate of 9.33% was gathered with a Dauphin Island dataset (validation set of 592 samples).

Comparing the classification results with Support Vector Machines (SVMs) and Convolutional Neural Networks (CNNs), it was revealed that, despite the huge complexity of SVM and CNN models, the accuracy of the ensemble of preferable models is superior to that of the Support vector machines and convolutional neural networks models.

The Maas database results were compared with the Viola Jones classifier [9] proposed by [10] (the same data is used), achieving less accuracy and higher FN_v than the selected MOGA model. In general, the present approach compares favourably with other similar (although with not the same data and not the same specifications) published approaches [10-14], achieving slightly improved or similar accuracy, albeit with a much smaller complexity model.

As a result of the research work developed under this PhD thesis, the following publications were produced:

- H. Harkat, A. Ruano, M.G. Ruano, S.D. Bennani, « GPR target detection using a neural network classifier designed by a multi-objective genetic algorithm », submitted to Applied Soft Computing.
- H. Harkat, A. Ruano, M.G. Ruano, S.D. Bennani, « Classifier Design by a Multi-Objective Genetic Algorithm Approach for GPR Automatic Target Detection », accepted to the 3rd IFAC Conference on Embedded Systems, Computational Intelligence and Telematics in Control, (CESCIT 2018).
- H. Harkat, A. Ruano, M.G. Ruano, S.D. Bennani, « GPR hyperbolas classification using a multi-objective genetic approach», 9th European Symposium on Computational Intelligence and Mathematics. Faro (Portugal), October 4th – 7th, 2017.
- H. Harkat, Y. Elfakir, S. D. Bennani, G. Khaissidi and M. Mrabti, « Ground penetrating radar hyper-bola detection using Scale-Invariant Feature Transform », 2016 International Conference on Electrical and Information Technologies (ICEIT), Tangiers, Morocco, 2016, pp. 392-397. doi: 10.1109/EITech.2016.7519626
- H. Harkat, S. D. Bennani, A. Mansouri and A. Slimani, « Inversion of GPR data: 2D forms reconstruction for cavities based on born approximation, TSVD regularization and image

processing algorithms », 2016 International Conference on Information Technology for Organizations Development (IT4OD), Fez, Morocco, 2016, pp. 1-6.
doi: 10.1109/IT4OD.2016.7479317

- H. Harkat, S.D. Bennani, « Inversion des Données GPR par Algorithmes Génétiques : Cas d'une Couche Dispersive », COLLOQUE INTERNATIONAL, TELECOM'2015 & 9èmes JFMMA, Meknès.
- H. Harkat, S.D. Bennani, A. Slimani, « Time-Frequency analysis of GPR signal for cavities detection application », Mediterranean Conference on Information & Communication Technologies 2015, Saïdia. doi: 10.1007/978-3-319-30301-7_12.
- H. Harkat, S.D. Bennani, A. Slimani, « Time-Frequency Analysis of GPR Signal Using a Modified Hilbert Huang Approach », International Conference on Wireless Technologies, Embedded and In-telligent Systems 2015, Fez.
- H. Harkat, H., S. Dosse Bennani, « Ground Penetrating Radar Imaging for Buried Cavities in a Dispersive Medium: Profile Reconstruction Using a Modified Hough Transform Approach and a Time-Frequency Analysis », (2015) International Journal on Communications Antenna and Propagation (IRECAP), 5(2), pp. 78-92. doi: http://dx.doi.org/10.15866/irecap.v5i2.4978.
- H. Harkat, S.D. Bennani, A. Slimani, « Time-Frequency Analysis of GPR Signal Using a Modified Hilbert Huang Approach », Revue Méditerranéenne des Télécommunications, Vol 5, No 2 (2015).
- H. Harkat, S.D. Bennani, J. Belkadid, « A Modified Approach Based On The Hough transform For Target Localization In GPR B-Scans », The 2014 International Symposium on signal, Image, Video and Communications - ISIVC 2014, Marrakech.

5. Thesis Structure

This thesis is organized into 7 chapters. Chapter 2 gives an extensive state of the art for GPR data treatment and inversion which includes data inversion procedures, in general, clutter removal approaches, fitting techniques, and classification strategies. A review of features extraction and selection techniques is also presented in this chapter.

Chapter 3 provides a brief overview of the theoretical background that was needed to develop this work. Firstly, an overview of GPR electromagnetic theory and FDTD mathematical formulation are given. Secondly, a review of Convolutional neural networks, artificial neural

networks and learning algorithms, Support Vector Machines, Multi-Objective Genetic Algorithm, and AproxHull is done.

Chapter 4 gives a detailed explanation of the proposed 1D Hough transform, illustrated with several examples. Moreover, a novel technique for velocity correction is explained, since the velocity provided by the Hough transform is quietly different than the true velocity of the wave. The velocity calculated by the Hough transform is the mean value between the different transitions of the wave. This technique reconstructs the permittivity of the medium, given the time of propagation manually read from the time-frequency representation of A-scans.

Chapter 5 starts describing how our datasets were produced from the radargram images, with the help of geologists [10]. To select an adequate RBF neural network classifier architecture, several scenarios were conducted in MOGA. This chapter shows the results of the validation set and compares the proposed approach with support vector machines and Convolutional neural networks. Several experiments are carried out with Support vector machines with different feature sets and scenarios, to demonstrate the capability of MOGA and the efficiency of the two-stage feature selection strategy. Finally, the results are compared with existing classifiers present in the current literature.

Chapter 6 demonstrates the discrimination power and the use of the designed classifiers over the radargrams. A sliding window approach is adopted, but the parametrization of the approach is a bit difficult since it relies only on one trial. The 1D Hough schema is applied over the selected regions. Results are compared with the case of direct application of 1D Hough schema over the big radargrams, and the literature approaches that use this stage (application of Hough over specific regions).

Conclusion and future works are presented in chapter 7

2. State of Art

2.1. Introduction

In radargram profiles, hyperbola apexes represent all the information about the nature of the medium and the localization of the target. Through the literature, several approaches were deployed to reconstruct the original profiles. The majority of the approaches adopt the idea of target discrimination before applying a fitting technique to build signatures skeleton from edge points.

In fact, it is known that various researchers have developed new GPR signal and image classification algorithms in pursuit to discriminate targets/anomalies from the background or false alarms. Though presented in many applications, feature engineering continues to be an active area of research in this field; it is the key of any machine learning system. Besides, several classifier architectures were proposed to classify both hyperbola signatures and 1D A-scan signals, for several civil and military applications, to design automatic systems for clustering this data and recognizing types of buried targets.

This chapter is intended to present an extensive state of the art review about the research done in this field starting from the pre-processing phase and discrimination step, and then ending by the classification and fitting methodology.

The outline of this chapter is as follows: Section 2.2 gives a general overview about the approaches employed for GPR data inversion. Section 2.3 presents the GPR signal denoising techniques. The work done in the features engineering section related to GPR is summarized in section 2.4. Finally, the different classification approaches tested with GPR data are described in section 2.5.

2.2. GPR data inversion approaches

An extensive volume of research has been undertaken to solve this problematic GPR data inversion. Some researchers had focused on qualitative reconstruction while others had taken quantitative research studies to reconstruct the exact parameters. Reconstruction of parameters includes defining the scatterer, in terms of localization, depth, and physical properties, which are examined while deploying various methodologies inspired on electromagnetics, signal

processing, and, lastly, on computational intelligence. Accordingly, the previous research can coarsely be regrouped into four derived categories: Electromagnetic field inversion techniques, fitting techniques, frequency based techniques and classification based techniques.

2.2.1. Electromagnetic field inversion techniques

Electromagnetic field inversion based techniques for GPR processing are very promising techniques. This is due to the fact that theoretical foundations are available from several bibliographic sources, whose propagation equations are facilitated by the use of computational techniques.

Recent developments in the field include novel algorithms [14, 15] used for the resolution of the electromagnetic equations of the reflected field, without taking into account the classical approximations to linearize the problem. In fact, this problem was usually formulated as an ill-posed problem with a non-linear characteristic [16] which could only be applied for weak scatterers, in order to have a valid physical solution [17]. These assumptions are known as Born and Rytov approximations for penetrable targets and by Kirchhoff approximation for non-penetrable ones. Several linear algorithms were proposed using these approximations [18-20], paired with a regularization schema [18-21] that presented some constraint on the solution based on prior knowledge of the field or on iterative procedures.

Actually, the linear methods ignore some scattering effects and assume a qualitative reconstruction. However, non-linear methods take into account these effects and are generally based on iterative schemes. Hence, they minimize the difference between the measured and modelled reflected fields in order to perform a quantitative reconstruction of the objects. These methods employ iterative schemes to obtain the unknown parameter, as exemplified by the Born iterative method [22], the modified gradient method [23], and the global optimization approaches which rely on genetic and evolutionary algorithms [24].

Another category of methods is based on the inversion of the electromagnetic fields and on Migration algorithms [25]. They are inspired by seismic prospecting which exploits either the time or space reversibility in order to relocate the events that are recorded in the interface located at the subsurface level. Hence, the resolution of the images increases and the area of the objects is focused. Examples of tested migration algorithms for GPR sensing are the diffraction summations [26], the Kirchhoff wave-equation migrations [27], the phase-shift migrations [28], and the frequency-wave number (F-K) migrations [29]. Furthermore, for lossy media, since the

Fourier Transform cannot be used, the inverse filtering process is just reformulated as a linear inverse problem [30].

It is observed that electromagnetic field based inversions of GPR signals are not broadly exploited in fielded or real-time data analysis systems for reasons which oppose to the goals of these designed software.

This type of inversion models depends on computationally expensive iterative estimation algorithms when scatterers are small. If in addition large-scale estimates are envisaged, the referred solutions will preclude real-time applications.

Furthermore, previous knowledge of the incident field is inherent, and for surface GPR sensing, far-field theoretical expressions broadly vary from the measurements of the radiation patterns. Besides, antenna configurations and the parametrization of emitted wavelets rely on medium physical constants. Therefore, the estimated results of the inversion techniques will be completely unsuitable for complex cases.

2.2.2. Fitting techniques

Fitting techniques aim to fit a series of data points extracted from edges of the profiles that are possibly subject to constraints. A hyperbola is then defined which consequently reconstructs the skeleton of these curves. The research done at this level concentrates on two main algorithms: the Hough Transform [31] and its derivatives, and the polynomial fitting techniques.

Hough Transform, that relies on a voting process over the possible locations (i.e. points belonging to hyperbolas) of the hyperbolas, had been tested by several authors for different types of GPR data [32, 33], but it presents many drawbacks. In fact, for hyperbolic arcs, a 3D array is needed, increasing dramatically the computing time. Moreover, the discrete nature of the voting process causes several doubts regarding the peak (i.e. the maxima of Hough accumulator) generation step of the Hough algorithm. It might generate several derivatives of the same peak, lying in the same level close to each other. It also spreads the peak into several cells around its true position, disabling peak distinction. Therefore, it limits the accuracy of the Hough transform especially when the image space contains disturbances and noise. An extreme case is that when the resolution (i.e. the size of the accumulator) is set so high that the peak will be too flat to be accurately detected [33]. Besides, the trade-off between the execution time for the curve identification and spatial resolution are exponentially related.

Several derivatives of the Hough transform algorithm have been proposed in pursuit to overcome these issues either by using an improved version of Hough, such as the randomized Hough transform [34, 35] which revealed significant results, or, by pre-processing the profiles before applying the original generalized Hough transform [36]. Others propose to work only with a 1D Hough algorithm. Wang and Su [37] use a correlation between adjacent A-scans to extract hyperbola curves. Two parameters were derived by a fitting strategy while the third parameter was given by the 1D Hough transform. Presented from this perspective, an algorithm using a 1D Hough transform visibly reduces the computational problem and increases the probability of detection [38].

Concerning the polynomial fitting techniques, many conventional algorithms were tested for GPR data as slightly improved versions of least square [39-42]. Otherwise, Chen and Cohn [43] proposed a probabilistic mixture model, using an improved version of the orthogonal fitting algorithm proposed in [44], to detect multi-hyperbola and to ensure the robustness against noise. However, the iterative process of such orthogonal fitting algorithm makes it highly computationally consuming.

In general, these latter techniques are considered faster than the Hough Transform and had attained good achievements for highly sensitive applications, whereas the produced performance tend to decrease due to the lack of data. The trivial way to avoid this technical issue was by deploying iterative refinement algorithms which in turn present a computational cost. This persistently apparent dilemma must be solved with an appropriate combination of accuracy, computational time, and a reasonable amount of data.

2.2.3. Frequency based techniques

Frequency based techniques tend to estimate the parameters of the targets based on spectral estimation techniques. In fact, GPR signals time distributions are usually not very representative in the presence of interferences with closely spaced and multiple events. Inspired by the seismic field, there are two kinds of techniques: Fourier based, and high-resolution signal processing techniques. The aim of these techniques is to extract properties based on an analysis of the phase and amplitude representation of the 1D signals. The first group of techniques is characterized by high precision while the second group is characterized by their high resolution.

Researchers had applied Fourier based techniques, such as the fast Fourier method and the stationary Fourier transform. Others tend to apply time-frequency techniques, like the short

time Fourier transform (STFT), the S-transform [45, 46], and smooth-pseudo-Wigner-Ville [47, 48], usually coupled with an optimization algorithm, like the stochastic hill-climbing algorithm in a modified version [45]. This type of algorithm aims to minimize the difference between the analytic solution and the measured one, via a well-defined cost function, until convergence to an optimal solution of the problem is achieved.

The second group of techniques englobes the super-resolution spectral estimation techniques, namely Multiple Signal Classification (MUSIC) [49], Estimation of Signal Parameters by Rotational Invariance Techniques (ESPRIT) [50], Polynomial Versions of MUSIC (root-MUSIC) [51] and Min-Norm (root-min-norm) Algorithms [52].

A third supplementary hybrid category of techniques could be added to frequency based techniques, which combines frequency based techniques and high resolution ones. An example of these approaches was introduced by Shrestha and Arai [53] that combines fast Fourier transform (FFT) and the MUSIC algorithm revealing very interesting results in comparison with simple applications if MUSIC or FFT algorithms by their owns.

2.2.4. Machine learning based techniques

Machine learning based approaches deploy two types of algorithms. Unsupervised clustering techniques and classification approaches are used to localize hyperbola regions to further apply a fitting algorithm, or to directly classify the material types starting from a 1D signal.

Clustering is an important approach to data mining since it separates the population under study into groups with similar parameters. Several clustering algorithms have been tested based on the perspectives of K-means [54], fuzzy k−Nearest Neighbours (K-NN) [55], fuzzy interference [56], and Hidden Markov models (HMM) [13].

Nonetheless, there are some unsupervised clustering approaches proposed recently as direct alternatives to fitting techniques; the idea is to cluster the pixels to orderly reconstruct the structures using an internal iterative fitting strategy. A hyperbola fitting and a rearrangement of the clusters is done alternately until convergence of the process to optimal solution is achieved. Based on relevant work in this area, some high rated algorithms could be cited: GameRec clustering method relying on geometrical background knowledge [57], PLAD iterative geometric clustering [58], and a SPABC simultaneous perturbation artificial bee colony algorithm [59].

The first two referred techniques are robust against noise and they are able to find clusters of hyperbola shapes without a beforehand knowledge of occurring cluster numbers.

The later technique deploys an evolutionary algorithm, "artificial bee colony" and had drawn interesting results as well. Jafrasteh and Fathianpour [59] first designate the candidate hyperbola by using the approach proposed by Capineri, Grande and Temple [33]. The set is then given to a bee colony algorithm. The individuals (in the case hyperbolas) with higher fitness value are chosen as a solution after an iterative process previously established by the inherent reproducing techniques used by any genetic based algorithm.

Regarding classification approaches, they could be segregated into two main categories: template matching approaches and features based. The first categories aim to conduct classifications making use of algorithms to match a specific template while the features based techniques rely on a features extracting step before the classification step is undertaken.

The template matching approaches usually involve a fitting technique deploying variant similarity criterions, to be mentioned the 2-D spatial convolution, normalized cross correlation [60], and a L1-norm [61].

Concerning the features based techniques, a variety of features have been used to correctly classify GPR data. The studies executed in this scope are detailed in section 1.3.

Concerning the classifier, several architecture types have been experimented. The different approaches may be regrouped as follows: Bayesian models, function based, and trees based models.

The first group reveals the approaches with Hidden Markov models [62] and naïve Bayesian classifier [63]. The second group categorizes approaches with Multi-layer perceptron [12] and SVM [64]. The last group describes classification approaches with decision tree [65, 66] and the cascade classifier namely Viola Jones [9].

The main objective is to compare the works carried out in this thesis with already existent approaches. Therefore, section 2.4 will present further descriptions about some approaches that draw higher performance for GPR data classification.

2.3. GPR Signal Denoising

2.3.1. Overview

GPR is a high resolution technique, but it suffers from the presence of noise from different sources including from the electronic equipment during the acquisition phase which, in particular, causes a coherent noise interfering with the informative signal in radargrams.

Another source of noise, non-shielded antennae noise, causes a surface-scattering effect, thus generating clutter noise. In fact, reflections from above-surface and low-depth objects, such as trees, power lines, buildings, and electrical and water underground installations tend to obscure real subsurface reflections. Due to the characteristic of high air velocity, the arrival time of such noisy reflections could intersect within the time window of the interest signal. It is an inherent problem at low frequency when a shielding operation is difficult.

Moreover, there is a type of coherent noise, ringing noise, which occurs as horizontal and periodic events. When it is strong and not properly eliminated, some deeper signatures may be completely masked.

Some basic corrections are normally applied before trying to process radargrams for noise removal:

- **Refocusing traces:** This first step consists on removing the continuous component from each trace. There are several ways to do it:

  Average amplitude of the trace is calculated and then subtracted at each sampled amplitude. Average amplitude is subtracted within a sliding window of a defined size, so it has to be greater than the temporal size of the transmitted signal. It is also sufficiently low in order to take into account any change in the average value over the acquisition time of the trace.

  A high-pass frequency filter is performed to suppress the low frequencies.

- **Spectral filtering:** It is a recommended step, since:
  - The time pulse radar emits a wave with a wide frequency spectrum. The data is affected by noise from different sources, resulting from a lack of calibration of the electronic device. Thus, each trace is centred on a 'DC value', and it is also possible to use a f_k filter as well.
  - Each trace is affected by periodic electronic noise of low amplitude sometimes masking the sought signals if the latter are of small amplitude. This noise produces a masking effect for the most recent reflections captured or in lossy environments. In this case, it is recommended to carry out frequency band-pass filtering in the range between f/3 and 2f where f is the centre frequency of the transmitted signal. Moreover, it is necessary to filter electronic noise as well.

- **Signal amplification:** Since the attenuation of the radar signal is very important, it is often difficult to see other things than the direct wave in the raw data. Often, only the visual correlation trace to trace allows to localize a hyperbola. To allow this visual interpretation, one can play on the palette of colours, or on the value at which the maximum signal is truncated. This truncation decreases the resolution of events near to the surface. Being said, the best solution is to play with the amplification of the signal by multiplying each trace by a gain function F_g. There are several ways to apply a gain:
  - The user can choose a function F_g sum of a linear term and an exponential term in function of the travel time t.
  - Another way is to apply a gain with automatic control: a constant is determined to amplify the amplitude inside of a window, user defined, as a function of its maximum amplitude (defined by its quadratic mean) and of the top window. A post-gain analysis of a radar can only be done on the arrival times of interesting reflections.

However, either radargrams or A-scans require an additional pre-processing step after these corrections are made in order to remove different types of noise resulting from multiple sources. In addition, this step could have a direct effect on the quality of the reconstruction procedure.

2.3.2. General signal denoising methodologies

The denoising approaches tested in the literature could be mainly classified into three groups: Wavelet based techniques, statistical relying algorithms, and empirical mode decomposition based approaches.

The wavelet based technique make use of multi-resolution analysis tools to decorrelate an useful signal and noise. It has been introduced to GPR data processing recently while most of the related studies are trying to establish efficient noise suppression algorithms in a time-frequency mode. Recent research results divulged in this area include discrete wavelet transform [67], Gabor transform [68], Curvelet transform [69], and Shearlet transform [70]. The Curvelet and Shearlet methods are designed with the aim of combining scale with orientation information in the same representation. They are multiscale and multidirectional expansions that yield to optimal sparse representation, which describe the data with superposition of a small number of components while allowing a high-resolution reconstruction.

Besides, various thresholding techniques have been tested with GPR data, such as hard, soft, firm shrinkage, and non-negative garrote thresholds [71].

For the second group of techniques, denoising is carried out through an empirical mode decomposition (EMD) algorithm, designed as a time-domain detrending tool. It decomposes a signal into several sub-signals, usually known as intrinsic mode functions (IMF), of varying frequency content. IMFs have two distinct properties: the total number of extrema (minima and maxima) must not be different from the total number of zeros crossings by more than one, and it must not have a non-zero mean.

Several studies carried out, relying on this approach to denoise profiles, Battista, Addison and Knapp [72] and Chen and Jeng [73], try to take advantage of some improved versions of the original EMD algorithm, EEMD, and CEEMD [74]. On the other hand, other studies try another interesting approach called BPD-EMD. This approach combines a Basis Pursuit Denoising technique [75] with EMD. The results prove that due to computational costs the BPD-EMD should not be adopted with high dimensionality and very noisy signals.

2.3.3. Clutter removing approaches

Clutter removing techniques are organized among others as two principal clusters: statistical signal processing based approaches and classical filtering procedures.

Starting from the assumption that clutter signal is much stronger than a target component, present in the GPR profiles, the former approaches deploy a linear transformation to model this problem. In this perspective, several statistical signal processing techniques were deployed, such as principal component (PCA), independent component analysis (ICA) [76, 77], morphological component analysis (MCA) [78], method of independent factor analysis (FA) [79], and SVD [80]. All of these techniques have their own advantages and drawbacks in the image processing field. For example, ICA and PCA have multiple advantages, and when compared with each other, the ICA shows significant improvement to respect to the PCA and SVD decomposition technique.

Furthermore, several filtering techniques are proposed through the literature. Some conventional filters are tested as Kalmen filter [81] and as a symmetrical filtering approach [82]. Other approaches [14, 15, 83], take advantages of filtering properties in time or frequency domains. The main method within these works is a classical algorithm performing a process in a time domain with a digital filter with non-optimized coefficients. Another technique consists

of a windowing signal in the time domain, with an appropriate choice of the window size, to further filter the spectrum as a convolution calculation taken place in the frequency domain. Nevertheless, the last approach has a major drawback which is the difficulty in choosing the window size in order to avoid the elimination of target signatures buried at shallow depths. Besides, since the clutter is non-stationary, the size of windows must be variables depending on the A-scan traces analysed. Potin, Duflos and Vanheeghe [84] had proposed to work with digital filters that are specially adapted to B-scan, trying to correct the issues addressed before.

Additional interesting approaches [85, 86] could be classified under classical filtering procedures such as polynomial phase and Autoregressive–Moving-Average (ARMA) model based deconvolution. Kempen and Sahli [86] had chosen to treat the clutter as a parametric ARMA model combined with a Linear Predictive Coding (LPC) method for the purpose of estimating the parameters of the model while in pursuit to filter the data at the same time.

2.4. GPR Features Extraction Techniques

Engineering features are the key to any machine learning technique. It enhances signal characteristics for object detection and classification, even though image features must give sufficient amounts of information about the targets while avoiding irrelevant and redundant attributes. Besides, the features should provide invariance characteristics made for changes in illumination, rotation, scale and/or translation. Moreover, the extraction algorithm should be as less complex as possible, so that the approach becomes feasible for a large amount of data.

To achieve these goals, many conventional feature types have been tested with the aim of giving an accurate representation of GPR hyperbola patches. Therefore, researchers had the vision to apply feature extraction methods. They are robust in dealing with many of aforementioned challenges without needing prior information about targets. In addition, they are fast enough to be applied in real-time applications that need high precision such as mine discrimination.

The work done in this field could be sub-classified into two basic main categories: model and non-model based techniques.

For the model-based approaches, a priori shape model is designed to match image sequences in the database to this predefined model. Once a match occurs, the corresponding features are extracted. These techniques have the ability to extract detailed and accurate sample descriptions as well as having the capability to model the clutter well.

However, for the non-model based method, feature correspondence between samples is based on prediction, shape, texture, and colour.

The model-based class regroups a minority of approaches [87, 88], using an autoregressive model (AR) to distinguish between different target types and clutter.

An autoregressive model could be defined as a random process modelling approach in which the current value of the output is expressed as the sum of: its mean value, a current value of a white noise process, and a linear aggregate of the previous output values. The number of output values used is designed as the model order (n), while it has (n+2) parameters: n coefficients plus the mean and the variance of the white noise.

Deiana and Anitori [87] had the objective to experiment the sensitivity of the AR model in order to distinguish the behaviour of the AR model coefficients with different types of mine and soil characteristics while at different depths. The comparison of two patterns is performed using the Itakura-Saito distortion measurement [89]. Nabelek and Ho [88] has taken the previous work as a reference to improve the proposed AR model. They proved that applying the AR model, in the down-track direction, moves the vehicle-mounted radar during the scan and, thus, it is better than a fitting model along depth for a single A-scan. When using this strategy, they found that the mean-square fitting error along the scan reveals some microscopic pattern features that can be used for the classification of those target types. These patterns give a large fitting error between adjacent A-scans, at the edge boundaries of the targets, and a relatively small error between A-scans over the targets. The resulting model fitting residual square errors are used to form a feature vector for classification by using a K-NN classifier.

The third approach [90] explores the application of a more practical and statistical time-frequency model for A-scans. The proposed model uses an HMM model with AR state densities where each state models a number of spectral and energy characteristics and the underlying HMM structure flexibly describes the time varying nature of these characteristics.

The Bayesian formulation of the model takes advantage of non-parametric techniques relying on a Dirichlet process [91], stick-breaking priors [92-94] and uncertain-order autoregressive modelling. The aim is to develop a model able to handle a relatively complex spectral data and to perform an automatic order selection of the model in each stage of the designed structure while making very limited assumptions. In fact, to range all the spectral properties across the collected A-scans in a unique feature vector, the non-parametric Bayesian time-series model is used within a larger statistical model. This is carried out by developing a Dirichlet process mixture of HMMs with uncertain-order autoregressive state densities. A

clustering of the A-scans is performed according to the deduced model parameters. This model is then used to perform the classification of A-scans compared to the known landmine types.

For the non-model based approaches, they can be broadly classified as four approaches: gradient based feature, binary based features, pixel level features, and frequency feature.

2.4.1. Gradient feature

State-of-the-art gradient feature extraction methods regroup the approach deploying: Edge Histogram Detector (EHD), Histogram of Oriented Gradient (HOG), Scale Invariant features (SIFT), and (SURF). These features had been chosen mainly due to their easy computational implementation, robustness against moderate occlusions, and for being relatively insensitive to viewpoint changes.

Torrione, Morton, Sakaguchi and Collins [95] and Torrione, Morton, Sakaguchi and Collins [66] had proposed to apply the conventional HOG algorithm and followed by a comparison of the obtained results with an EHD technique. However, Reichman, Collins and Malof [96] proposed an adapted schema of the original HOG pre-screener [66] that makes it more suitable for processing GPR data, especially since the original algorithm was developed for human and pedestrian detection.

The major formal improvements are summarized in figure 2.1.

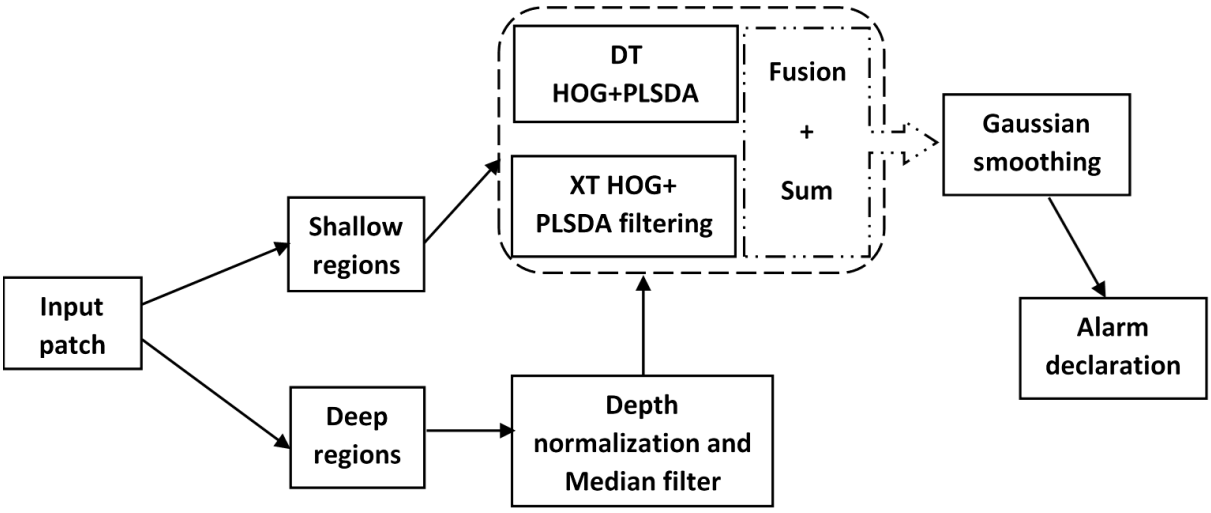


Figure 2. 1: Adapted HOG schema. PLSDA is the linear partial least squares discriminant analysis classifier used for pre-screening [97]

They took into consideration the extension of the original HOG, as executed in [97], to identify deeply buried targets more precisely. In fact, a signal collected by a GPR, for deeply buried targets, is considerably attenuated in comparison to the one measured from shallowly buried targets. Results reveal the presence of low energy regions which affect the calculation of gradients in those patches and could affect the features extraction step.

In GPR target detection, high contrast is an important cue for detecting the presence of targets. Due to these facts, they had suggested modifying the histogram normalization step which controls the suppression of image contrast. The normalization of standard HOG (figure 2.2) tends to obscure the position of targets in a region where the majority of the neighbouring cells exhibit high contrast values. As so, they had proposed to normalize the features utilizing only a row of three contiguous temporal cells (i.e. 1x3 instead of 3x3 neighbouring) sourced from the side on which data had already been collected.

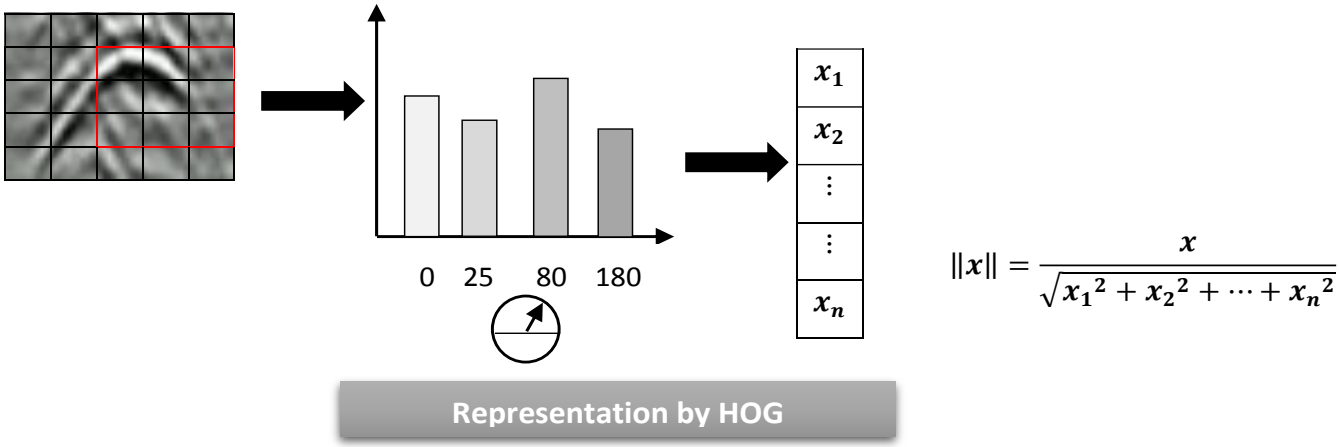


Figure 2. 2: Original HOG normalization process [97]

Moreover, about the EHD technique, despite the original algorithm tested by Frigui and Gader [98] and Frigui and Gader [99], an adaptive approach was proposed by Frigui, Fadeev, Karem and Gader [100], to correct some issues in order to make the descriptor more optimal for use with GPR data.

Frigui, Fadeev, Karem and Gader [100] had proposed to use adaptive threshold values based on the response given by edge filters at background areas for windows at different depths, as well as for different orientations, in order to decide if the edge in a region is strong enough.

2.4.2. Binary features

This category regroups two algorithms: Local Binary Pattern (LBP) and Binary Robust Independent Elementary Features (BRIEF) extraction techniques.

The original LBP algorithm was tested by Harris, Alvey, Ho and Zare [101]. GPR B-scans are first split into several spatial and depth regions. LBP features are then computed in order to capture the edges of each target in the leftmost spatial region and the falling edges in the rightmost spatial region. In this context, the region in the centre of the patches was used in this context, to capture the entire pattern for limited targets. Nonetheless, depth regions were considered to diminish the effect of noise.

Otherwise, the original BRIEF algorithm was discussed by Sakaguchi [102], who states that BRIEF features perform worse as an object descriptor, but it performed well for instant matching in comparison with SIFT and HOG descriptors. This is due to the fact that the BRIEF descriptor relies on a binary comparison of neighbouring pixels, and it has a limited amount of information about absolute energy in the image which represents the higher values in target regions.

2.4.3. Pixel level features

This category regroups three basic techniques: Principal Component Analysis (PCA) based, variance and covariance features, grey level co-occurrence matrix (GLCM) based features, texture features coding method (TFCM), and moments.

PCA [103], normally used for dimensionality reduction, is deployed by Moysey, Knight and Jol [12] and Park, Kim and Ko [104] as a features extraction technique within a geological context, to describe the GPR patches.

It consists of projecting radar patches into a set of basis images in which each one models a reflection pattern in the image.

The basis images are obtained by performing a PCA over the patches extracted from random locations of the big images. In fact, patterns are gathered in the eigenvalues and eigenvectors of the covariance matrix. Pairs of eigenvectors and eigenvalues could appropriately represent every pattern of a data point in the eigenvector direction.

PCA was used in the context of A-scan signals classification by Rodriguez, Pantoja, Travassos, Vieira and Saldanha [105].

Regarding the variance and covariance features, Moysey, Knight and Jol [12] deployed these features that provide acceptable results within the used database with the aim of describing textural properties of GPR samples.

Variance describes the magnitude of amplitude variations, and it does not retain information about the spatial arrangement of values in the samples.

Nevertheless, covariance, used as a measure of spatial correlation, is independent of the samples absolute location, being the reasoning for this technique to be useful for GPR data discrimination.

Another deployed approach found in literature is the GLCM matrix based features [106, 107].

A co-occurrence matrix is a 2D array representing relative appearance frequencies of quantity pairs that are grey level values of the image pixels both disassociated by a distance and lying in a given direction (figure 2.3).

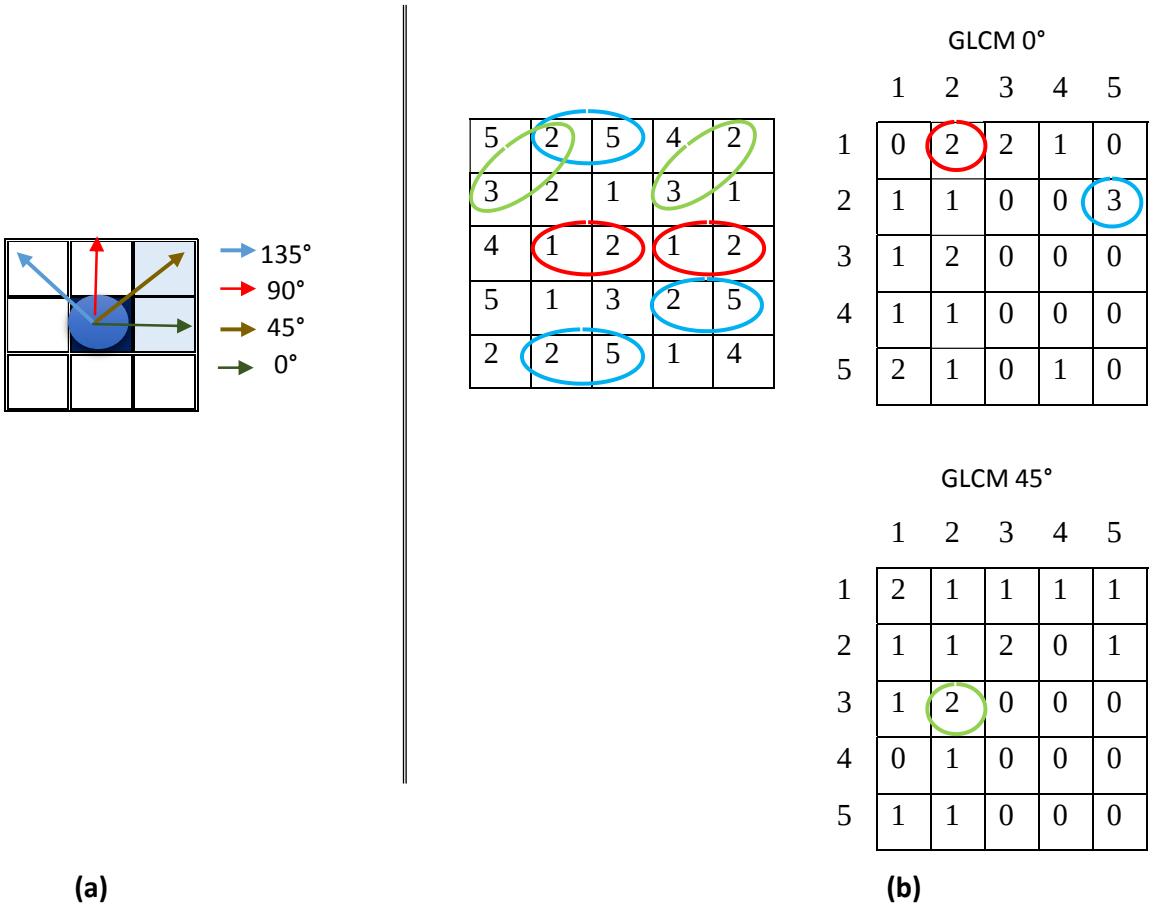


Figure 2. 3: (a) Possible directions to calculate the GLCM, (b) GLCM matrix calculation

Haralick, Shanmugam and Dinstein [108] had defined different textural feature measures from GLCMs matrix. Contrast, energy, entropy, correlation, dissimilarity, variance, cluster Shade, and homogeneity were the most common attributes used in the geophysical interpretation. Zhao, Forte and Pipan [107] state that the number of grey levels has a visible effect on the resolution and should be adequately chosen to optimize the given results.

A few papers in the literature describe the TFCM technique [109, 110] for GPR data feature extraction.

The TFCM [111] transforms an input image into a texture feature number (TFN) image.

The algorithm starts by calculating the differences among horizontal, vertical, and diagonal neighbouring directions. An example is illustrated in figure 2.4.

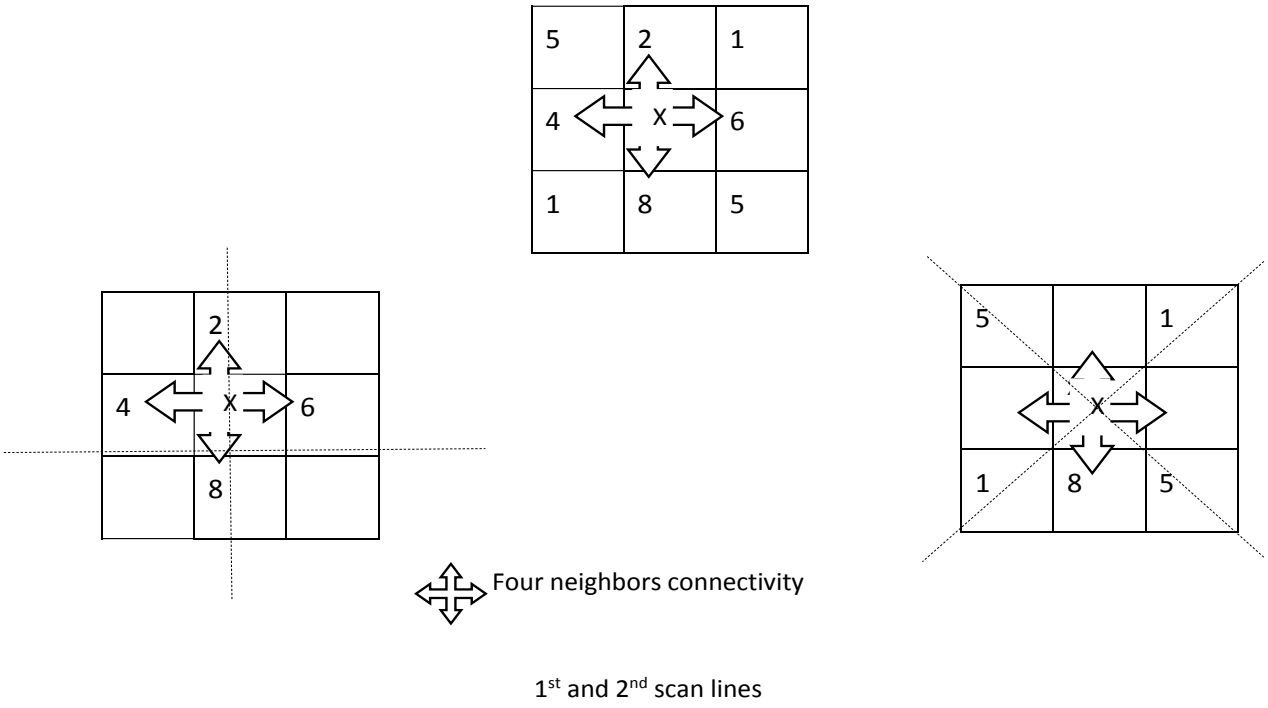


Figure 2. 4: Horizontal, vertical, and diagonal connectivity sets of a TFCM matrix

A thresholding step is then established for the two-element difference vectors output of the previous step. The TFCM maps the vectors to a grey-level class, depending on the amount of variation in each vector. The mapping is employed to obtain a 2-D texture feature number images. After constructing the co-occurrence matrices of texture feature number images, a feature vector is calculated.

Furthermore, Gilmore, Peters, LoVetri and McNeill [112] and Klęsk, Kapruziak and Olech [113] deployed moments as features to describe the GPR patterns.

Gilmore, Peters, LoVetri and McNeill [112] had taken advantage of the invariance to scale, translation and rotation proprieties of Hu's moments [114], and to classify hyperbolas in B-scan images. Otherwise, Klęsk, Kapruziak and Olech [113] propose an extension of the formulation of a 2D statistical central moment to the 3D space to be then applied to C-scan profiles. The given results were compared with the one given by HOG applied to B-scan images. It was obviously clear that the 3D moment outperforms the HOG approach. Nevertheless, the 3D moment features are highly computationally demanding.

Another approach [11] which should be included in this context, describes the decomposition of images based on a wavelet transform before extracting central moments.

However, the moment features present some disadvantages for this application. In general, they are not suitable to recognize noisy and obscured objects, and in some cases, they become completely irrelevant. This problem has been outlined since our contribution for a new feature extraction technique that wasn`t already tested in the field and is based on moments. However, high order statistics (HOS) cumulants features are quite different than moments.

Despite the fact that the moments features have a lower computational time and complexity than cumulants, high order cumulant features are more suitable for certain applications due to some mathematical properties:

- Cumulants are additive, and following this property for large statistical quantities, they are not significantly affected by Gaussian noise which does not happen with statistical moments.
- Higher order cumulants (higher than two) are semi invariant.

2.4.4. Frequency feature

The frequency based features are regrouped into three main categories: wavelet based approaches, Fourier based techniques, and time-frequency based approaches.

The wavelet based approaches involve interesting studies deploying the basic wavelet packet decomposition [115], Haar-like features integrated in the Viola Jones framework [9] for B-scan image characterization, and discrete cosines transform (DCT) [64, 116] usually applied to A-scan signals.

Nevertheless, Lu, Pu and Liu [117] proposed a more complicated schema (figure 2.5) that first uses a discrete wavelet transform (DWT) to extract approximation coefficients. Fractional Fourier transform (FRFT) is then applied to convert given coefficients by DWT into a fractional

domain. This is done to construct the feature vector that was extracted from the resulting curve envelope.

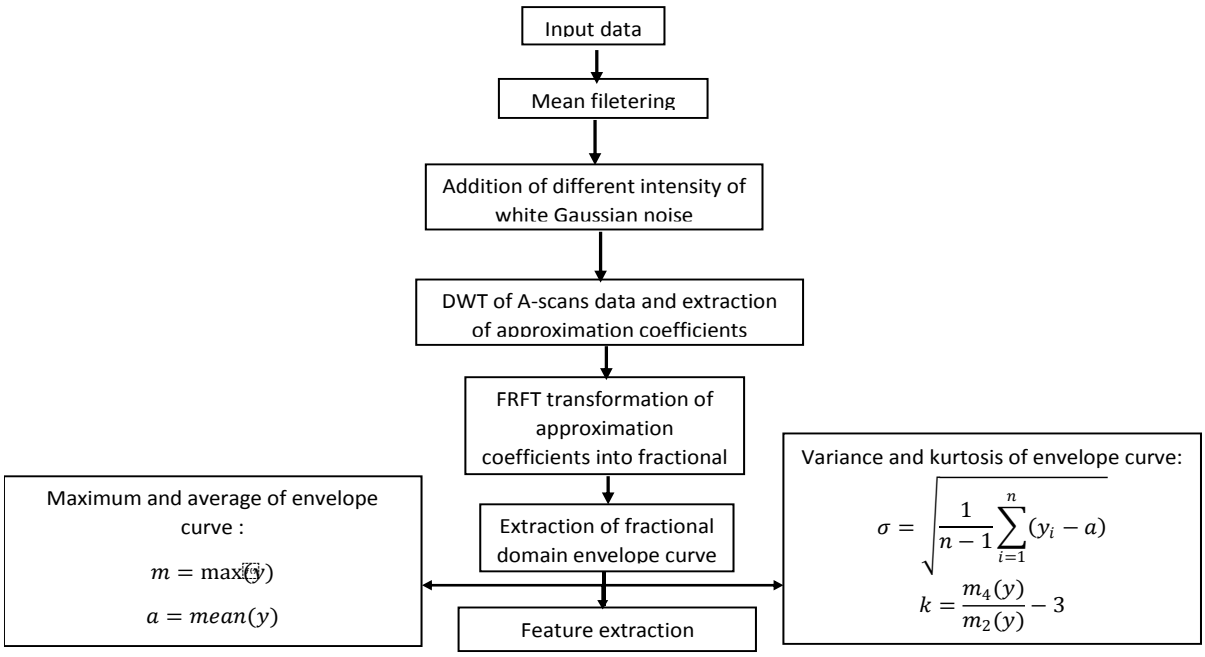


Figure 2. 5: The features extraction schema proposed by Lu, Pu and Liu [117]

Furthermore, the Fourier based approaches englobe approaches for A-scan information extraction that uses the discrete Fourier transform [118], the Periodogram [119, 120] and the Bispectrum [121, 122] based features.

Shao, Bouzerdoum, Phung, Su, Indraratna and Rujikiatkamjorn [118] had derived the amplitudes of the local maximum peaks within a specific frequency range which are arranged as a feature vector in the order of peak points for each trace.

Al-Nuaimy, Huang, Nakhkash, Fang, Nguyen and Eriksen [119] estimate the Welch power spectral based on periodogram representation. In fact, an important portion of information from the phase of the radar signal is lost when this representation is used.

Motivated by this reality, Balan and Azimi-Sadjadi [121] and Strange, Ralston and Chandran [122] extracted the magnitude and phase features based on the third order spectrum, namely the Bispectrum. The purpose is the classification of A-scans [122] and B-scans [121] signals, and in this case, the phase information is retained. The features extracted using the Bispectrum draw significant results in comparison with the ones of the Periodogram.

Another hybrid approach, based on both wavelet and Fourier transform, Mel-frequency cepstral coefficients (MFCCs) [123], mainly used for speech recognition, are commonly extracted through cepstral analysis.

The input image is first decomposed into multiple frames, and then a windowing procedure is accomplished to compute the Fourier transform (FT) thereafter. The magnitude of the FT spectrum is warped by the Mel-scale analysis [124]. Afterwards, a DCT is applied to the logarithmic of this spectrum. Fig. 2.6 resumes the operations of extraction of MFCCs from a GPR patch.

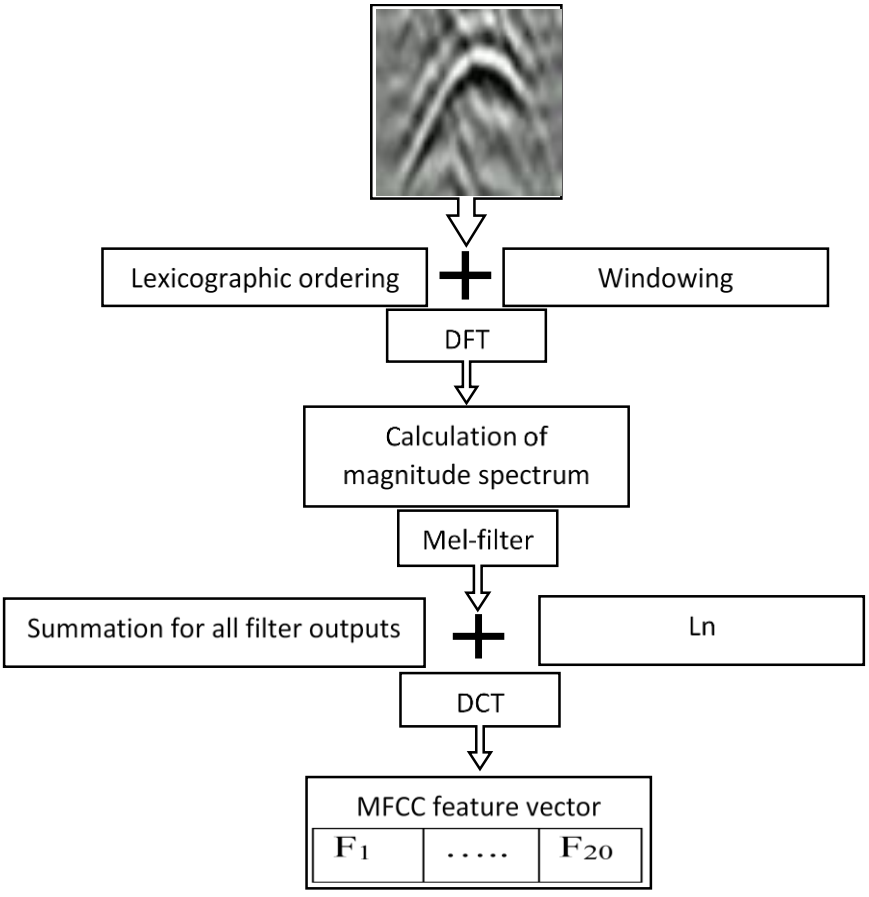


Figure 2. 6 : The features extraction schema of MFCC proposed by Khan, Al-Nuaimy and Abd El-Samie [123]

However, these features are highly sensitive to data time shifting. As so, polynomial coefficients [125] are joined to these features to reduce a shifting effect between the training and the testing portion of the data. Polynomial coefficients complete the given information about samples as a slope and curvature of MFCC introduced as a time waveform.

A comparison study was conducted using seven methods for extracting these features, from: the time domain signals, the DWT of these signals, both the original signals and the DWT of these signals, the DCT of the time domain signals, both the original signals and the DCT of these signals, the discrete sine transform (DST) of the time domain signals, and both the original signals and the DST of these signals.

It was clear that the extracted features from the DCT achieve the highest recognition rate. This is due to the properties of DCT, from which energy compaction enables accurate feature extraction.

About the time-frequency based approaches, several time-frequency distributions were tried in this scope, including the Short-Time Fourier Transform (STFT) and the higher resolution approaches relying on the famous Wigner-Ville distribution (WVD).

Savelyev, van Kempen and Sahli [126] and Shao, Bouzerdoum and Phung [127] proceed to classify A-scan signals based on different time-frequency approaches. Savelyev, van Kempen and Sahli [126] had proposed to consider the centre of mass as the first singular value of the WVD as a feature vector revealing promising results. Although, Shao, Bouzerdoum and Phung [128] proceeded to a feature selection approach using a sparse kernel algorithm which filters some points from the time-frequency distribution.

2.5. GPR Classification Approaches

Over the last years, many researchers have developed new GPR signal and image classification algorithms for discriminating targets from the background or, in general, false alarms in GPR B-scans for many applications. Hereafter, the approaches obtaining the best results in the field are described.

Al-Nuaimy, Huang, Nakhkash, Fang, Nguyen and Eriksen [120], as several other authors, have proposed to work directly on the classification of A-scan signals based on spectral information (Periodograms) and a three layers MLP, trained with the backpropagation learning algorithm. Hence, after the hyperbola regions are located, the Hough transform is applied to reconstruct the hyperbola parameters. Using this procedure, the computational time of the Hough transform is reduced, and the number of false alarms is minimized.

Gamba and Lossani [11] have pre-processed the image in several phases. After that, every chosen region is presented to a two-layers MLP. The classifier was trained with the backpropagation technique with a momentum term. Mudigonda, Kacelenga and Edwards [129]

performed a comparative study of a convolutional neural network (CNN) and a holographic neural network (HLN) for landmine detection applications. The HLN model attenuates some problems presented by the classical MLP architecture, such as over-training and redundancy of information.

Moysey, Knight and Jol [13] classified GPR facies signatures based on texture features extracted from radar windows. They tested the performance of an MLP classifier with 25 hidden nodes, with a variety of texture feature types. They demonstrated that a set of texture features outperform others. These set of features are:

- The instantaneous amplitude and frequencies defined at specific points in the images.
- The Variance and covariance of pixels in the window.
- The Fourier-Mellin and R transforms over the selected window.
- The Principal component vectors of the radar windows.

Accuracy values greater than 93% were obtained by this approach, within these features independently supplied to the MLP.

Hui-Lin, Wei-Ping and Yu-Hao [12] proposed to recognize different types of mines based on GPR echoes and a novel feature gathering method [130] that first decomposes the signal in terms of Wavelet Packets (WP). Clutter from the signal then separates, and finally the central moments (1 to 4) of every subspace retained after the wavelet decomposition are calculated. The moment features used by the authors present, however, some disadvantages for this application. In general, they are not suitable to recognize noisy and obscured objects. In some cases, they become completely irrelevant. After the most relevant features are selected by a genetic algorithm approach, a constructive neural network approach is used for determining the "optimal" architecture. No details were provided about the nature of the problem, the size of the database, the data splitting strategy, the features vector size, the accuracy of detection, or any other performance criterion. Furthermore, as it was mentioned before, it is more suitable to work on two-dimensional profiles or higher ones. The reason for this is to avoid the lack of information in 1D profiles that are completely misleading in noisy environments and for low depth applications.

Maas and Schmalzl [10] used the Viola Jone's algorithm [9] developed initially for face recognition tasks, based on Haar-like features and the Gentle Adaboost learning algorithm, to

first restrain the hyperbola regions. Subsequently, they applied the Hough transform on these regions to isolate the shape structure.

Besaw and Stimac [131] trained a Deep Belief Network (DBN) to recognize background A-scan signals, hence, they classify landmine targets from background ones. The authors supplied a normalized autocorrelation as input given to the DBN (they used the Viterbi algorithm to track the minimum and maximum of each A-scan).

In a second work, Besaw and Stimac [132] used the previously explained method as a pre-screening procedure in order to find the potential targets. Those targets were given to a CNN to build a features vector.

Sakaguchi, Morton, Collins and Torrione [133] proposed to use a supervised CNN with the aim of extracting features from GPR images, and to classify the corresponding CNN output features using a SVM approach. This procedure was compared with the oriented gradient histogram approach combined with an SVM classifier. The results of the comparison showed that the CNN-SVM approach outperforms the HOG-SVM approach [67].

Gader, Mystkowski and Yunxin [134], Frigui, Ho and Gader [135], Frigui, Missaoui and Gader [63], and Hamdi and Frigui [14] worked with ensemble Hidden Markov (HM) classifiers with different assumptions. Hamdi and Frigui [14] achieved the best results. They tried to solve the problem of signature variations, due to differences in landmine types and the nature of the soil, by modelling every class with an independent HM channel mixture with a final decision stage.

Another approach that was recently proposed by Qin and Huang [65] used Discrete Cosines Transform (DCT) extracted features combined with an SVM classifier to recognize voids, for a simulated database, based on a numerical technique for antenna radiation modelling. The approach gives promising results with high ratios of added noise. Nevertheless, further testing should be conducted on real profiles to validate the approach.

2.5.1. Neural networks approaches

Neural network approaches englobe studies using Multi-Layer Perceptron, a parallel layer perceptron [106], a radial basis function (RBF) network, self-organizing maps [136], and deep learning neural networks.

The Multi-Layer Perceptron approaches [13, 120, 137, 138] used a network trained with a backpropagation algorithm [139].

Few approaches are proposed through the literatures using the RBF network [12, 140]. RBF networks are typically favored over MLP due to:

- Training of RBFs normally employ unsupervised clustering algorithms and least-square solutions that are much faster than backpropagation, and other nonlinear minimization employed by MLPs.
- Approximation of a function for MLPs are done by hyperplanes, while for RBFs Gaussian kernels are employed. Therefore, the extrapolation error is minimal for RBF structure, and it provides better classification capabilities and speedy convergences [141].
- The mapping strategy of a RBF network is more insightful than the one of a MLP.

Deep learning techniques, specifically CNNs, have also produced interesting results in the field.

Besaw and Stimac [131], Besaw and Stimac [132], Lameri, Lombardi, Bestagini, Lualdi and Tubaro [142], Reichman, Collins and Malof [143], and Sakaguchi, Morton, Collins and Torrione [133] experimented different architecture types.

Besaw and Stimac [131] used a deep belief network to classify GPR A-scan data.

Sakaguchi, Morton, Collins and Torrione [133] explored the high capability of CNN to classify GPR images with a basic architecture consisting of a single convolutional layer, for the feature extraction component, and a sigmoid activation function. The Aggregation is performed by taking the max value within non-overlapping blocks across. The output is then used as input in the fully connected part of the CNN. The training is done using a stochastic gradient descent (SGD) optimization algorithm. These authors had also tried to replace the fully connected part by using an SVM classifier in order to further improve the performance.

Besaw and Stimac [132] tested the architecture described in figure 2.7. In an extended version of this work, Besaw [144] tested a slightly modified version of the one depicted in Figure 2.7. The CNN architecture used consisted of two convolutional layers. In addition to a Softmax classification layer, each convolutional layer was followed by a max-pooling stage and a single layer of fully connected hidden nodes. An activation of a REctified Linear Unit (ReLU) type is employed.

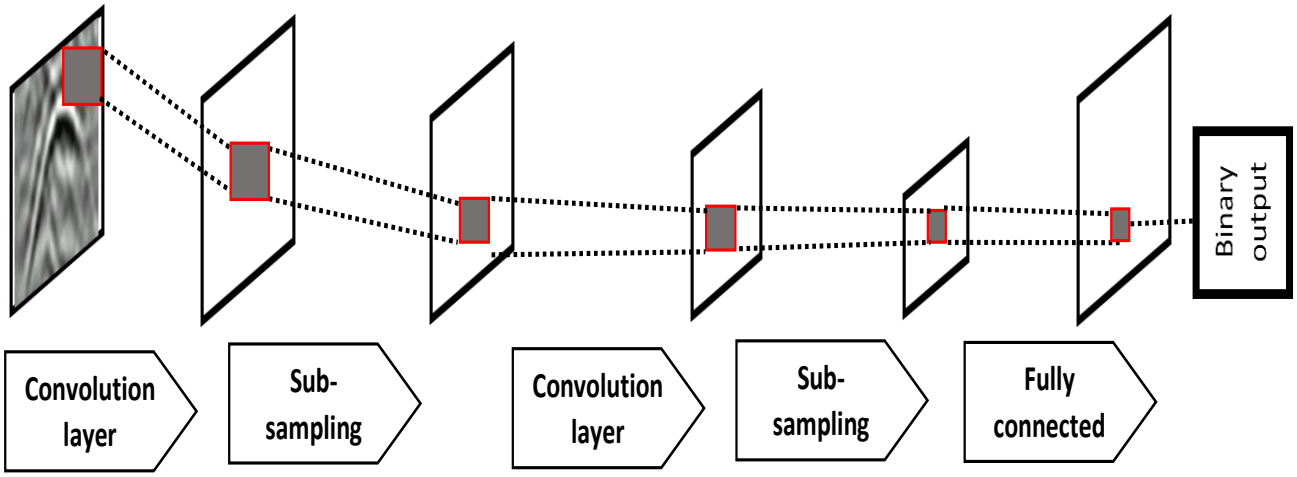


Figure 2. 7: The CNN architecture presented in [144]

Each convolutional layer count consists of 20 filters and 100 nodes in the fully connected hidden layer (FC(100)). Max pooling was performed after each convolutional with a mask of 2x2 (Pool ('Max',2x2, 2)).

A dropout strategy was integrated with a rate of 0.5 to avoid overfitting. Following the hidden layer, a 2-way Softmax classification layer was used in the fully connected stage. Other than the used dropout regularization technique, several other techniques were tested in this work. These techniques include mini-batch training and L2 regularization. Moreover, for fast convergence, other training methods were tested as well. They include the classical backpropagation algorithm, with a momentum term, and decaying the learning rate.

Lameri, Lombardi, Bestagini, Lualdi and Tubaro [142] tested three different network architectures. The first one is composed by 2 convolutional layers with 20 kernels of size 5 × 5 (Conv(5,20)), a ReLU activation layer, a 2 × 2 max-pooling layer (Pool ('Max',2x2, 2)), and, lastly, followed by two fully connected layers of 500 and 2 neurons (FC(500)&FC(2)), respectively. The second architecture is a smaller version of the first one consisting of 3x3 convolutional kernels and 250 neurons on the first fully connected layer. Regarding the third architecture, it is identical to the first architecture, but rather than having two convolutional layers, there is only one.

Reichman, Collins and Malof [143] tested the approach of learning transfer. It is an evolutionary technique found in the literature that exploits the fact that the filters learned at the early stages of the network are not related to the used database. The authors first trained a network on a Cifar10 dataset [145]. The GPR network had the same number of filters in the

first layers as the Cifar10 network. Pre-training was done by transferring the first four convolutional layers. The different architectures tested are described in table 2.1.

Table 2. 1 : The CNN architectures tested in [143]

<table>
<tr><td colspan="11">Patches sizes: 18x18</td></tr>
<tr><td>A</td><td rowspan="2">Conv (3,16)</td><td rowspan="2">Conv (3,16)</td><td rowspan="2">Pool('Max',2x2,2)</td><td rowspan="2">Conv (3,32)</td><td rowspan="2">Conv (3,64)</td><td>Pool ('Max',2x2,2)</td><td>Conv (3,4)</td><td>Conv (3,64)</td><td>Pool ('Max',2x2,2)</td><td>FC (32)</td><td>FC (32)</td><td rowspan="2">FC (16)</td><td rowspan="2">Soft-max</td></tr>
<tr><td>B</td><td colspan="6">FC (64)</td></tr>
<tr><td>C</td><td>Conv (3,24)</td><td>Conv (3,24)</td><td colspan="8">FC (16)</td></tr>
</table>

They employ a convolutional layer with 16 filters of 3 × 3 pixels (Conv(3,16)), a pooling layer within a window of 2×2 pixels (Pool ('Max',2x2, 2)), fully connected layers with 32 neurons (FC(32)), and a ReLu unit placed after each convolutional layer.

A stochastic gradient descendent (SGD) training algorithm is used with momentum term of 0.9 and a L2 regularization factor of 0.0001. A dropout rate of 0.5 is defined for the fully connected layers of the CNN. The learning rate is variable. A mini-batch size is fixed to 32 samples. A number of epochs on training is 16. The proposed learning transfer technique obtained very promising results.

2.5.2. GPR classifier model design

Regarding the inevitable problems presented in the classifier design, in the current literature, the model parameters are typically chosen by trial techniques which depend on the testing results.

In most of the previous approaches, there is no employed algorithm for a complete model design that includes data and feature selection, topology determination, and parameter estimation.

A few works [12, 14] could be reported concerning classifier topology determination.

Hui-Lin, Wei-Ping and Yu-Hao [12] used a constructive algorithm for neural network parameter estimation and topology determination. It presented significant disadvantages compared with othr tools such as evolutionary algorithms. The most evident of them is that the

network is constrained to use a set of features selected previously. In addition, the execution time and the model complexity were not taken into consideration as a design criterion for the proposed approach.

2.5.3. Feature selection

Feature selection is a commonly used operation in the machine learning field. It selects a subset of the extracted features from the data before applying the classification algorithm.

The best chosen subset is the less dimensional one that achieves higher accuracy.

A few feature selection algorithms could be reported in the GPR data classification field: Genetic algorithm (GA) [12], RELIF [146, 147], sparse kernel [128], Principal Components Analysis (PCA) [148], and Iterative Search Margin-Based Algorithm (SAMBA) [146].

PCA is a dimensionality reduction algorithm, which is a bit different from a feature selection process as it is a unsupervised algorithm..

In fact, these algorithms could be classified in two main categories: GA as a wrapper technique; and filter methods, including RELIF [149] based on a statistical relevance of the subset of features [149], SAMBA [150] relying on separation margins between classes, and the sparse kernel approach [128].

Wrapper methods try to estimate the quantity of information held by a feature of a subset by a predictor or a learning algorithm. The aim is to optimize the performance of the predictor by assessing the generalization performance.

However, filter methods compute a kind of a score for each feature of the subset and rank them depending on this score, via an evaluation function that tends to achieve the goal of maximizing the information held by the subset. The most well-known evaluation functions are mutual information, margin, and dependence measures. These basic filter methods present a disadvantage for GPR data application which is the problem of the separability of dependencies between features, especially when clutter noise is present. Nevertheless, variable ranking methods, such as RELIEF, try to overcome this issue.

While trying to overcome the issue of higher computation time, an approach has been proposed to take advantage of the GA based methods by performing a two-stage feature selection procedure based on another filter approach relying on mutual information [151].

2.6. Conclusions

Though there are many issues to address, GPR data inversion is a very wide area of research focused on regrouping and interconnecting these issues. The goal is to eventually arrive and achieve a step of initial profile reconstruction and data interpretation.

Fuelled by this up-to-date literature review chapter, a detailed study about the different approaches is employed in order to resolve the main setbacks presented, aiming to easily compare our contributions with the existing studies in the field. All will be presented in the next following chapters. In conclusion, major issues and proposed solutions are highlighted in comparison with the existing approaches.

.

3. Theoretical Foundations

3.1. Introduction

GPR is an electromagnetic sensing technique involving the Maxwell wave's propagation theory. Data inversion is the process of solving mathematical equations presented by a wave propagation within this theory which, in turn, is referred to as an "inverse problem". In order to reconstruct the investigated subsoil model, an inverse problem is used. To invert the data, it is necessary to resolve the primary issue that involves the mathematical modelling of wave propagation in pursuit to find and approximate a solution from this equation, under certain assumptions. For this thesis, the modelled profiles were used to conduct the first tests of the developed algorithms.

Within this chapter, all the details of fundamental materials for the thesis are provided. The chapter gives a short description of GPR electromagnetic background theory, and explains, in detail, the modelling process using finite difference time domain (FDTD) discretization technique in section 3.2.

The classical Hough pattern recognition approach used for hyperbola signatures reconstruction is presented in section 3.3. In section 3.4, a time-frequency analysis is presented, due to its functionality as a tool for permittivity inversion since the velocity calculated by Hough transform is the mean value for two mediums.

In order to reduce the computational time of Hough transform, a classification approach was used in order to narrow down the hyperbolas into certain regions before applying the Hough transform. Feature extraction techniques and selection procedure were deployed to realize this task and therefore are detailed in section 3.5. Moreover, section 3.6 gives a background about the classifiers while the theory concerning the multi-objective genetic optimization algorithm used to design the neural network is resumed in section 3.7.

3.2. GPR Modelization

Ground radar is a geophysical prospecting method relying on the principle of propagation of electromagnetic waves (EM) of frequencies varying in the [1MHz-3GHz] band. Electromagnetic waves are reflected or diffracted at the boundaries of objects that exhibit

differences in electrical and magnetic properties. The dielectric permittivity, electrical conductivity, and magnetic permeability are the three parameters which determine the reflectivity of layer boundaries and the depth of penetration [150].

3.2.1. The propagation of electromagnetic waves

3.2.1.1. Maxwell's equations

The phenomenon of the propagation of electromagnetic waves in any medium is governed by Maxwell's equations [151], which are expressed in terms of the components of the electric and the magnetic fields. In a homogeneous and isotropic medium, these equations are in the form:

$$\begin{cases} \mathrm{rot}\vec{E} = -\frac{d\vec{B}}{dt} \\ \mathrm{rot}\vec{H} = \vec{J} + \frac{d\vec{D}}{dt} \\ \mathrm{div}\vec{B} = 0 \\ \mathrm{div}\vec{D} = q \end{cases} \quad (3.1)$$

Where: $\vec{E}$ is the electric field (V/m)
$\vec{H}$ is the magnetic field (A/m)
$\vec{E}$ is the intensity of the magnetic induction (T)
$\vec{J}$ is the density of the electric current (A/m2)
$\vec{D}$ is the electric displacement current (C/m2)
q is the electrostatic charge density (C/m3)

The first two equations are the equations of Maxwell Faraday and Maxwell Ampere, respectively [152].

The deep layers of the subsoil are generally considered as a homogeneous and isotropic medium, or the propagation laws can be modelled by the following wave equations:

$$\Delta\vec{E} - \mu_0\sigma\frac{d\vec{E}}{dt} - \mu_0\varepsilon\frac{d^2\vec{E}}{dt^2} = \vec{0} \quad (3.2)$$

$$\Delta\vec{H} - \mu_0\sigma\frac{d\vec{H}}{dt} - \mu_0\varepsilon\frac{d^2\vec{H}}{dt^2} = \vec{0} \quad (3.3)$$

Where Δ is the Laplacian operator.

Equation (3.2) has as a harmonic solution:

$$E_x = E_0 e^{j\omega t - \gamma z} \quad (3.4)$$

The propagation factor of the electromagnetic waves y is given by the formula:

$$\gamma = \alpha + j\beta \quad (3.5)$$

Where :

$$\alpha = \omega\sqrt{\frac{\mu_0\varepsilon_0 \mathrm{real}(\varepsilon_r)}{2}\left(\sqrt{1+\tan^2\delta}-1\right)} \quad (3.6)$$

$$\beta = \omega\sqrt{\frac{\mu_0\varepsilon_0 \mathrm{real}(\varepsilon_r)}{2}\left(\sqrt{1+\tan^2\delta}+1\right)} \quad (3.7)$$

The loss angle δ is defined by the formula:

$$\tan\delta = \frac{\sigma}{\omega \mathrm{real}(\varepsilon)} \quad (3.8)$$

In the case where $\tan\delta = 1$, there is the equality between the conduction and the displacement currents. The propagation of the electromagnetic wave is favoured only in the case where $\tan\delta < 1$, whereas for $\tan\delta > 1$, the diffusive process dominates (this is the case for low frequencies) [153].

3.2.1.2. Propagation through an interface

The propagation of electromagnetic waves is linked to permeability, permittivity, and conductivity, thus, a change in soil characteristics may cause a wave re-transmission. Let two homogeneous and isotropic media be characterized respectively by the parameters $(\mu_1, \varepsilon_1, \sigma_1)$ and $(\mu_2, \varepsilon_2, \sigma_2)$, and separated by a plane interface (Figure 3.1).

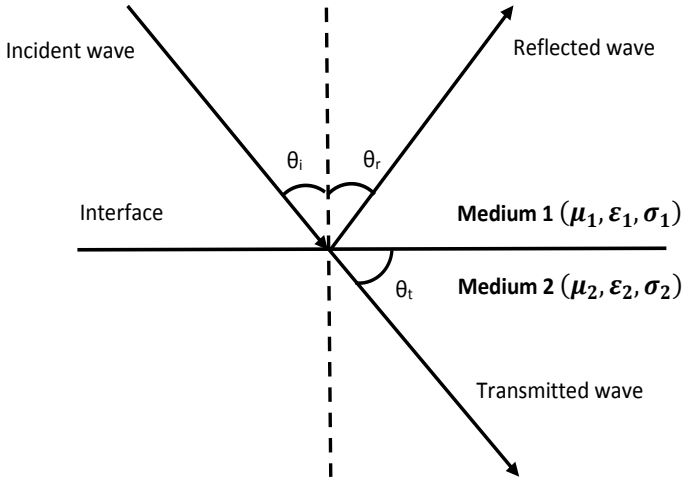


Figure 3. 1 : Diagram showing the phenomena of wave propagation at the interface between two different media

The incident wave on the interface of the two media is reflected in medium one and transmitted in medium two. The directions of the incident, reflected and transmitted waves are given by the laws of Snell-Descartes:

$$\frac{\sin\theta_i}{v_1} = \frac{\sin\theta_r}{v_1} = \frac{\sin\theta_t}{v_2} \quad (3.9)$$

Where: θ_i is the incidence angle

θ_r the reflection angle

θ_t the transmission angle

v_1 and v_2 are the electromagnetic wave velocity in the first and second medium respectively.

The refractive index n_{21} is given by the formula:

$$n_{21} = \frac{\sin\theta_i}{\sin\theta_t} = \frac{v_1}{v_2} \tag{3.10}$$

For the case where $v_2 > v_1$, there is a critical angle θ_c, corresponding to a transmission angle equal to 90°, which cause a phenomena called critical refraction of the wave. The critical angle θ_c is given by:

$$\theta_c = \arcsin\left(\frac{v_1}{c_0}\right) \tag{3.11}$$

Where c_0 is the celerity.

This angle correspond to a critical offset:

$$x_c = \frac{2hv_1}{\sqrt{c_0^2 - v_1^2}} \tag{3.12}$$

In this particular situation, where the angle of incidence is $\theta_i < \theta_c$, the refracted wave moves along the interface at the speed v2. In this case, there is a total reflection of the wave.

The amount of radar energy retrieved is indicated by the coefficients of reflection and transmission. The application of Maxwell`s equations to the interface between two media, coupled with the law of Snell-Descartes, allows us to drawback to the conditions of continuity of the electric and Magnetic fields between the two media. The formula of the reflection and transmission coefficients is established depending on the polarization of the incident wave.

When the polarization of the magnetic field is parallel to the incidence plan, it is the TM mode (Magnetic Transverse). When the polarization of the electric field is perpendicular to the incidence plan, it is the TE mode (Electrical Transverse).

The TE mode is the mode most used when acquiring surface radar. In this mode, the coefficient of reflection and transmission, known as Fresnel coefficient, between two perfect dielectric materials (non-conductive and non-magnetic), are given by:

$$R_\perp = \frac{\sqrt{\varepsilon_1}\cos\theta_i - \sqrt{\varepsilon_2\left(1 - \frac{\varepsilon_1}{\varepsilon_2}\sin^2(\theta_i)\right)}}{\sqrt{\varepsilon_1}\cos\theta_i + \sqrt{\varepsilon_2\left(1 - \frac{\varepsilon_1}{\varepsilon_2}\sin^2(\theta_i)\right)}} \tag{3.13}$$

$$T_\perp = 2\frac{\sqrt{\varepsilon_2\left(1 - \frac{\varepsilon_1}{\varepsilon_2}\sin^2(\theta_i)\right)}}{\sqrt{\varepsilon_1}\cos\theta_i + \sqrt{\varepsilon_2\left(1 - \frac{\varepsilon_1}{\varepsilon_2}\sin^2(\theta_i)\right)}} \tag{3.14}$$

However for TM mode the reflection and transmission coefficient are formulated as:

$$R_{\parallel} = \frac{\sqrt{\varepsilon_2}\cos\theta_i - \sqrt{\varepsilon_1\left(1-\frac{\varepsilon_1}{\varepsilon_2}\sin^2(\theta_i)\right)}}{\sqrt{\varepsilon_2}\cos\theta_i + \sqrt{\varepsilon_1\left(1-\frac{\varepsilon_1}{\varepsilon_2}\sin^2(\theta_i)\right)}} \quad (3.15)$$

$$T_{\parallel} = 2\frac{\sqrt{\varepsilon_2\left(1-\frac{\varepsilon_1}{\varepsilon_2}\sin^2(\theta_i)\right)}}{\sqrt{\varepsilon_2}\cos\theta_i + \sqrt{\varepsilon_1\left(1-\frac{\varepsilon_1}{\varepsilon_2}\sin^2(\theta_i)\right)}} \quad (3.16)$$

The Angle of incidence is related to the separation of the antennas in acquisition process [150].

3.2.2. The acquisition configurations

GPR surveys are classified into reflection and transillumination measurements (Annan, 2009). For reflection measurements, there are two commonly deployed configurations: common offset (CO) and common midpoint (CMP). However, for transillumination or boreholes measurement configurations, antennas are placed on the ground or installed into trenches or drilled wells.

3.2.2.1. Common-offset (CO) survey

In a common-offset survey, the transmitter and receiver scan the survey area with a fixed spacing constant while acquiring the data at each measurement location (figure 3.2.a). For every position, the transmitter gathers a 1D profile containing the emitted wave followed by the reflected echoes that are spaced in time, depending on the characteristics of the mediums. For a single survey line, the acquired GPR data corresponds to a 2D reflectivity map of the subsurface below the surveyed area called a radargram. By gathering multiple radargrams in parallel lines, 3D maps can be constructed.

3.2.2.2. Common midpoint (CMP) survey

In a common midpoint survey, a transmitter and receiver are placed on the ground and separated by a variable distance while keeping a centre position of the antennas constant. While varying separation distance between antennas, various signal paths with the same point of reflection are obtained (figure 3.2.b). The data can then be used to estimate radar signal velocity distribution versus subsurface depth [153, 154].

When the transmitter is fixed, and only the receiver is moved away from the transmitter, the survey is called a wide-angle reflection and refraction (WARR).

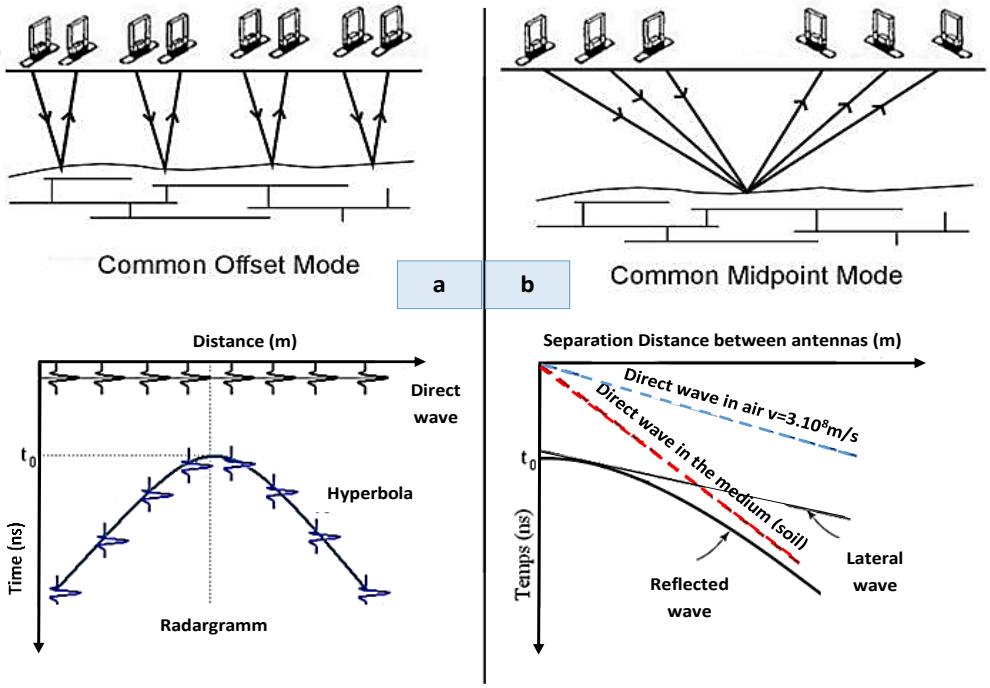


Figure 3. 2: a) Radargram (low) acquired in common offset mode above one object (top) and b) in common midpoint mode above a flat interface. For the case of common offset configuration (i.e. case a)) down a flat interface the obtained signature is a straight line

3.2.2.3. Transillumination measurements

There are two commonly used configurations: Zero-offset profiling (ZOP) and multi-offset gather (MOG).

ZOP uses a configuration where the transmitter and receiver are moved in two parallel boreholes with a constant distance. It is the easiest configuration to locate velocity anomalies [154].

For MOG configuration, transmission signals are gathered through the volume between boreholes while varying angles. Tomographic imaging, constructed from the survey data, can provide a complete map of dielectric properties over the survey volume.

3.2.3. FDTD modelisation

Solving Maxwell equations is modelled by the finite difference method (FDTD) with UPML (uniaxial perfectly matched layers) boundary conditions.

Surface to surface prospecting model is considered, i.e. by TMz mode. The two equations of Maxwell-Ampere and Maxwell-Faraday are given by:

$$\begin{cases} \nabla\vec{H} = (j\omega\varepsilon_0\varepsilon_r)\bar{\bar{S}}\vec{E} \\ \nabla\vec{E} = (j\omega\mu)\bar{\bar{S}}\vec{H} \end{cases} \tag{3.17}$$

Where: ε_0 is the vacuum permittivity.
ε_r is the relative permittivity.
μ is the permeability.
ω is the pulsation of the wave.
$\vec{E}, \vec{H}$ are the electric and the magnetic fields respectively

$\bar{\bar{S}}$ is the diagonal tensor defined by equation (3.18) :

$$\bar{\bar{S}} = \begin{bmatrix} \frac{S_y S_z}{S_x} & 0 & 0 \\ 0 & \frac{S_x S_z}{S_y} & 0 \\ 0 & 0 & \frac{S_x S_y}{S_z} \end{bmatrix} \tag{3.18}$$

Allowing a non-real unitary part K, the components of the diagonal elements of multiplication $\bar{\bar{S}}$ are given by [155] :

$$\begin{cases} S_x = K_x + \frac{\sigma_x}{j\omega\varepsilon_0} \\ S_y = K_y + \frac{\sigma_y}{j\omega\varepsilon_0} \\ S_z = K_z + \frac{\sigma_z}{j\omega\varepsilon_0} = 1 \end{cases} \tag{3.19}$$

To absorb outgoing waves, the UPML is incorporated into Maxwell's curl equations through the tensor $\bar{\bar{S}}$ (1). In fact, UPML works by incorporating fictitious loss terms, σ_x, σ_y and σ_z, while simultaneously matching impedance. In this manner, fields decay without being artificially reflected back into the problem space. For optimal performance, the loss terms should increase gradually into the UPML. A performed grading profile is proposed as follow [155, 156] :

$$\begin{cases} \sigma_i = \sigma_{i,max}\left(\frac{i}{d}\right)^m \\ K_i = 1 + \left(K_{i,max} - 1\right)\left(\frac{i}{d}\right)^m \\ \sigma_{i,max} = -\frac{(m+1)ln(R)}{2\eta d} \\ i = x,y;\ K_{i,max} = 1;\ d = 10\ cells; \\ m = 4;\ R = e^{-16};\ \eta = \sqrt{\frac{\mu}{\varepsilon}} \end{cases} \tag{3.20}$$

Where d is the thickness of the UPML, R is the desired reflection error, and η is the intrinsic impedance.

Two auxiliary variables $\vec{P}$ and $\vec{D}$, which permit modelling the presence of a dispersive material. These variables are defined by equations (3.21) [155].

$$\begin{cases} P_x = \varepsilon_0 \varepsilon_r(\omega) {S_x}^{-1} E_x \\ P_y = \varepsilon_0 \varepsilon_r(\omega) S_x {S_y}^{-1} E_y \\ P_z = \varepsilon_0 \varepsilon_r(\omega) S_y E_z \\ D_z = \varepsilon_0 S_y E_z \end{cases} \tag{3.21}$$

The implementation of dispersion is made through "Auxiliary Differential Equation method" technique (detailed in [155, 157, 158]), and the model implemented is Debye one, given by:

$$\varepsilon_r(\omega) = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + j\omega\tau} + \frac{\sigma}{j\omega\varepsilon_0} \tag{3.22}$$

Where : ε_∞ : is the relative permittivity at high-frequency.

ε_s : is the relative Permittivity at low-frequency.

τ : is acharacteristic relaxation time of the medium.

σ : is the conductivity of the medium.

Thus, the discretization of the field equations [155, 157, 159] :

$$\begin{cases} P_{z\,i,j}^{n+1} = (C_{1ez}) P_{z\,i,j}^{n} + (C_{2ez}) \left[\dfrac{H_{y\,i+\frac{1}{2},j}^{n+\frac{1}{2}} - H_{y\,i-\frac{1}{2},j}^{n+\frac{1}{2}}}{\Delta x} - \dfrac{H_{x\,i+\frac{1}{2},j}^{n+\frac{1}{2}} - H_{x\,i-\frac{1}{2},j}^{n+\frac{1}{2}}}{\Delta y} \right] \\ D_{z\,i,j}^{n+1} = \dfrac{(a_1) P_{z\,i,j}^{n+1} + (a_2) P_{z\,i,j}^{n} + (a_3) P_{z\,i,j}^{n-1} - (b_2) D_{z\,i,j}^{n} - (b_3) D_{z\,i,j}^{n-1}}{b_1} \\ E_{z\,i,j}^{n+1} = (C_{3ez}) E_{z\,i,j}^{n} + (C_{4ez}) \left[(C_{5ez}) D_{z\,i,j}^{n+1} - (C_{6ez}) D_{z\,i,j}^{n} \right] \end{cases}$$

(3.23)

Accordingly, the field equations $\vec{B}$ are given by [155]:

$$\begin{cases} B_{x\,i,j+\frac{1}{2}}^{n+\frac{3}{2}} = (D_{1hx}) B_{z\,i,j\frac{1}{+2}}^{n+\frac{1}{2}} - (D_{2hx}) \left[\dfrac{E_{z\,i,j+1}^{n+1} - E_{z\,i,j}^{n+1}}{\Delta y} \right] \\ H_{x\,i,j+\frac{1}{2}}^{n+\frac{3}{2}} = (D_{3hx}) H_{x\,i,j+\frac{1}{2}}^{n+\frac{1}{2}} + (D_{4hx}) \left[(D_{5hx}) B_{x\,i,j+\frac{1}{2}}^{n+\frac{3}{2}} - (D_{6hx}) B_{x\,i,j+\frac{1}{2}}^{n+\frac{1}{2}} \right] \\ B_{y\,i,j+\frac{1}{2}}^{n+\frac{3}{2}} = \left(D_{1hy}\right) B_{z\,i,j\frac{1}{+2}}^{n+\frac{1}{2}} - \left(D_{2hy}\right) \left[\dfrac{-\left(E_{z\,i,j+1}^{n+1} - E_{z\,i,j}^{n+1}\right)}{\Delta y} \right] \\ H_{y\,i,j+\frac{1}{2}}^{n+\frac{3}{2}} = \left(D_{3hy}\right) H_{y\,i,j+\frac{1}{2}}^{n+\frac{1}{2}} + \left(D_{4hy}\right) \left[\left(D_{5hy}\right) B_{y\,i,j+\frac{1}{2}}^{n+\frac{3}{2}} - \left(D_{6hy}\right) B_{y\,i,j+\frac{1}{2}}^{n+\frac{1}{2}} \right] \end{cases}$$

(3.24)

Where the discretization coefficients are regrouped in table 3.1.

$$\begin{cases} D_{z\,i,j}^{n+1} = \frac{(a_1)P_{z\,i,j}^{n+1} + W_1^{n+1}}{b_1} \\ W_1^{n+1} = (a_2)P_{z\,i,j}^{n+1} - (b_2)D_{z\,i,j}^{n+1} + W_1^{n+1} \\ W_2^{n+1} = (a_3)P_{z\,i,j}^{n+1} - (b_3) \end{cases} \quad (3.25)$$

Where the coefficients a_i and b_i; i=1,2,3; are given in table 3.1.

Table 3. 1 : The coefficients of the discretization of the fields $\vec{E}$ and $\vec{B}$ with UPML boundary conditions

C_{1ez}	$\frac{2\varepsilon K_x - \sigma_x \Delta t}{2\varepsilon K_x + \sigma_x \Delta t}$	a_1	$\frac{\varepsilon_0}{2\Delta t} + \frac{\varepsilon_0 \tau}{\Delta t^2}$	D_{1hx}	$\frac{2\varepsilon K_y - \sigma_y \Delta t}{2\varepsilon K_y + \sigma_y \Delta t}$	D_{1hy}	1
C_{2ez}	$\frac{2\varepsilon \Delta t}{2\varepsilon K_y + \sigma_y \Delta t}$	a_2	$-\frac{\varepsilon_0 \tau}{2\Delta t^2}$	D_{2hx}	$\frac{2\varepsilon \Delta t}{2\varepsilon_0 K_y + \sigma_y \Delta t}$	D_{2hy}	Δt
C_{3ez}	$\frac{2\varepsilon K_y - \sigma_y \Delta t}{2\varepsilon K_y + \sigma_y \Delta t}$	a_3	$-\frac{\varepsilon_0}{2\Delta t} + \frac{\varepsilon_0 \tau}{\Delta t^2}$	D_{3hx}	1	D_{3hy}	$\frac{2\varepsilon K_x - \sigma_x \Delta t}{2\varepsilon K_x + \sigma_x \Delta t}$
C_{4ez}	$\frac{1}{\varepsilon_0(2\varepsilon K_y + \sigma_y \Delta t)}$	b_1	$\frac{\varepsilon_0 \varepsilon_s + \sigma\tau}{2\Delta t} + \frac{\varepsilon_0 \varepsilon_\infty \tau}{\Delta t^2} + \frac{\sigma}{2}$	D_{4hx}	$\frac{1}{2\varepsilon\mu}$	D_{4hy}	$\frac{1}{\mu(2\varepsilon K_x + \sigma_x \Delta t)}$
C_{5ez}	2ε	b_2	$-\frac{2\varepsilon_0 \varepsilon_\infty \tau}{\Delta t^2}$	D_{5hx}	$2\varepsilon K_x + \sigma_x \Delta t$	D_{5hy}	$2\varepsilon K_y + \sigma_y \Delta t$
C_{6ez}	2ε	b_3	$-\frac{\varepsilon_0 \varepsilon_s + \sigma\tau}{2\Delta t} + \frac{\varepsilon_0 \varepsilon_\infty \tau}{\Delta t^2} + \frac{\sigma}{2}$	D_{6hx}	$2\varepsilon K_x - \sigma_x \Delta t$	D_{6hy}	$2\varepsilon K_y - \sigma_y \Delta t$

The convergence criterions used are given by equations (3.26) [158]

$$\begin{cases} dx = dy = dz = \frac{1}{\sqrt{\varepsilon_0 \mu_0 \varepsilon_{r_{max}} \mu_{r_{max}}}} \frac{1}{10 f_{max}} \\ dt = \frac{dx\sqrt{\varepsilon_0 \mu_0 \varepsilon_{r\_min} \mu_{r\_min}}}{2} \end{cases} \quad (3.26)$$

dt is the time discretization step, dx, dy, and dz are the spatial discretization steps. A regular mesh is chosen.

3.3. Generalized Hough Transform

3.3.1. Overview

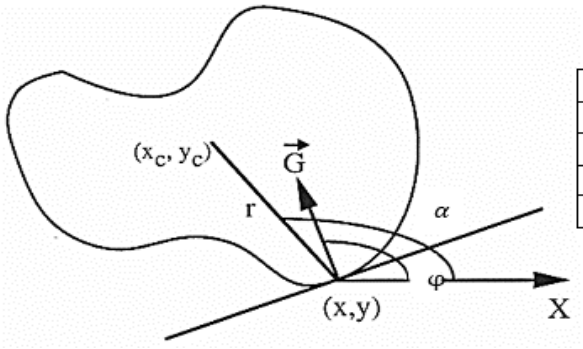
The Hough transform [31] is universally used for shape detection due to its capability to handle multiple occurrences of parametric and nonparametric shapes. The Hough transform has a generalized form, applicable to the arbitrary nonparametric class of shapes, known as the Generalized Hough Transform (GHT) [2, 160].

An object model is represented by an equation:

$$f(x, y, \beta_1, \dots, \beta_k) = 0 \quad (3.27)$$

x and y represent the pixel position in the image, while $(\beta_1, \dots, \beta_k)$ is a set of k parameters defining the shape. In the k-dimensional parametric space, designed as the Hough space and spanned by the variables $(\beta_1, \dots, \beta_k)$, this equation (3.27) can represent some different shapes with x and y considered as two parameters.

Being said so, every object is modelled in terms of a lookup R-table that can be used to perform mapping between an image space and a Hough space. An example is presented in figure 3.3.



$\varphi_1 = 0$	$(r,\alpha)_{1_1}$	$(r,\alpha)_{1_2}$	...	$(r,\alpha)_{1k_1}$
...	...	...	...	...
φ_j	$(r,\alpha)_{j_1}$	$(r,\alpha)_{j_2}$	...	$(r,\alpha)_{jk_1}$
...	...	...	...	...
$\varphi_n = \pi$	$(r,\alpha)_{n_1}$	$(r,\alpha)_{n_2}$	...	$(r,\alpha)_{nk_1}$

Figure 3. 3: An example of R-table for GHT

The algorithm is summarized in table 3.2. The parameters r and α are given by the formulas:

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2} \tag{3.28}$$

$$\alpha = tan^{-1}\frac{y - y_c}{x - x_c} \tag{3.29}$$

Table 3. 2: Generalized Hough Transform algorithm

Step 1	• Create a table with n entries, each one indexed by an angle $\varphi_i \epsilon \left[0 : \frac{180}{n} : 180°\right]$, $i = 1 : n$ n is the resolution of the gradient orientation.
Step 2	○ Set a reference point (x_c, y_c) inside the 2D shape that could represent the gravitational center for some defined shapes. ○ For each point (x, y) on the boundary of the shape (or the edge), calculate the two parameters r and α, and the gradient orientation $ang(\overrightarrow{G})$. ○ If the angle φ is the closest to $ang(\overrightarrow{G})$. Then add the correspondent pair (r, α) to this table.
Step 3	○ Create a Hough array $H(x_c, y_c)$ and initialize it to zero

Step 4	○ For each image pixel (x, y) with $\|G(x,y)\| > T_g$, find the table entry with its corresponding angle φ_j closest to $ang(\overrightarrow{G})$. ○ For each of the k_j pairs $(r, \alpha)_i \quad i = 1, \dots, k_j$ in this table entry, find : $\begin{cases} x_c = x + r\cos\alpha \\ y_c = y + r\sin\alpha \end{cases}$ ○ Increment the corresponding element in the H array by 1.
Step 5	• Localize the maxima in table H: points satisfying $H(x_c, y_c) > T_h$ represent the locations of the searched shape in the image.

3.3.2. Hyperbolas detection

The simplest case is the detection of lines. For hyperbolic arcs, the theory is more complicated and the algorithm application is more difficult. Since a 3D Hough array is deployed the computing time increases dramatically even because it depends on the image size too.

Let's have a scatterer with a size ρ comparable to the wavelength, embedded in a homogeneous medium with a constant velocity v, and at a depth d below the inspection surface (please refer to figure 3.4).

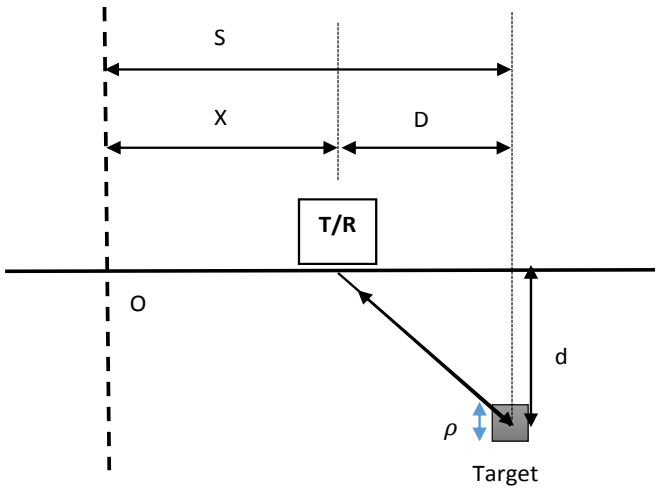


Figure 3. 4 : GPR B-scan image generation principal in surface-surface acquisition mode [33]

When the probe is at a horizontal distance D, measured along the inspection surface from the normal to the surface which passes through the scatterer, then the round-trip travel time or time of propagation of the wave (two ways), is:

$$T_p = 2t_p = 2\frac{\sqrt{d^2+D^2}}{v} \tag{3.30}$$

Since the horizontal position of the scatterer is not known in this prospection configuration type, the travel time is reformulated as:

$$T_p = 2t_p = 2\frac{\sqrt{d^2+(S-X)^2}}{v} \tag{3.31}$$

The horizontal and vertical axis are discretized as:

$$\begin{cases} T_p = jdt \\ S = i_s dx \\ X = i_x dx \end{cases} \tag{3.32}$$

Where dx and dt represent space and time discretization steps respectively.

Hence, equation (3.31) will be discretized as:

$$j^2 = \alpha + \beta(i-\gamma)^2 \tag{3.33}$$

Where the three parameters (α, β, γ) describe the apex of the hyperbolas:

$$\begin{cases} \alpha = \frac{4d^2}{dt^2 v^2} \\ \beta = \frac{4dx^2}{dt^2 v^2} \\ \gamma = i_s \end{cases} \tag{3.34}$$

Figure 3.5 represents the geometrical meaning of these three parameters (α, β, γ), which are related to the position (vertical and horizontal location below the surface of inspection) and the dimension of the scatterer by the following equations [33]:

$$\begin{cases} S = \gamma dx \\ d = 2\alpha^{0.5}\frac{dt}{v} \\ \rho = k\frac{\alpha}{\beta} \end{cases} \tag{3.35}$$

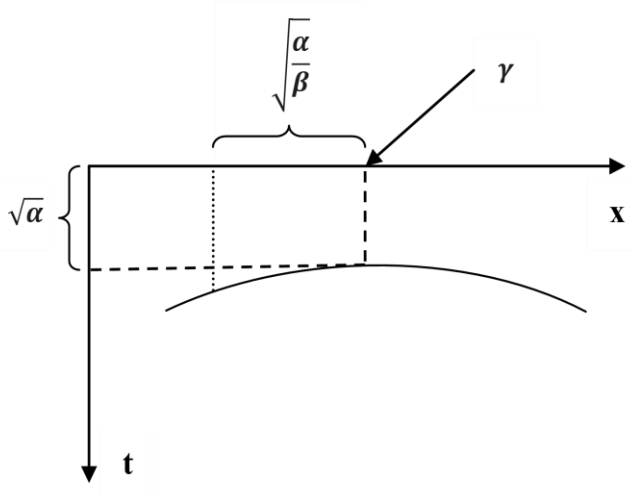


Figure 3. 5: Hyperbola structure defined by the three parameters, α, β, and γ

The Generalized Hough Transform algorithm is slightly modified to be projected for the case of hyperbolas characterized by the parameters(α, β, γ). Algorithmic steps are detailed in table 3.3.

Parameters (α, β, γ) are calculated by picking three distinct points (x, y) , (x_1, y_1) and (x_2, y_2), based on equations:

$$\gamma = \frac{[(x_2^2 - x^2)(y_1^2 - y_2^2) - (x_1^2 - x_2^2)(y_2^2 - y^2)]}{[2(x_2 - x)(y_1^2 - y_2^2) - (x_1 - x_2)(y_2^2 - y^2)]} \tag{3.36}$$

$$\beta = \frac{y_1^2 - y_2^2}{(x_1 - x_2)(x_1 - x_2 - 2\gamma)} \tag{3.37}$$

$$\alpha = y^2 - \beta(x - \gamma)^2 \tag{3.38}$$

Table 3. 3: Generalized Hough Transform algorithm for hyperbola detection

Step 1	Calculate the possible intervals of parameters (α, β, γ): • Starting from the deployed frequency of the survey that corresponds to a limit depth acquisition parameter α variation interval is deduced. ○ Given the length of the surface of prospection, the interval of variation of parameter β, and γ is calculated. ○ Create a Hough array $H(\alpha_i, \beta_i, \gamma_i)$ and initialize to zero
Step 2	○ For each point (x, y) on the edge of the original image, choose another neighbouring two points (x_1, y_1) and (x_2, y_2). ○ Given this three points calculate the values of parameters (α, β, γ), using equations (3.36), (3.37), and (3.38). ○ Increment the corresponding entry in the Hough array by 1.
Step 3	○ Localize the maxima in table H: points satisfying $H(\alpha_i, \beta_i, \gamma_i) > T_h$ represent the locations of the apexes of searched hyperbolas.
Step 4	○ To reconstruct the hyperbolas, given the apexes, for every probable point of the hyperbolas two other neighbouring points are picked to verify if the parameters are identical.

3.4. Time-Frequency Analysis

The Fourier Transform is not an adequate tool for analysing a non-stationary GPR signal. Information localized in time, such as spikes and high frequency bursts, could not be distinguished in the Fourier Transform distribution. These facts give rise to the use of time-frequency representation. There are two major types of time-frequency distributions: The non-quadratic ones, and the quadratic ones characterized by their higher resolution.

3.4.1. Non-Quadratics representations

3.4.1.1. Short Time Fourier transform (STFT)

The Short Time Fourier Transform (STFT) performs time localization by windowing the signal to segment some localized slices of the signal and then performs it`s Fourier Transform. The STFT $ST(\omega)$ of a signal sig(t) is defined as [47]:

$$ST_{sig}(\omega) = \int_{-\infty}^{+\infty} sig(t)w\,(t-\tau)e^{-j\omega t}dt \tag{3.39}$$

The squared magnitude of the STFT represent the spectrogram. The major drawback of STFT is that it uses a fixed window for analysing all the segments of the signal offering a constant resolution for all frequencies presented in the signal.

3.4.1.2. Scalogram

It is a quadratic linear time-frequency distribution with a wavelet based kernel. It is very similar to the STFT with a varying window length by the means of scaling the axis of the window. The Scalogram is defined as the energy density function of the wavelet transform (WT) [161]:

$$WT_{sig}(\alpha,\beta) = \frac{1}{\sqrt{\beta}}\int_{-\infty}^{+\infty} sig(t)w^*(\frac{t-\alpha}{\beta})\,dt \tag{3.40}$$

The WT is a simple convolution operation of signal sig(t) and a window w(t) shifted and dilated in time by the two quantities α and β. The parameter α can be chosen such that it is inversely proportional to the frequency to obtain a TF representation comparable to the STFT.

3.4.1.3. Stockwell Transform

The S transform of a signal sig(t) is defined as a WT with a specific mother wavelet multiplied by a phase factor:

$$SwT_{sig}(t,\tau) = \int_{-\infty}^{+\infty} e^{-2j\pi f\tau}w(\tau,f) \tag{3.41}$$

Where the window function is given by the formula:

$$w(t,\tau) = \frac{|sig(t)|}{\sigma\sqrt{2\pi}}e^{-\frac{t^2f^2}{2k^2}}e^{-2j\pi ft}, \qquad \forall\sigma > 0 \tag{3.42}$$

3.4.1.4. Hilbert Huang Transform (HHT)

HHT [162] was initially developed to analyse water-wave evolution, but it has drawn interesting results while analysing other non-stationary signals. It consists of two major stages: the empirical mode decomposition (EMD) and the Hilbert transform (HT) stage [162, 163].

The EMD decomposes the time series into a set of narrow-band sequences or functions called intrinsic mode function (IMF). The HT of these sequences represents local amplitude and frequency attributes. The iterative EMD procedure is summarized in figure 3.6 where an IMF must satisfy two properties:

- It must have a zero-mean.
- The number of extrema and zero-crossings must differ by at most one.

The process described in figure 3.6 continues until the final residue (the final residue h(t)=r(t)) r(t) is either a constant, a monotonic slice or contains only one extrema.

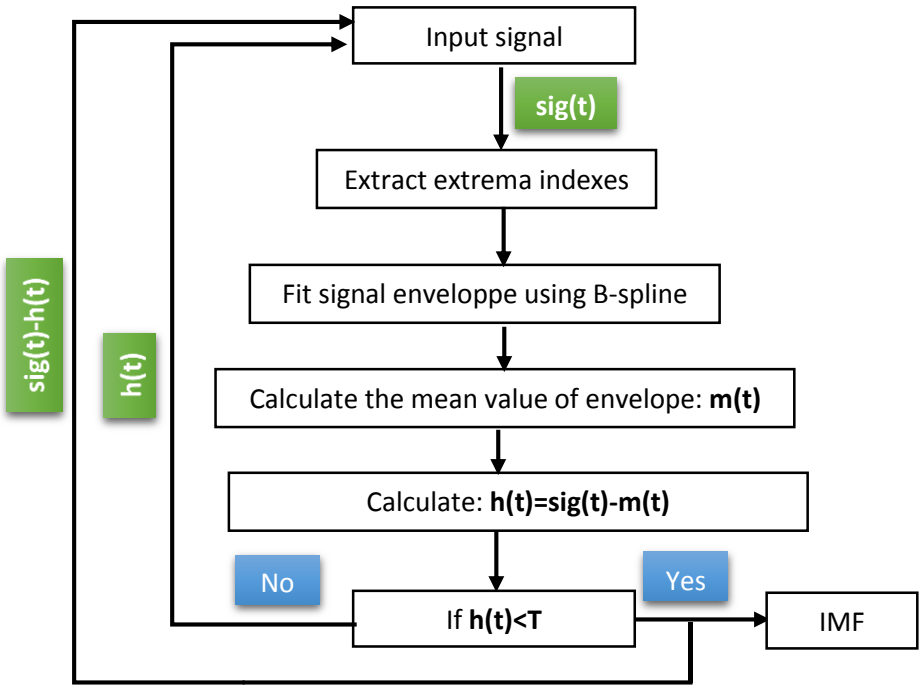


Figure 3. 6: EMD decomposition procedure

After decomposing a signal sig(t) through the EMD algorithm, the Hilbert transform is applied to each individual IMF. Each IMF is associated with its Hilbert transform by the corresponding formula:

$$HT_k(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{IMF_k(\tau)}{t-\tau} d\tau \qquad (3.43)$$

Hence, the analytic representation of the signal sig(t) is given by:

$$sig(t) = \sum_{k=1}^{n} IMF_k(t) + jHT_k(t) = \sum_{k=1}^{n} a_k(t)e^{j\theta_k(t)} + r_k(t) \quad (3.44)$$

Where $a_k(t)$ and $\theta_k(t)$ denotes the instantaneous amplitude and the instantaneous phase respectively for the IMF number k.

The Hilbert-Huang spectrum represents the amplitudes plot of all IMFs as a function of time and frequency [163].

3.4.2. Quadratics representations

Quadratic, bilinear, or Cohen's classes of time-frequency distributions are powerful tools for analysing non-stationary signals such as speech signal, ECG signal, and other biomedical ones. These representations are invariant to translation in the time-frequency plan. The quadratic, or Cohen's, class of time frequencies distributions could be formulated as:

$$C_x(t,f) = \iiint e^{j2\pi(\zeta t - \zeta f - f\tau)}\phi_x(\zeta,\tau)x(u+\tfrac{\tau}{2})x^*(u-\tfrac{\tau}{2})dud\tau d\zeta \quad (3.45)$$

3.4.2.1. Wigner Ville distribution (WVD)

The Cohen's class is a generic formulation of non-parametric time-frequency representations which include Wigner-Ville distributions [164].

The Wigner distribution in terms of the signal, s(t) or its spectrum, S(u), is:

$$\begin{cases} W_s(t,\omega) = \frac{1}{2\pi}\int s^*\left(t-\frac{1}{2}\tau\right)s\left(t+\frac{1}{2}\tau\right)e^{-j\tau\omega}d\tau \\ \quad = \frac{1}{2\pi}\int S^*\left(\omega-\frac{1}{2}\theta\right)S\left(\omega+\frac{1}{2}\theta\right)e^{-jt\theta}d\theta \end{cases} \quad (3.46)$$

The WVD presents a time-frequency resolution which is double of the one obtained for any linear transform, namely the Short Time Fourier Transform (STFT) [165].

Nonetheless, it suffers from severe interferences which are cross-terms. Cross-terms are areas of time-frequency energy density where there are false indications of the presence of signals activity in a given time and in certain frequency positions. Several reduced interference distributions were proposed [164] to overcome this major drawback and to enhance the representation of the signals. In fact, these proposed solutions deploy a smoothing kernel which decreases the influence of cross-terms at a cost of a given lower time-frequency resolution.

Some of the proposed distributions are the Smoothed Pseudo Affine Wigner Ville (SPAWD) and the Choi-Williams (CWD) Distributions. A particular interest is given to these distributions since they will be used to later analyse GPR signals.

3.4.2.2. Smoothed Pseudo Affine Wigner Ville distribution (SPAWD)

Cohen's class distributions are intended to obtain high quality expected properties like higher resolution and lower cross-terms density. This class of distribution is formulated by performing time and frequency shifting of the Wigner–Ville distribution with a kernel function. These kernels are low-pass filters in an ambiguity domain designed to smooth the WVD. Each specifically designed distribution corresponds to a different kernel formula: Bessel, Born–Jordan, Choi–Williams, and Zhao–Atlas–Marks distributions [166].
However, affine class of distributions is essentially time-scale functions with a Wigner–Ville distribution as a basis function.

The k-th pseudo affine Wigner distribution of a signal, s(t), PAWD can be formulated in terms of the Wavelet transform [167] of the signal, [167, 168] as:

$$PAWD^{k}(t,f) = \int \frac{\mu_{k(u)}}{\lambda_k(u)\lambda_k(-u)} E_w(t,a_+) * {E_w}^*(t,a_-)du \tag{3.47}$$

Where:

$$a_+ = \lambda_k(u)f, \quad a_- = \lambda_k(-u)f \tag{3.48}$$

In equation (3.47), the Wavelet transform is evaluated in time and scale. Since the scale parameter λ is a function of k, then k pseudo-affine distributions are generated.

For the Pseudo-Affine Wigner Distribution k=2 is settled [169].

The smoothed PAWD (SPAWD) is introduced by performing a smoothing in the frequency direction, as described in equation (4). In this case, the smoothing could be done by a frequency independent window such as the Gaussian window [167, 169]. In addition, when changing the range of the parameter μ, the smoothed PAWD allows a continuous transition between affine Wigner distributions and the Scalogram [167, 168]. Nevertheless, the smoothed distribution offers the higher resolution [164].

3.4.2.3. Choi-Williams distribution (CWD)

The Choi-Williams is a Cohen's class distribution with an exponential kernel.

The CWD of the signal s (t) is formulated as:

$$CWD(t,f) = \sqrt{\frac{2}{\pi}} \int \int_{-\infty}^{+\infty} \frac{\sigma}{|\tau|} e^{-\frac{2\sigma^2(u-t)^2}{\tau^2}} s\left(u+\frac{\tau}{2}\right) * s^*\left(u+\frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau du, \quad \sigma < 10 \qquad (3.49)$$

σ is a scaling factor, which controls the attenuation rate.

In fact, the smaller the value of the parameter σ is chosen the larger is the amount of smoothing and reduction of cross-terms interference. For a high value of σ, the CWD approaches the WVD [166].

CWD distribution does not meet some mathematical distribution properties for time and frequency. Otherwise, it diminishes the cross-terms interference of the WVD and provides a high time-frequency resolution.

3.5. Features Engineering

Feature engineering is the process of extracting and deploying domain knowledge of data to generate characteristics, which are used by machine learning algorithms to classify data. In this section, higher order statistics (HOS) cumulant features are described, as well as the feature selection algorithms, namely genetic algorithms and mutual information approach (MIFS).

3.5.1. Higher order statistics (HOS) cumulants

Let x(t) be a non-stationary signal. The second, third, and fourth-order cumulants [172, 173] are defined by the mathematical formulas:

$$\begin{cases} C_{2,x}(\tau_1) = E[x(t)x(t+\tau_1)] \\ C_{3,x}(\tau_1,\tau_2) = E[x(t)x(t+\tau_1)x(t+\tau_2)] \\ C_{4,x}(\tau_1,\tau_2\tau_3) = E[x(t)x(t+\tau_1)x(t+\tau_2)x(t+\tau_3)] \\ -E[x(t)x(t+\tau_1)]E[x(t+\tau_2)x(t+\tau_3)] \\ -E[x(t)x(t+\tau_2)]x(t+\tau_1)x(t+\tau_3)] \\ -E[x(t)x(t+\tau_3)x(t+\tau_1)x(t+\tau_2)] \end{cases} \qquad (3.50)$$

E[·],τ1, τ2, τ3 denote, respectively, the statistical expectation operator at the first, second and third time lag parameters. The second-order cumulants is the autocorrelation of the initial signal x(t).

By reducing the lag parameter to zero $\tau_1 = \tau_2 = \tau_3 = 0$ in equations (3.50) the variance $C_{2,x}(0)$, skewness $C_{3,x}(0,0)$and kurtosis $C_{4,x}(0,0,0)$ are calculated.

In general, cumulants are non-linear combinations of moments up to the desired order.

Some properties of cumulants that are the principal motivation to use these HOS features, are:

- Additivity propriety: If $S(\tau) = x(\tau) + y(\tau)$, where $x(\tau)$ and $y(\tau)$ are mutually independent processes, then $C_{kS}(\tau) = C_{kx}(\tau) + C_{ky}(\tau)$.
- Robustness against Gaussian noise: If $x(\tau)$ is Gaussian, then $C_{kx}(\tau) = 0$ for k > 2, starting from the third degree cumulants.
- Robustness against Gaussian coloured or altered noise: if $S(\tau) = x(\tau) + y(\tau)$, where $y(\tau)$ is Gaussian and independent of $x(\tau)$, then for k> 2, $C_{kS}(\tau) = C_{kx}(\tau)$. Additionally, it is very simple to recover higher-order cumulants of a non-Gaussian signal in the presence of coloured or altered Gaussian noise.

3.5.2. Feature selection

In machine learning, feature selection is an important problem to consider for obtaining the desired performance of models. Feature selection denotes choosing a subset of all available features so that the selected subset has the strongest relation to the model output. Two feature selection approaches are presented in this context: Genetic algorithms, and mutual information.

3.5.2.1. Genetic algorithms

Genetic algorithms (GA) is a heuristic randomized searching methodology which follows the idea of natural selection proposed by Darwin's theory [174].

The algorithm (figure 3.7) begins with a population of individuals encoded as chromosomes, which are simply a set of genes represented as bits. Individuals are selected according to their fitness value; individuals with higher fitness values have more chances of being selected. Crossover and mutation are principal operations for producing a new population. In fact, crossover accelerates the search process early by accomplishing the task of an evolution of the population. In comparison, mutation tries to restore the loss of information that could happen to the population through local or global movement introduced in the search space [4]. The algorithm is iteratively repeated several times until stopping criteria, related to fitness values, are met, or when an optimal set of the feature is found. Genetic algorithms outperformed other random and local search methods due to their capability to perform a search over subspaces where domain knowledge is inexistent.

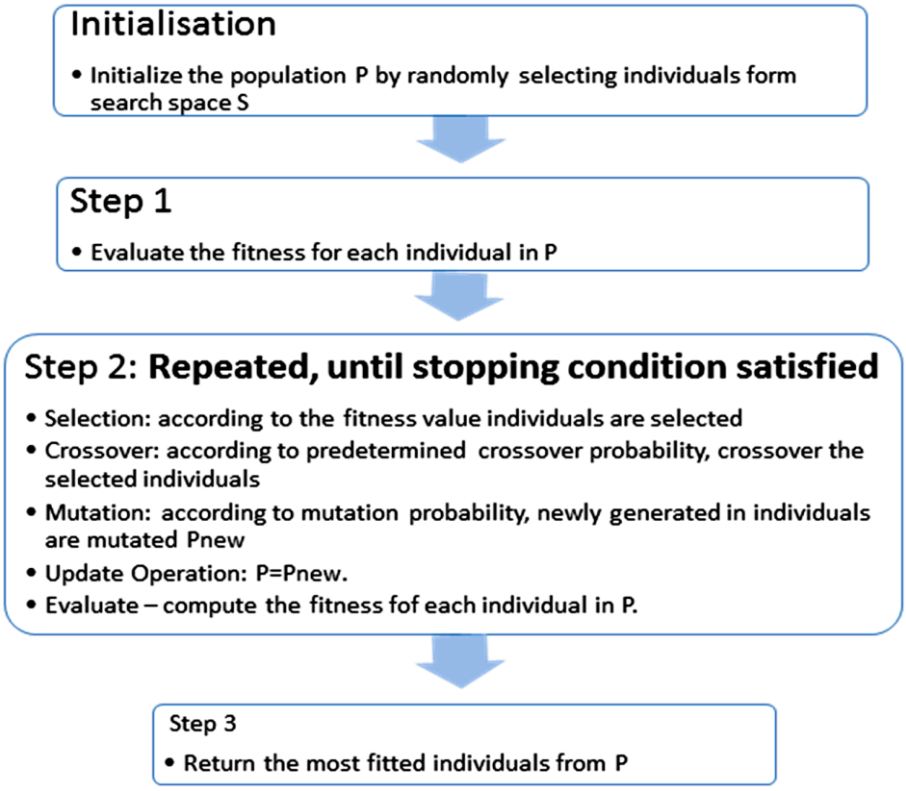


Figure 3. 7: Pseudo code for the GA based selection algorithm [4]

3.5.2.2. Mutual information approach

Mutual Information is a metric for quantifying the dependency of arbitrary variables, as well as the information contained, derived from the theory of information [175]. It is a simple tool to filter a feature while taking in consideration the degree of importance of information given in high dimensional classification problems.

MIFS algorithm, initially developed by [6], was chosen for a possible pre-feature selection task due to some motivating points:

- The use of MI, to filter features, widely reduce the uncertainty of output and the error of classification.
- This technique is completely independent from the considered classifier model and architecture.
- It takes into consideration the dependency between the set of given features
- Entropy and conditional entropy notions are defined with a direct relation to mutual information theory. Assume a classification problem where L is the set of output classes. The uncertainty in the classes will be quantified by the entropy measure:

$$H(l) \;=\; -\sum_{l\epsilon L} P_r(l)\, log P_r(l) \tag{3.51}$$

Where Pr (l) is the probability for class l ∈ L to occur.

By introducing a sample condition on feature vectors in the dataset, the uncertainty on the class, with a prior knowledge of a specific feature vector v ∈ V, is defined as a conditional entropy:

$$H(\boldsymbol{L}|V = \boldsymbol{v}) \;=\; -\sum_{l\epsilon L} P_r(l|\boldsymbol{v})\, log P_r(l|\boldsymbol{v}) \tag{3.52}$$

Knowing the probability distribution of the feature vector, the uncertainty becomes:

$$H(\boldsymbol{L}|V) \;=\; -\sum_{v\epsilon V} P_r(\boldsymbol{v})\, H(L|V = \boldsymbol{v}) \tag{3.53}$$

Therefore, Mutual Information is defined as:

$$MI(\boldsymbol{L}, V) = H(\boldsymbol{L}) - H(\boldsymbol{L}|V) \tag{3.54}$$

The MIFS algorithm for feature selection based on MI, introduced in [6], used in this work, is summarized in Figure 3.8.

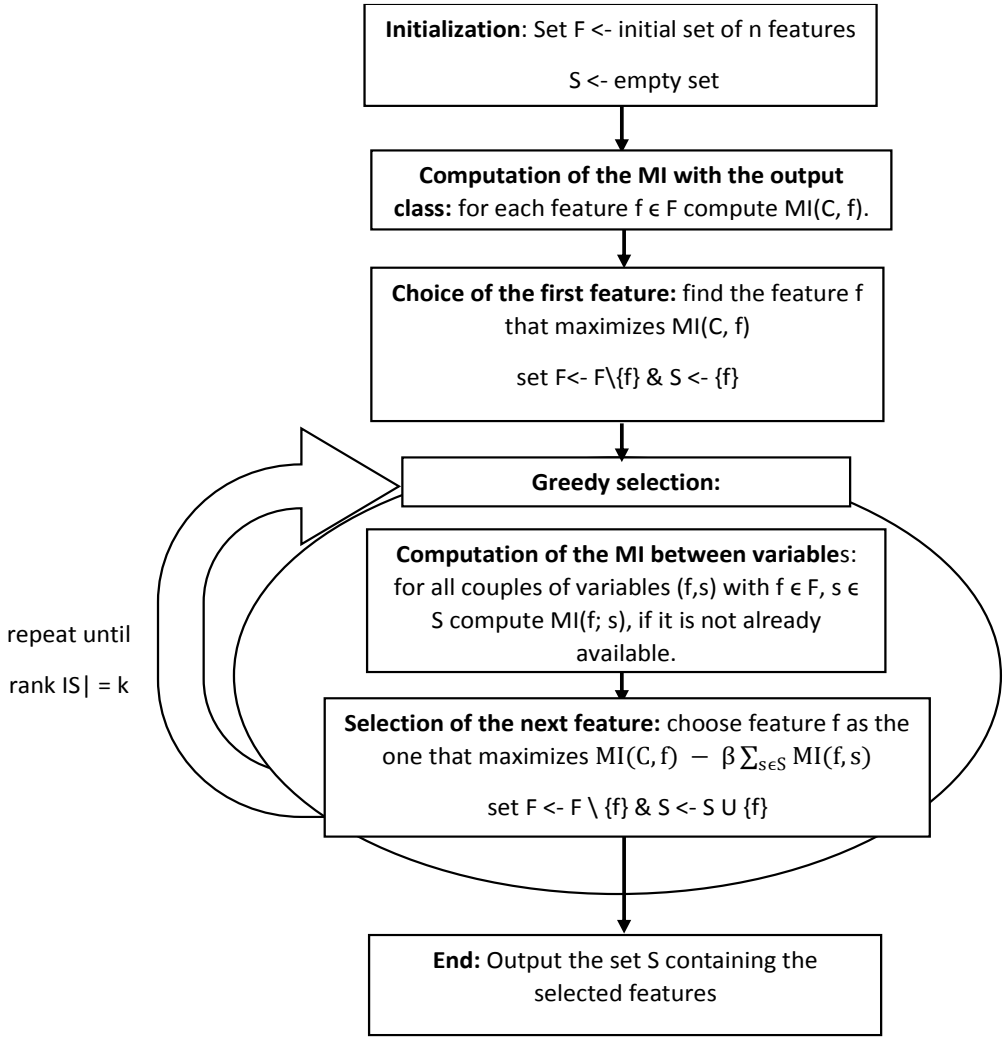


Figure 3. 8: MIFS algorithm [6]

The Parameter k represents the desired number of outputs, β is the regularization parameter. In practice, it is assumed that a value for β in the range [0.5 , 1] is appropriate for many ordinary classification tasks [6].

3.6. Classification Approaches

3.6.1. Neural networks

Neural Networks are models where the processing elements, denoted as neurons, are grouped in distinct layers: input, output, and middleware or hidden layers.

Different NNs architectures exist, but in this work we shall use Radial Basis Functions (RBFs) neural networks.

3.6.1.1. RBF neural networks architecture

RBF neural networks (figure 3.10) were first introduced by Broomhead and Lowe [176], where they were used as functional approximations for data modelling.
In contrast with other neural Multilayer Perceptron, only one hidden layer is used. The architecture of a RBF is shown in figure 3.9.

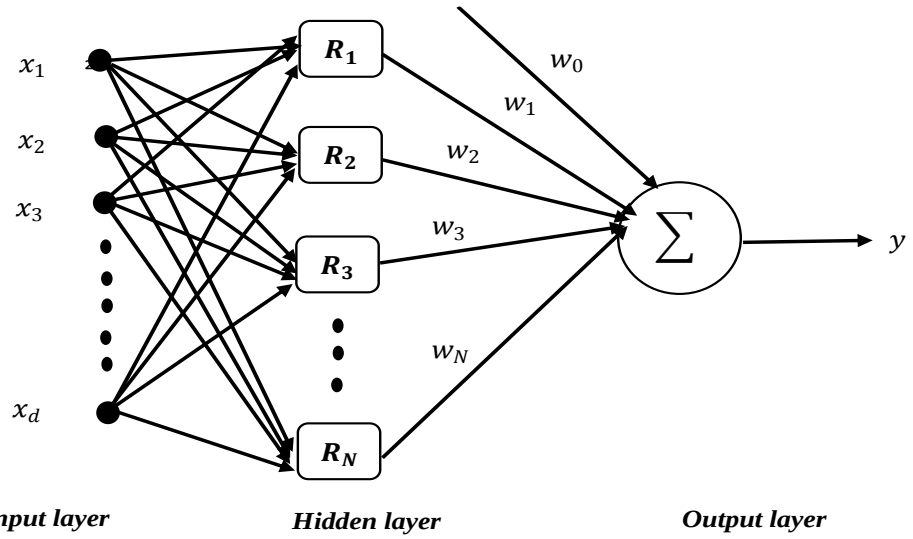


Figure 3. 9: RBF network architecture

The input of the network is a vector of real numbers $\boldsymbol{X} = [x_1, x_1, \ldots x_d]$, and the output y is a scalar function of the input vector given by:

$$y(X) = \sum_{i=1}^{N} \boldsymbol{w_i} R(\|\boldsymbol{X} - \boldsymbol{c_i}\|) \quad (3.55)$$

Where *N* is the number of neurons in a hidden layer, $\boldsymbol{w_i}$ and $\boldsymbol{c_i}$ are respectively the weights and the centers of the hidden neurons. *R* is the radial activation function.

The norm most used is the Euclidean norm and the radial basis function is usually a Gaussian:

$$R(\|\boldsymbol{X} - \boldsymbol{c_i}\|) = e^{-\frac{|X - c_i|_2^2}{2\sigma_i^2}} \quad (3.56)$$

Where σ_i is denoted as the spread parameter. The values of $\mathbf{w_i}$, $\boldsymbol{c_i}$, σ_i (collected in a parameter vector p) are determined by a training algorithm.

3.6.1.2. RBF learning strategies

There are essentially three learning strategies to estimate the parameters of an RBF neural network [177-180].

3.6.1.2.1. The first method

The first approach for centre selection consists in choosing center locations from the input data. The simplest procedure is to pick a desired number of centers at random, or selecting randomly the centers from the input data. However, if the distribution of the input data is not representative of the particular problem, typically large networks are needed for a satisfying performance, exhibiting asloa poor behaviour on unseen data.

Another procedure consists of choosing the center locations from input data, but based on a defined criterion. The orthogonal least squares (OLS) learning algorithm [181] selects a suitable set of locations from the set of input data.

In both approaches, the spreads are typically obtained using heuristics and the output linear weights are computed as a Least Squares solution.

3.6.1.2.2. The second method

The second method considered is a two stage hybrid learning approach. The selection of the center locations is carried out by a clustering algorithm. In this work, the optimal adaptive k-

means algorithm (OAKM) [182] is employed. The spreads of each activation function are then determined [183] using equation 3.67:

$$\sigma_i = \frac{d_{max}}{\sqrt{2n}}, i = 1, \dots, n \tag{3.57}$$

Where d_{max} is the maximum distance between the centers.

In the same way as the first method, the output linear weights are computed as a Least Squares solution.

3.6.1.2.3. The third method

In this approach, the center locations, the spreads, and the output linear weights are all determined under a supervised learning procedure. Although different methods are available, the following training algorithm is used in this work.

An improved version of the Levenberg-Marquardt (LM) algorithm [184, 185] for training individuals in each generation. This version of the LM algorithm [180, 186] takes into consideration the linear and non-linear separability of the neural network parameters. By employing this version of LM, the dimensionality of the problem is reduced, hence the computational time is reduced, and usually a higher accuracy and a fast convergence is obtained compared with other supervised methods.

The cost function to be minimized in this problem has the form:

$$\Omega(w) = \frac{1}{2}\sum_{i=1}^{m} e_i^2(w), w = [w_0, \dots, w_n] \tag{3.58}$$

m is the number of data samples and $\boldsymbol{e}_i(\mathbf{w})$ is the error of the network parameterized by the weights $\mathbf{w}$ while feeding the ith input pattern.

The output of the models can be represented as:

$$y = \phi(v)u \tag{3.59}$$

$\boldsymbol{\phi}$ represents the output matrix of the last nonlinear hidden layer, v represent the nonlinear parameters (centers and spreads) and u the linear parameters. When equation 3.59 is replaced in equation 3.58.

$$\Omega(w) = \frac{\|t-\phi(v)u\|_2^2}{2} \tag{3.60}$$

$\mathbf{t}$ is the vector of target values. For any value of $\boldsymbol{\upsilon}$, the minimum of cost function Ω with respect to $\mathbf{u}$ can be obtained using the least squares solution determined with application of pseudo-inverse:

$$\hat{u}(v) = \phi(v)^+ t \tag{3.61}$$

By replacing equation (3.61) in equation (3.60), a new criterion is computed, as shown in equation (3.62), which only depends on the nonlinear parameters.

$$\psi(\upsilon) = \|t - \phi(\upsilon)\phi(\upsilon)^{+}t\|_2^2 \tag{3.62}$$

To minimize equation (3.62), it`s gradient and Jacobian with respect to $\boldsymbol{\upsilon}$ can be determined by computing first the optimal value of the linear parameters (using equation 3.61), replacing this in the model and subsequently performing the usual calculation of the derivates. Using the criterion stated in equation (3.62) presents some advantages in comparison with the use of the standard criterion depicted in equation (3.60) [187]. It reduces the dimensionality of the problem, so when the Levenberg-Marquardt is used, the computational time is lower, and a smaller number of iterations is needed for convergence to a local minimum (i.e. it provides a faster rate of convergence).

3.6.1.3. Training criterion termination

The training process is terminated when a high-level accuracy is attended or before overfitting occurs.

The term, overfitting, refers to a special case when a model is too adjusted to the training data. In this case, it does not achieve a convenient generalization and will perform badly for the unseen data.

Three basic approaches are usually used to stop training:

- The first consists of fixing a maximum number of iterations for the training process. The main problem of this approach is that it does not take in consideration how well the parameters are adapted to the training data and how well it generalizes.
- For the second approach, it is necessary to check whether conditions that are mentioned in equation 3.63 are simultaneously met by the end of each iteration [187].

$$\begin{cases} \Omega[k-1] - \Omega[k] < \theta[k] \\ \left\| w[k-1] - w[k] < \sqrt{\tau}(1 + \|w[k]\|) \right\| \\ \|g[k]\| \leq \sqrt[3]{\tau}(1 + \Omega[k]) \end{cases} \tag{3.63}$$

τ corresponds to the number of desired correct digits in the training criterion, $k-1$ and k denote two consecutive iterations, while **w**, **g** and Ω refer to weights, gradient and cost function respectively. With small values of model is overfit within few training iterations.

- The third approach, namely early stopping, permits to avoid the overfitting situation.

For this situation, the dataset is split into two parts, training and test sets. The model is trained and by the end of each iteration, the cost function is evaluated for training and test sets. If the parameter updates are reducing the cost for both training and test data samples, the training process continues. Otherwise, if the cost of training data samples is still decreasing but the cost of test data samples is increasing, this means that the model is overfitting. At this point, early stopping will stop the training process.

3.6.2. CNN classifier

3.6.2.1. CNN architecture

A CNN is a deep learning approach, which is designed to take advantage of the 2D structure of an input image. The architecture consists of a number of convolutional layers fused with a subsampling process and then followed by a number of fully connected layers as observed in a MLP structure.

This network is connected through tied weights followed by some form of pooling which results in translation invariant features. The main benefit of CNNs is that they have fewer parameters than fully connected networks. This is as such since they are based on the principal of weights sharing, which means that weights and biases are the same for adjacent layers. In consequence, training the network is an easier task.

There are some main operations that are performed through many CNN architecture, and that needs to be defined:

- Convolution operation: Aims to extract features from the input image. Convolution stores the spatial relationship between pixels by performing learning features through small squares of input data.
- Non-Linearity: Rectified Linear Unit (ReLU) is a non-linear operation that usually follows the convolution stage. ReLU replaces all pixels with negative values in the feature map by zero in order to introduce non-linearity in the map.
- Pooling Operation: Pooling, subsampling or down-sampling, aims to reduce the dimensionality of the feature maps by filtering the most important features while depending on different pooling types. The basic ones are as followed: Max, Min, Average, and Sum.
- Fully Connected Layer: It is a simple MLP with a softmax activation function in the output layer. The purpose of such a stage is to classify the input image into various classes based on the features produced by the antecedent convolutional layers.

Figure 3.9 represents an example of the architecture VGG16 [188], with basic layers introduced before, and with the corresponding number and size of filters deployed for each layer.

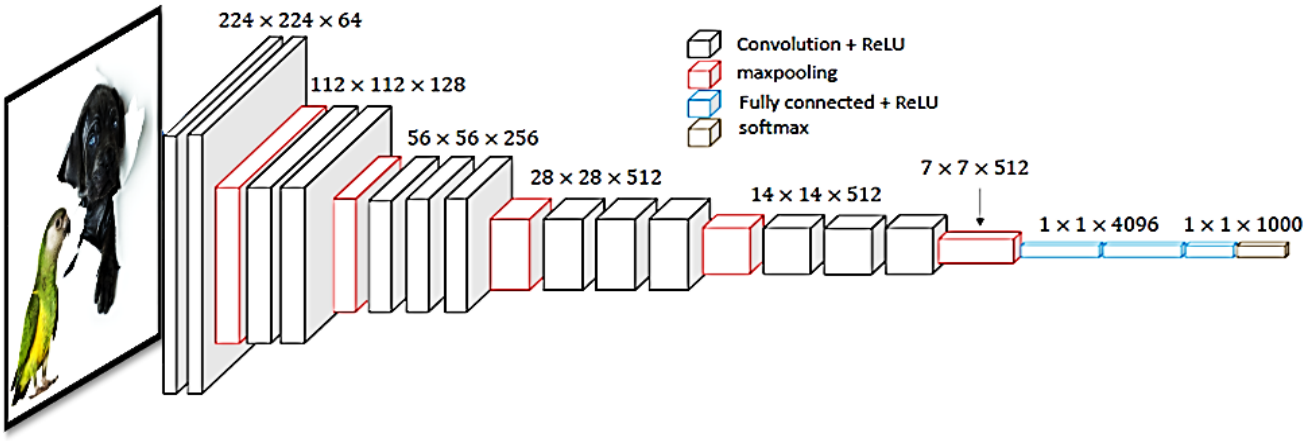


Figure 3. 10: VGG16 architecture [188]

3.6.2.2. Backpropagation training algorithm

The backpropagation algorithm used to train CNN is summarized in table 3.4.

Table 3. 4 : Backpropagation algorithm used for CNN training

	Backpropagation algorithm
Step 1	randomly initialize biases and weights arrays
Step 2	Network takes an image as input, and gives as output, the probabilities for each class, by a simple forward propagation process of the error. Let us have an example of the output probabilities for a two-class problem: [0.2, 0.1], this output probabilities are random, since the initialization, process of training was random.
Step 3	Calculate the total error at the last output of the network: $$\boldsymbol{Total\ Error} = \sum \frac{(\boldsymbol{target\ probability - output\ probability})^2}{2}$$
Step 4	Use Backpropagation to compute the gradients of the error for all weights and biases update them by gradient descendent formulas with new values with the aim to minimize the output error, until convergence of the process. Weights are updated depending on their contribution to the network error. The process converges when output probabilities will be something like [0.1, 0.7], which are close to the target labels [0, 1].
Step 5	**Repeat steps 2-4 with all the training images.**

3.6.3. SVM classifier

Support Vector Machine (SVM) is a classification technique that employs the concept of hyperplanes [189] in a multidimensional space that clusters data following the different labels introduced in the data.

To construct an adequate hyperplane, SVM employs an iterative training process to minimize an error function [189] defined according to the type of SVM used: SVM classifier type I, SVM classifier type II, SVM regression model type I, and SVM regression model type II.

- For SVM classifier type I:

$$error(x_i, y_i) = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n}\delta_i \tag{3.64}$$

Subject to $$y_i(w'K(x_i) + b) \geq 1 - \delta_i, \quad \delta_i \geq 0, \quad i = 1, \ldots . n \tag{3.65}$$

The parameter C is the capacity or the penalty parameter, w is the vector of weights, b is a constant, δ_i represent the slack variables, and n is the size of the training data. Besides x_i and y_i represent the input data and the class labels respectively.

- For SVM classifier type II:

$$error(x_i, y_i) = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} - m\varphi + \frac{1}{n}\sum_{i=1}^{n}\delta_i \tag{3.66}$$

Subject to $$y_i(w'K(x_i) + b) \geq \varphi - \delta_i, \quad \delta_i \geq 0, \quad i = 1, \ldots . n, \varphi > 0 \tag{3.67}$$

- For SVM regression model type I:

$$error(x_i, y_i) = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n}\delta_i + C\sum_{i=1}^{n}\delta_i^* \tag{3.68}$$

Subject to $$\begin{cases} w'K(x_i) + b - y_i \leq \delta_i^*{}_i \\ -w'K(x_i) - b + y_i \leq \delta_i \\ \delta_i, \delta_i^* \geq 0, \quad i = 1, \ldots . n \end{cases} \tag{3.69}$$

- For SVM regression model type II:

$$error(x_i, y_i) = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} - C\left(m\rho + \frac{1}{n}\sum_{i=1}^{n}(\delta_i + \delta_i^*)\right) \tag{3.70}$$

Subject to $$\begin{cases} w'K(x_i) + b - y_i \leq \varepsilon + \delta_i^*{}_i \\ -w'K(x_i) - b + y_i \leq \varepsilon + \delta_i \\ \delta_i, \delta_i^* \geq 0, \quad i = 1, \ldots . n, \ \ \rho > 0 \end{cases} \tag{3.71}$$

The kernel K is a function used to map the input data into the feature space by the transformation ρ. There are four types of kernels that are typically used:

$$K(x_i, x_j) = \rho(x_i)'\rho(x_j) = \begin{cases} Linear & x_i'x_j \\ Polynomial & (x_i'x_j + c)^d \\ Gaussian & e^{-\frac{|x_i - x_j|_2^2}{2\sigma^2}} \\ Sigmoid & tanh(x_i'x_j + c) \end{cases} \quad (3.72)$$

Where σ is the spread kernel parameter. The penalty parameter C should be determined, typically using a trial-and test method, for each data set.

The Gaussian, or RBF kernel, is the one chosen in this study because of its higher performance for non-linear mapping. Besides, an SVM with RBF kernel is operationally similar to a RBF neural network.

3.7. Multi-Objective Genetic Algorithm

3.7.1. Multi-objective evolutionary optimization

Many real-world applications are formulated as an optimization problem that involves more than one objective and are resolved through a multi-objective evolutionary optimization algorithm.

This process [5] could be drawn as the outcome of a mutual interaction between an artificial selector and an evolutionary search algorithm:

- The evolutionary search algorithm reproduces a new set of possible solutions in consonance with the fitness value given to the current set of solutions by the artificial selector.
- The artificial selector revaluates again the current solution and redefines new preferences.
- The evolutionary search algorithm considers the recently acquired changes and applies necessary corrections.

The artificial selector block is a simple cost assignment strategy, and the evolutionary search algorithm block is a search algorithm for finding strong solutions for the problem to optimize [190].

3.7.2. MOGA algorithm

3.7.2.1. RBF optimization problem in MOGA

MOGA is a framework developed in Matlab, C and Python. It is based on a genetic algorithm approach for solving problems with several objectives to be optimized. In this context, MOGA is used for the design of an optimized RBF network architecture.

In MOGA a classification problem using a RBF network could be formulated as:

$$min\ \varepsilon_k(F, \mathrm{i}, n, w)_{i,n,k=1,..,5} \tag{3.73}$$

Where ε_k, k=1,..,5 corresponds to the objectives functions to be minimized, given the features vector F, the range of the number of features $d \in [dmin, dmax]$ and the range of number of neurons $n \in [nmin, nmax]$.

The features vector is an $S_d \times$ Nb_F matrix with S_d patterns (size of the database) and Nb_F features computed for every pattern.

The variable i corresponds to the indices of features in F, and w refers to the RBF parameters that minimize the mapping error.

The objectives functions ε_k, k=1,..,5 used for this type of problem are the network complexity and the false positives (FP) and negatives (FN) for training (tr) and testing(tt), i.e., (μ, FP_{tr} FN_{tr} FP_{tt} FN_{tt}).

The complexity of the model can be defined as:

$$\mu = (d + 1) * n \tag{3.74}$$

Where d is the number of input features, and n the number of neurons in the hidden layer.

Each potential solution of the problem is encoded as a chromosome. As any genetic algorithm based process, selection, mutation crossover represent the key operations for individuals' reproduction.

As so, MOGA, introduced by Fonseca and Fleming [5], uses a Pareto-ranking based approach, described after, to establish the survival condition of individuals in the population, which is necessary for a selection process.

The non-dominated set, also called the Pareto-optimal set, represents the solutions that could not be improved without having a critical effect on other objectives. This is typically the output of MOGA (summarized in Figure 3.11).

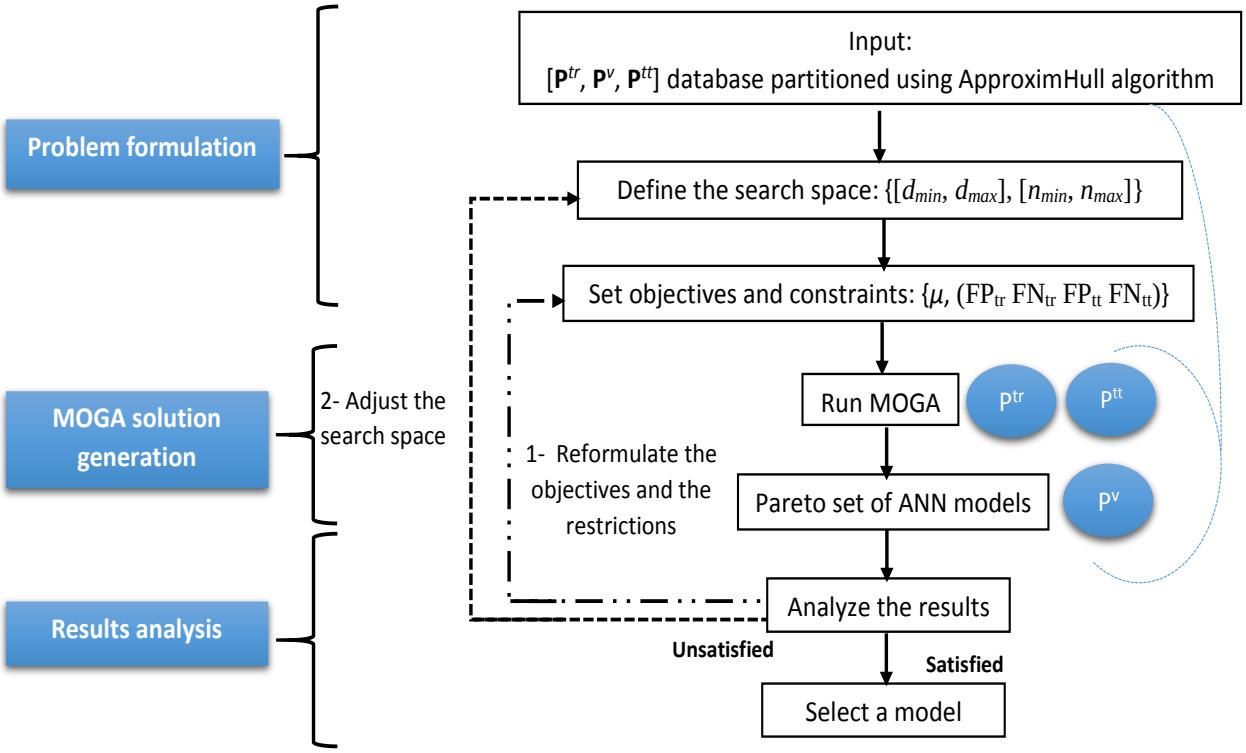


Figure 3. 11: MOGA approach [186]

If goals and/or preferences are assigned to the objectives, preferable individuals from the non-dominated set can be obtained, which decreases the number of possible solution, from where the user must select a model.

The algorithm consists of three major steps: problem definition, solution generation, and an analysis of results. [$\mathbf{P}^{tr}$, $\mathbf{P}^{v}$, $\mathbf{P}^{tt}$] represents the data, partitioned using AproxHull [191] strategy, given to MOGA for generating RBF models. Every set (training, validation, and testing) is supplied as an $S_d \times Nb\_F$ matrix with S_d patterns (size of the database) and Nb_F features. The first $\mathbf{P}^{tr}$ data set is used by MOGA for RBF ANN parameter estimation, using the modified LM algorithm described before. The $\mathbf{P}^{tt}$ second set is used either to stop the training process by using an early stopping criterion. The last $\mathbf{P}^{v}$ set is used to assess the performance of generated models after the execution of MOGA.

The training phase is repeated an user-defined number of times because the results of such a gradient-based training process depend on the parameters of initialization. As so, MOGA proposes several options for choosing the best training results (please refer to the example of figure 3.12).

The first one consists of selecting the training trial that minimizes all objectives in a better way (**Best compromise trial**). For the second option, the chosen trial is the one whose value is the closest one to the average of objective values for all training trails (**Trial closet to average**). Two other options are proposed too. They consist of selecting the trial that minimizes better one of the objectives (**Best trial minimize objective 1/objective 2**) [192].
For the example presented in figure 3.12, the number of trials, NT, is set to 10.

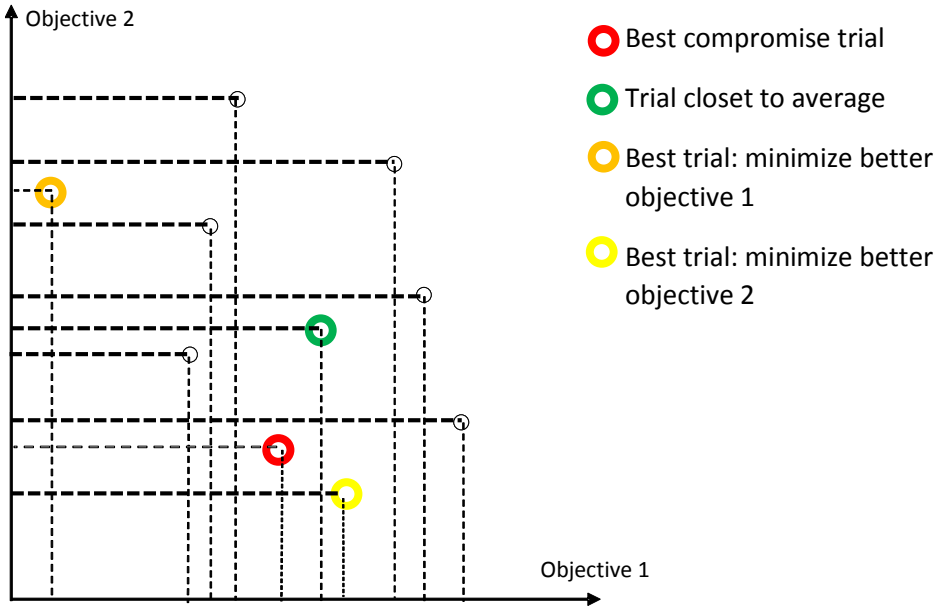


Figure 3. 12: An example of the selection strategies proposed by MOGA to identify the best trial in a set of $NT = 10$ trails [192]

The training termination for MOGA gives two options: maximum number of iterations specified by user or early-stopping procedure (the first and the third approach) [183].

3.7.2.2. Pareto ranking approach

The Pareto ranking approach used by MOGA is based on the principle of dominance of the solutions among each other while also taking into account restrictions and priorities.

Three situations are presented: (1) there are no restrictions and objectives that have the same priorities, (2) restrictions are present and objectives have the same priorities, (3) restrictions are present and some objectives are prioritized over others.

In the following, multi-objective cases will be reduced to two objectives approaches in order to demonstrate the ranking strategy with example samples. Figure 3.13 represents the same example where solutions are ranked depending on studied cases 1, 2 or 3.

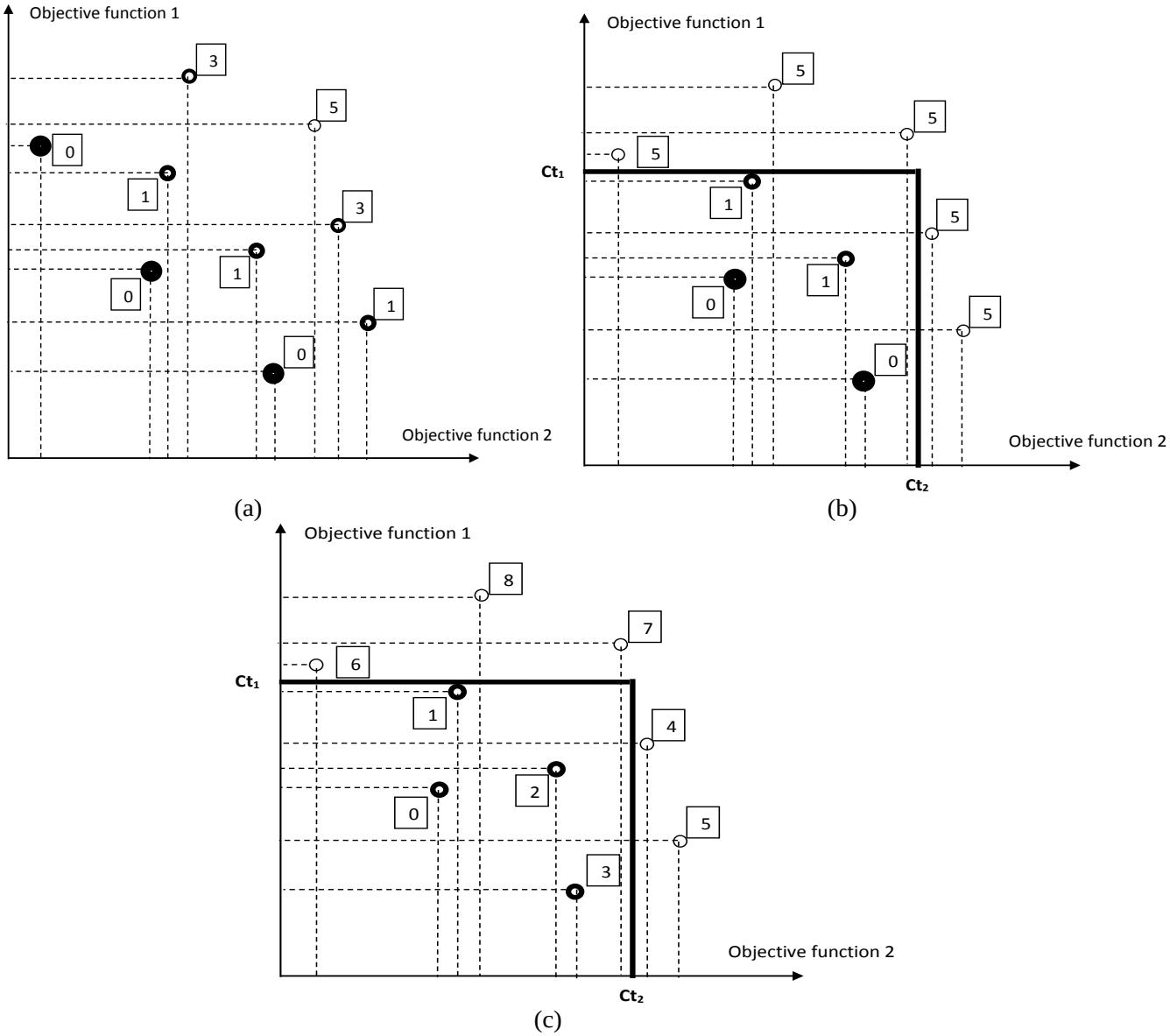


Figure 3. 13 : An example of Pareto ranking based approach (Fon+seca and al., 1995) for the three cases presented: a) for case 1, b) for case 2, and c) for case 3, Ct_1 and Ct2 are constraints on objectives one and two, respectively.

Case 1:

Step 1: Assign rank 0 to first set of non-dominated solutions

Step 2: Remove rank 0 solutions from contention

Step 3: Find next set of non-dominated solutions, remove them from contention, and assign rank equal to the number of n solution's dominating individuals

Step 4: Continue the process until the entire population is attributed a rank.

Case 2:
Step 1: Assign rank 0 to first set of non-dominated solutions that meet all defined constraints
Step 2: Remove rank 0 solutions from contention
Step 3: Find next set of non-dominated solutions that meet all defined constraints, remove them from contention, and assign rank equal to the number of n solution's dominating individuals
Step 4: Continue the process for all individuals of population that meet all defined constraints are attributed a rank.
Step 5: The rest of solutions that does not meet constraints are penalized by the higher rank.

Case 3:
Step 1: Find set of non-dominated solutions that meet prioritized constraints
Step 2: Assign rank depending on degree of satisfaction of other non-prioritized constraints
Step 3: The rest of solutions that does not meet prioritized constraints are penalized by a higher rank, depending on how far they are from the prioritized constraints values

After ranking the population, a fitness value must be assigned to select the best individuals for the new generation. The original rank-based fitness assignment procedure is slightly modified. It is summarized as follows:

- Step 1: Sort individuals by rank.
- Step 2: Assign fitness by interpolating from the best to the worst individuals according to an exponential function.
- Step 3: Assign average value of fitness to individuals with the same rank, so that they will have the same chance to be selected for the next generation.
- Besides, sharing and mating restriction processes were introduced in the algorithm in order to promote and maintain diversity in the population [5].

3.7.2.3. Approximate convex hull (AproxHull) data partitioning algorithm

In machine learning and data mining problems, two basic tasks have to be considered: feature selection and instance selection. The former is discussed in section 3.5.2. The latter refers to sample selection in which it is interesting to choose a subset of useful and informative data samples (denoted as S) among all existing data samples (denoted as D).The model obtained by using S usually exceeds the performance level that would be attained by using D. Using S, the runtime of the training process was decreased and the memory requirements of the learning

algorithm were reduced. MOGA uses an instance selection algorithm (or data partitioning algorithm), named AproxHull, that relies on the notion of convex hull points.

From a mathematical point of view, an object in a Euclidean space is characterized as convex when, for each pair of points within the object, every point on the straight line segment that joins them is also within the object.

A set $\boldsymbol{S}$ is convex if the relation is verified:

$$\forall u, v \epsilon S \; and \; t \epsilon [0,1], \quad \big((1-t)u + tv\big) \epsilon S \tag{3.75}$$

Moreover, if $\boldsymbol{S}$ is a convex set, and the relation is satisfied:

$$\forall u_1, u_2, \dots, u_r \epsilon S \; and \; \forall \frac{\{\lambda_1, \lambda_2, \dots, \lambda_r\}}{\lambda_i} < 0, \qquad \Sigma \lambda_i u_i = 1 \tag{3.76}$$

The vector $\Sigma \lambda_i u_i = 1$ is called a convex combination of {u_1,u_2,..., u_r}. According to the definitions above, the convex envelope of set **E** could be interpreted in terms of convex sets as follows [7, 191]: the minimal convex set containing **E**, or the intersection of all convex sets containing **E**, or even the set of all convex combinations of points in **E**.

Let`s consider all data samples residing on the hull of the convex set as a convex hull of set **E**. Each data sample within this hull is named a convex point or a convex vertex. Facets produce the connection between vertices. The dimension of the facet is equal to the dimension of the dataset. In fact, in a two-dimensional space, convex points are connected to each other within lines (two dimensional facets), but in a three dimensional space, convex points are connected through planes (three dimensional facets).

The AproxHull algorithm proposed by Khosravani, Ruano and Ferreira [8] is resumed in table 3.5.

Table 3. 5 : AproxHull algorithm [8]

Step	Process
Step 1	Data is scaled to the range [-1, 1]
Step 2	Identification of the maximum and minimum samples with respect to each dimension
Step 3	Generation of a population of $\boldsymbol{k}$ facets based on current vertices
Step 4	Identification of furthest points to each facet in the current population as new vertices of convex hull, if they have not been detected before.
Step 5	○ Update of convex hull by adding newly found vertices into current set of vertices. ○ Steps 3 to 5 are executed iteratively until one of the following two termination criteria is met: ▪ There are no newly found vertices in Step 4 ▪ Let $\boldsymbol{d}$ be the maximum of approximated distances of furthest points to the current convex hull in each iteration. If there are new vertices after execution of Step 4 and the difference between the maximum and minimum of $\boldsymbol{d}$ over $\boldsymbol{w}$ last iterations is less than a threshold, and there is fluctuation in value of $\boldsymbol{d}$ in this $\boldsymbol{w}$-sliding window.

3.8. Conclusions

In this chapter, all the background materials employed with the algorithms deployed for the feature extraction stage, of the designed classifiers, are presented. We first discussed the general FDTD discretization scheme deployed for synthesized data generation. Afterwards, an overview of the employed tools for data inversion was presented, including Hough transform, time-frequency representations, and machine learning algorithms.

4. GPR Data Inversion

4.1. Introduction

Due to the heavy computational burden of the Hough transform and the issues relevant to this fact, the Hough algorithm return random results. In order to correct this behaviour of the algorithm, it is necessary to reduce the dimensionality of the accumulator or of the resolution of images. When reducing the image resolution, apexes/vertex positions are affected since the pixel around this region is recalculated by interpolation. Thus, the practical solution is to reduce the accumulator dimensionality. The case of hyperbola identification requires a 3D accumulator, however, the proposed algorithm requires only a 1D accumulator array. It consists of three main steps: pre-processing, vertex localization, and the so-called 1D Hough transform. The pre-processing step, which requires an adequate choice of filtering and denoising parameters, is the core phase of this technique. Otherwise, a wrong choice of these parameters affect the results of the approach. Moreover, the velocity of the wave computed with a Hough transform, in general, is a mean value between the celerity and the velocity of a surrounding medium. In order to reconstruct the real velocity within the medium, a time-frequency approach is adopted. In this chapter, a detailed presentation of this technique coupled with suitable examples of different occurring cases is presented. Several tests using synthesized data (by FDTD) [38, 189, 190] has been done and are hereby presented. This chapter also includes tests with real cases, this is, using experimental data.

The outline of this Chapter is as follows: In Section 4.2, the modified Hough Transform technique is introduced. The three main phases are introduced in sections 4.2.2, 4.2.3 and 4.2.3. The time-frequency study completed to correct the velocity values and define the material type of targets is presented in section 4.3. This study englobes the results given by quadratic distributions and the non-quadratic ones that are explained in section 4.3.1 and 4.3.2, respectively, while a comparison of the obtained results is done in section 4.3.3. Conclusions are drawn in section 4.4.

4.2. Modified Hough Transform Technique

5.2.1. Proposed algorithm

Hough transform is a well-used technique for pattern recognition and extraction. Initially, it was used for detecting regular curves, but then after expanding its function, it become able to isolate other types of curves such as hyperbolas. Despite the simple cases, a hyperbola case requires a 3D space to implement the algorithm. Thus, three main shortcomings are highlighted:

- Problems residing in resolution limitation occur, and therefore are restricting algorithm accuracy for targets located in noisy environments.
- The algorithm, with the larger resolution of the B-scan images is time consuming. In addition, it produces a lot of false detections, or to be more precise, random output results.
- Vote spreading and peak splitting created issues such as making the true position of the peaks to be locally blurred.

A modified approach with only a 1D accumulator array is proposed (please see figure 4.1).

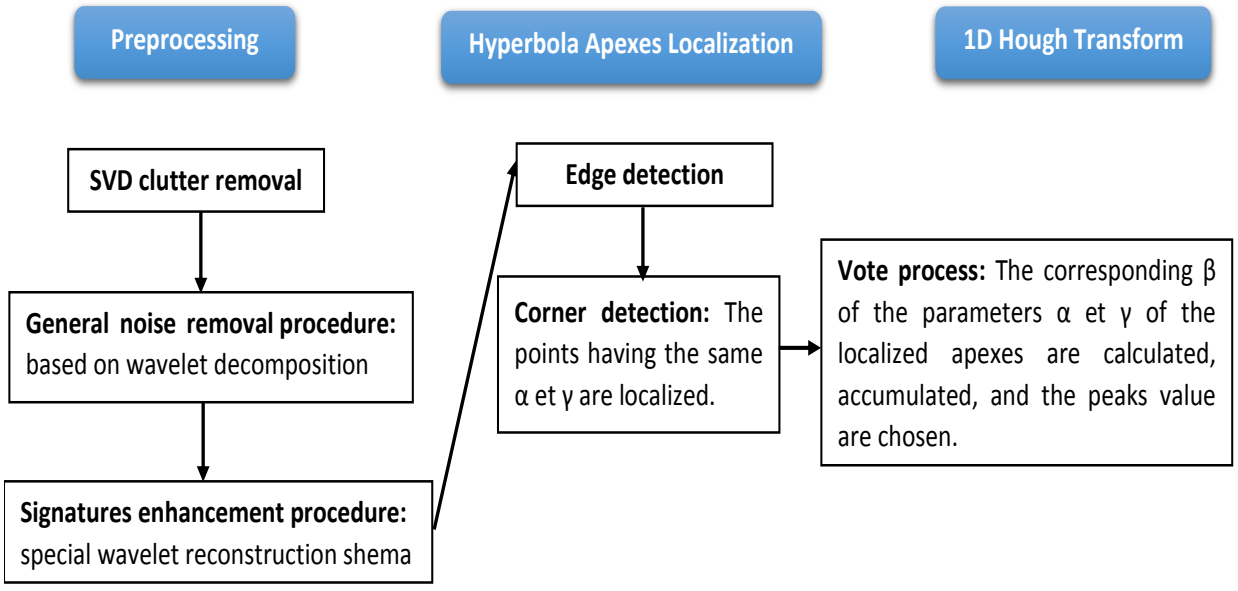


Figure 4. 1: A modified Hough Transform: consists of three stages, pre-processing, Apexes localization, and 1D Hough transform.

Divided into three main stages, the algorithm gives special consideration to the preprocessing step. The preprocessing step involves an SVD clutter removing technique as well as a wavelet denoising approach with special reconstruction aiming to enhance the hyperbolas.

In the second stage, a first round of search is done to find the possible locations of hyperbola apexes. In the third stage, the previously selected pixels are then filtered by the 1D Hough transform providing the exact locations of apexes.

The following paragraphs provide insights into the strategy.

5.2.2. Pre-processing

The pre-processing phase is the core of the introduced algorithm.

4.2.2.1. SVD clutter removal

On B-scans, the clutter is a non-attenuated signal characterized as horizontal bands while buried objects are represented by hyperbolas.

The clutter usually appears as three basic bands: The first band contains the peak caused by the coupling of GPR antennas recording various measuring points, while the two other bands are due to the reflections on the air-ground interface. Another component belonging to the clutter noise is represented by the small reflections recorded in depth, usually less than one meter. These reflexions are due to small objects and underground installations.

The challenging points here are that the reduction of the clutter algorithm must be able to eliminate the clutter while preserving the hyperbolic signatures of the objects being detected.

SVD is a statistical method capable of decomposing and reducing any set of data into a series of components that describe the main properties. This technique is used on B-scans to separate the signals from clutter.

Suppose $B = \{b_{ij}\} \in \mathbb{R}^{N_t x N_a}$ is a B-Scan image, resulting from the concatenation of N_a A-Scans, having N_t time samples. The SVD of this B-Scan, in its mathematical formulation, is given by the equation:

$$B = USV^T = \sum_{i=1}^{r} \gamma_i u_i {v_i}^T \quad (4.1)$$

$U = [u_1 \cdots u_{N_t}] \in \mathbb{R}^{N_t x N_t}$ and $V = [v_1 \cdots v_{N_a}] \in \mathbb{R}^{N_a x N_a}$ are two unitary matrices, and S is a diagonal matrix.

The singular values γ_I verify the condition: $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_r > 0$, while r is the rank of the B-Scan (usually equal to N_a).

The product $u_i v_i^T$ represents the singular B-Scan number i. This singular B-scan describes the behaviour of the waveform contained in the vector u_i for different positions of the GPR, while the vector v_i gathers the spatial information.

The singular values γ_i can be interpreted in terms of energy of the singular B-scan number i, the relative energy content of each singular B-Scan i, denoted by ξ_i, is given by:

$$\xi_i = \frac{\gamma_i}{\xi} \tag{4.2}$$

ξ is the global energy given by the following equation (4.3):

$$\xi = \sum_{i=1}^{N_t} \sum_{j=1}^{N_x} {y_{ij}}^2 = \sum_{i=1}^{r} {\gamma_i}^2 \tag{4.3}$$

The singular values γ_i are sorted in a descending order, thus the first singular B-scans captures most of the overall energy quantity.

SVD provides the basis orthogonal by maximizing the relative energy γi in the least squares sense [84, 191]. This reflects the fact that the first singular B-scans contain a maximum energy and provides the best correlated structure in time and space, while the last singular B-scans capture the random structures.

The B-scan B, seen as the sum of an informative signal and clutter, can be decomposed into a sum of several orthogonal matrices, given by SVD:

$$B = S + C = \sum_{i=1}^{r} \gamma_i u_i {v_i}^T + \sum_{i=r+1}^{n} \gamma_i u_i {v_i}^T \tag{4.4}$$

The challenge faced at this point is to make an adequate choice of the ranks r and n. In fact, the best approximation, in the least squares sense, is given by the truncation of the SVD.

The B-scan shown in Figure 4.3.a., synthesized by the FDTD technique, is decomposed into a sum of several singular B-scans according to the equation (4.1). The first six singular B-scans of this decomposition are shown in Figure 4.2, and the distribution of Amplitude of singular values is given in Figure 4.3.

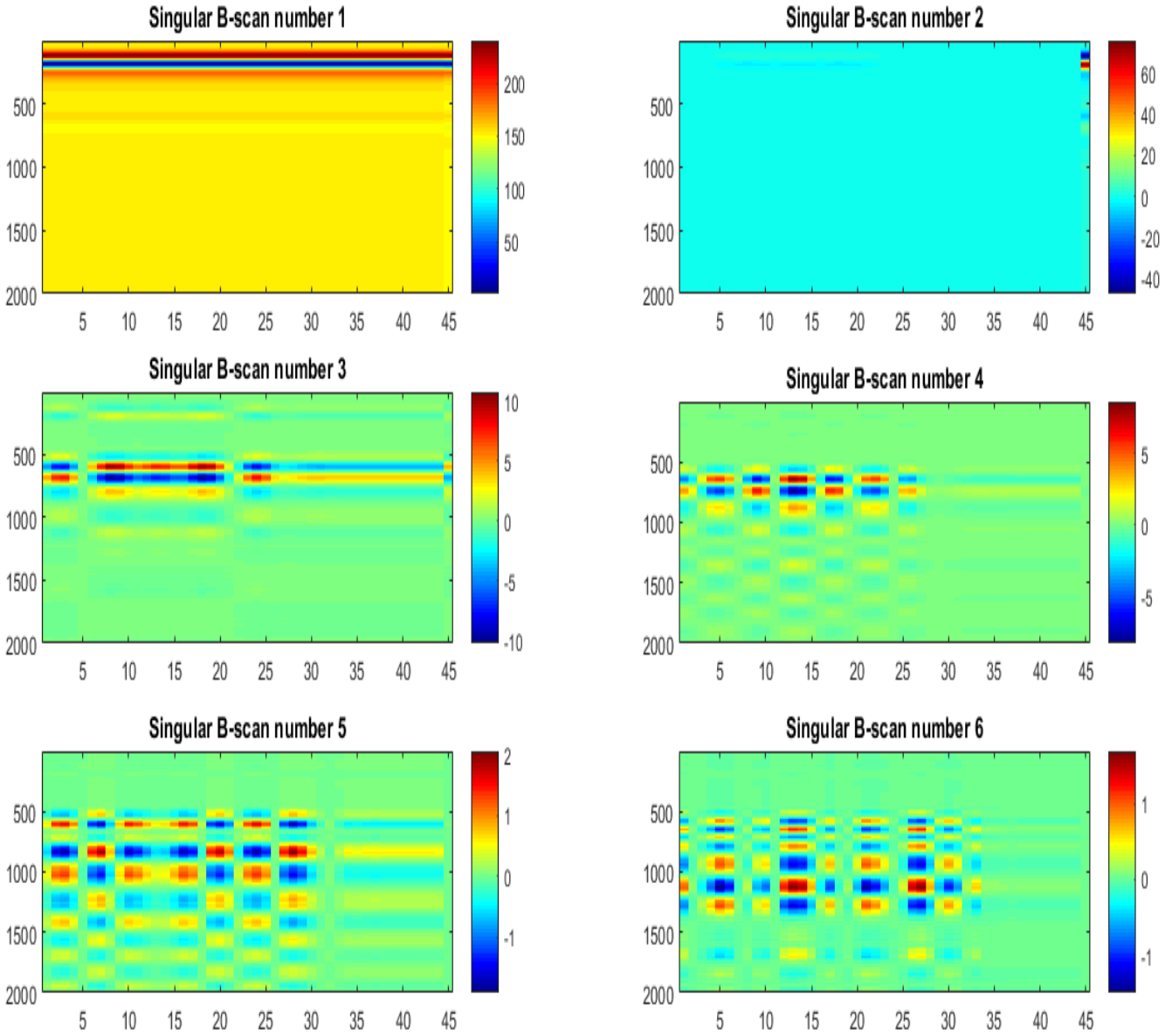


Figure 4. 2: The six first singular B-scans corresponding to the decomposition of B-scan in figure 3.4.a

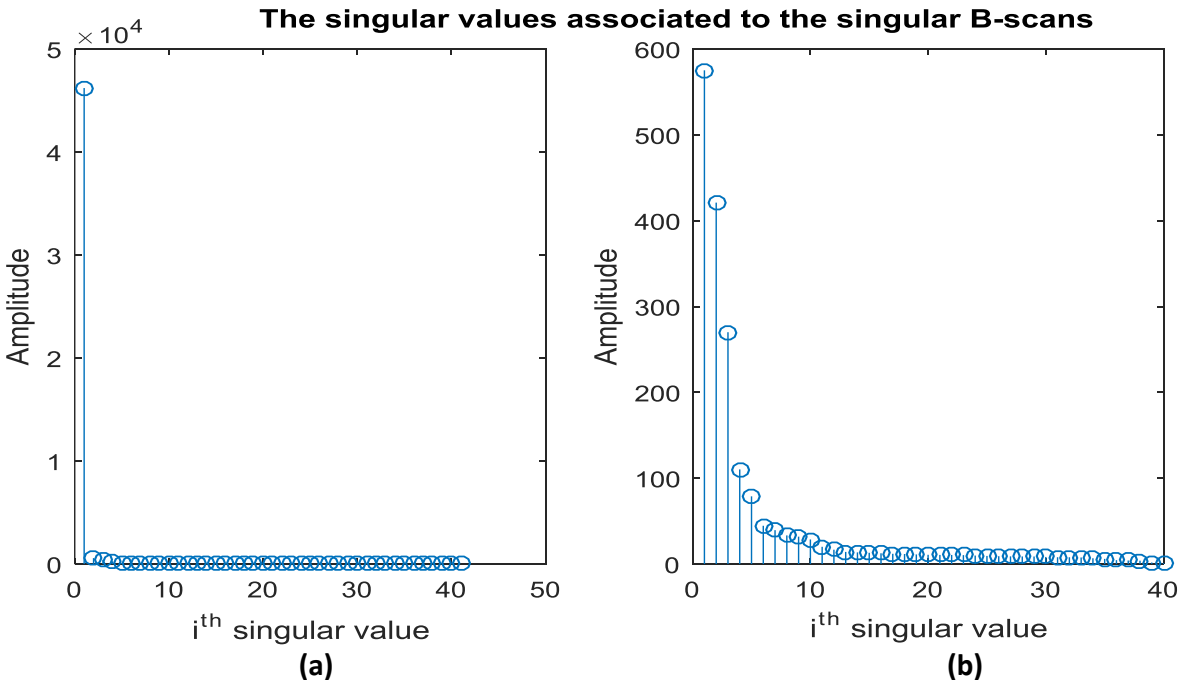


Figure 4. 3: The singular values distribution corresponding to the singular B-scans: (a) The clutter is the dominant component (b) the distribution after discarding the first singular value

It is perceptibly expressed (figure 4.2) that the first singular B-scan is constituted by the most correlated elements. This singular B-scan contains the three horizontal bands resulting from the coupling between the antennas of the GPR as well as the reflection from the air-ground interface. The second singular B-scan also contains mainly the elements of the clutter.

It is visually interpreted (figure 4.3) that the singular value associated with the first singular B-scan is largely higher than other values which can be expressed in terms of energy by the formula (4.2). The first singular B-scan, shown in figure 4.2, contains by itself 99.97% of the total energy of the B-scan B. It does not contain any clear signatures, and the same is true for the second singular B-scan, despite the fact that the second one does not contain a big amount of energy. These two singular B-scans can therefore constitute an estimate of the clutter C.

The SVD decomposition procedure was tested on experimental data as well, provided as grayscale images. An example of the decomposition is shown in figure 4.4 and the estimate of the clutter is presented in figure 4.5.c. For this case, the first singular value stores a large amount of energy representing 94.07% of the total energy of the B-scan. Thus, it could form on its own an estimate of the clutter C.

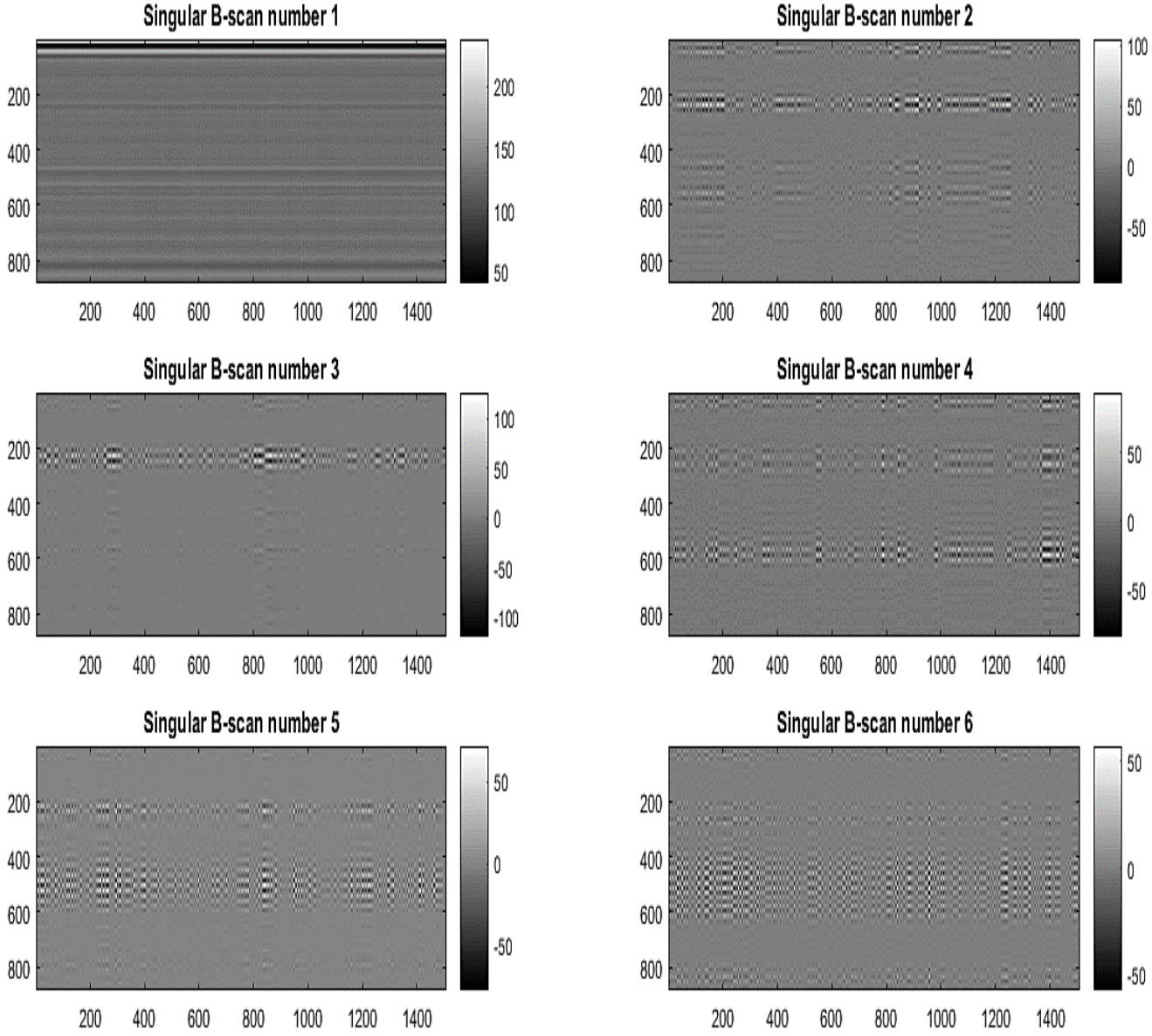


Figure 4. 4: The first six singular B-scans corresponding to the decomposition of an experimental B-scan acquired in the field (figure 3.6.a).

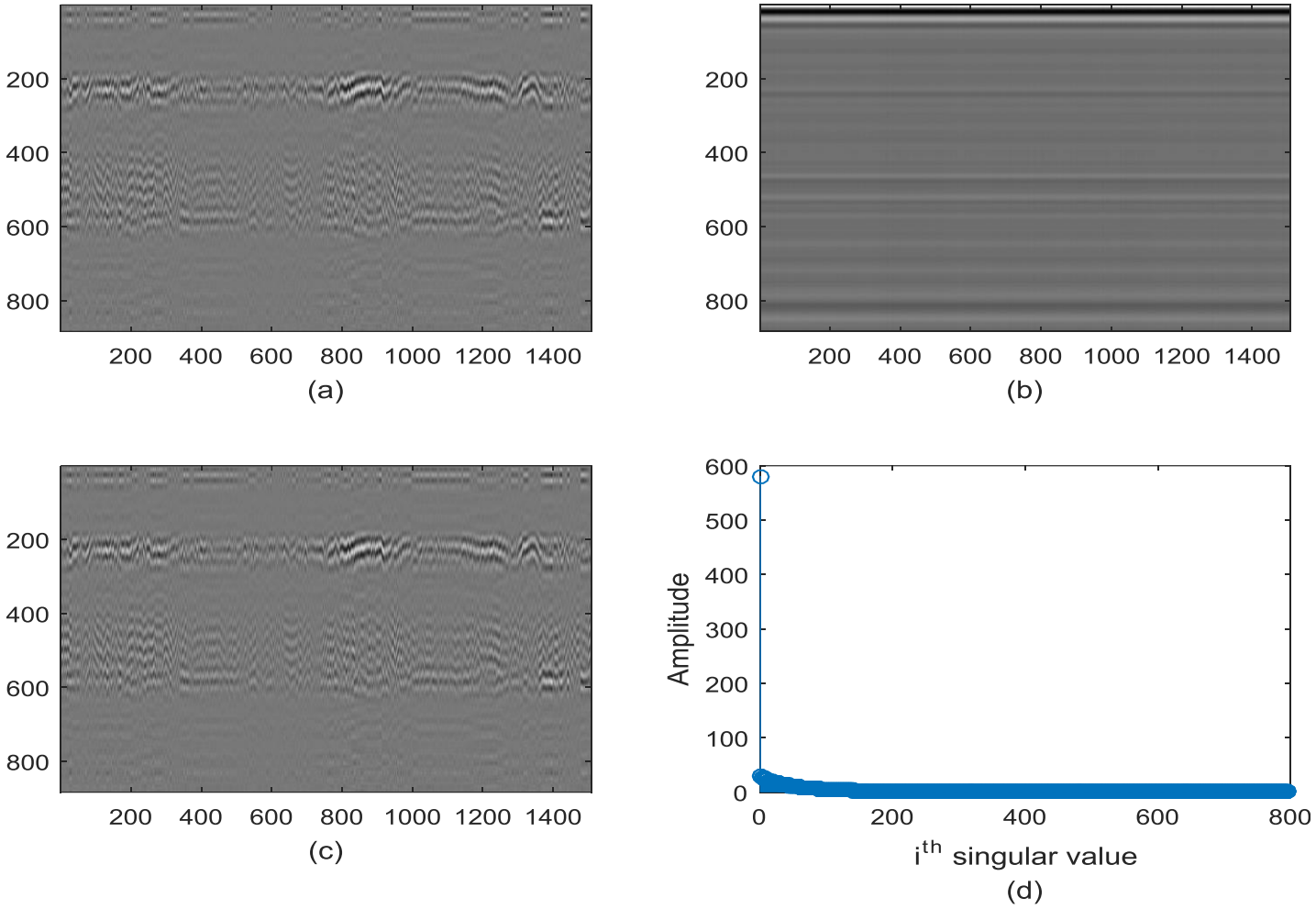


Figure 4. 5: (a) An experimental GPR radargram, (b) The estimate of clutter, (c) The B-scan after clutter removal by SVD, (d) The distribution of amplitude of singular values.

In general, a B-scan B could be rebuilt from a few singular B-scans. For example, the B-scan data of figure 4.6.a can be reconstructed from the sum of the first six singular B-scans already constituting 100% of the total response in terms of energy. Please refer to Figure 4.6.b and 4.6.c.

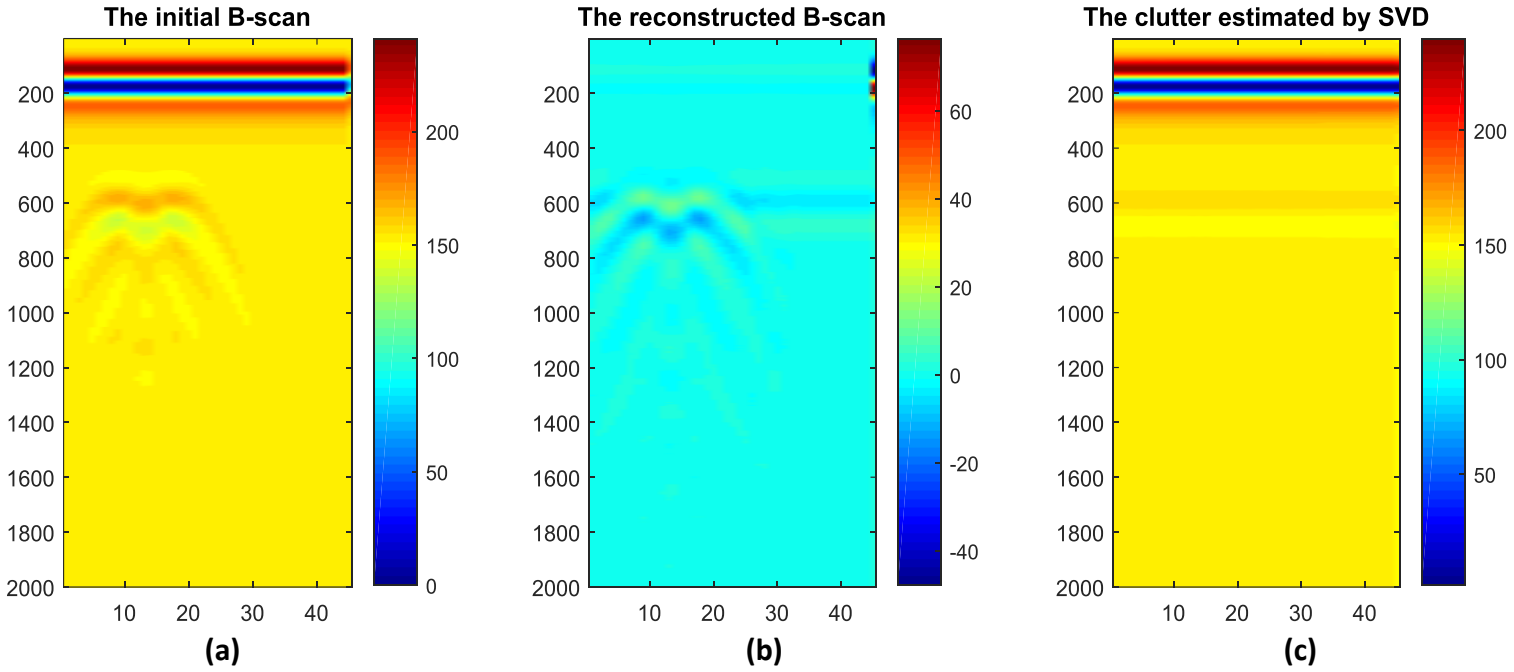


Figure 4. 6: (a) a synthetic B-scan , (b) B-scan after clutter removal, (c)The estimate of clutter by SVD

The clutter is notably reduced since the three bands on the radargram are almost eliminated, and only some residues are present for the last A-scans traces. The target responses were not affected by significant distortions and they are predominant in the total response.

Regarding the choice of the parameter r, it can be selected so that the sum of the relative energies of the r first singular B-scans represent about 99% of the global energy ξ.

This value was determined experimentally from the analysis of several radargrams within the same measurement scenario. It is, however, more difficult to put in place an automatic procedure for choosing the parameter n.

It can be expected that the first singular B-scan is a good estimate of the clutter since the relative energy of this one is in practice superior to the others. However, for a B-scan whose clutter is not relatively stable, it can be expected that the subsequent singular B-scans also contain components of clutter. Therefore, an inconvenient choice of the r and n ranks could cause some clutter residues of relative energy higher than that of the buried objects in the radargram.

In particular, this is the case for the B-scan of Figure 4.5.a which shows a variation of clutter in its central part. The bands of the clutter are not horizontal, and there are some little disturbing objects buried at a low depth. The decomposition into singular values of this B-scan, given in Figure 4.7, shows that the linear combination of the first three singular B-scans is necessary in

order to correctly estimate the clutter, and remove the disturbing signatures that overlap the true reflections. Otherwise, taking only the first (figure 4.7.b) or the first and the second (figure 4.7.c) singular B-scans, as an estimate of clutter, will introduce some residue on radargrams causing a blurring effect and overlapping some signatures.

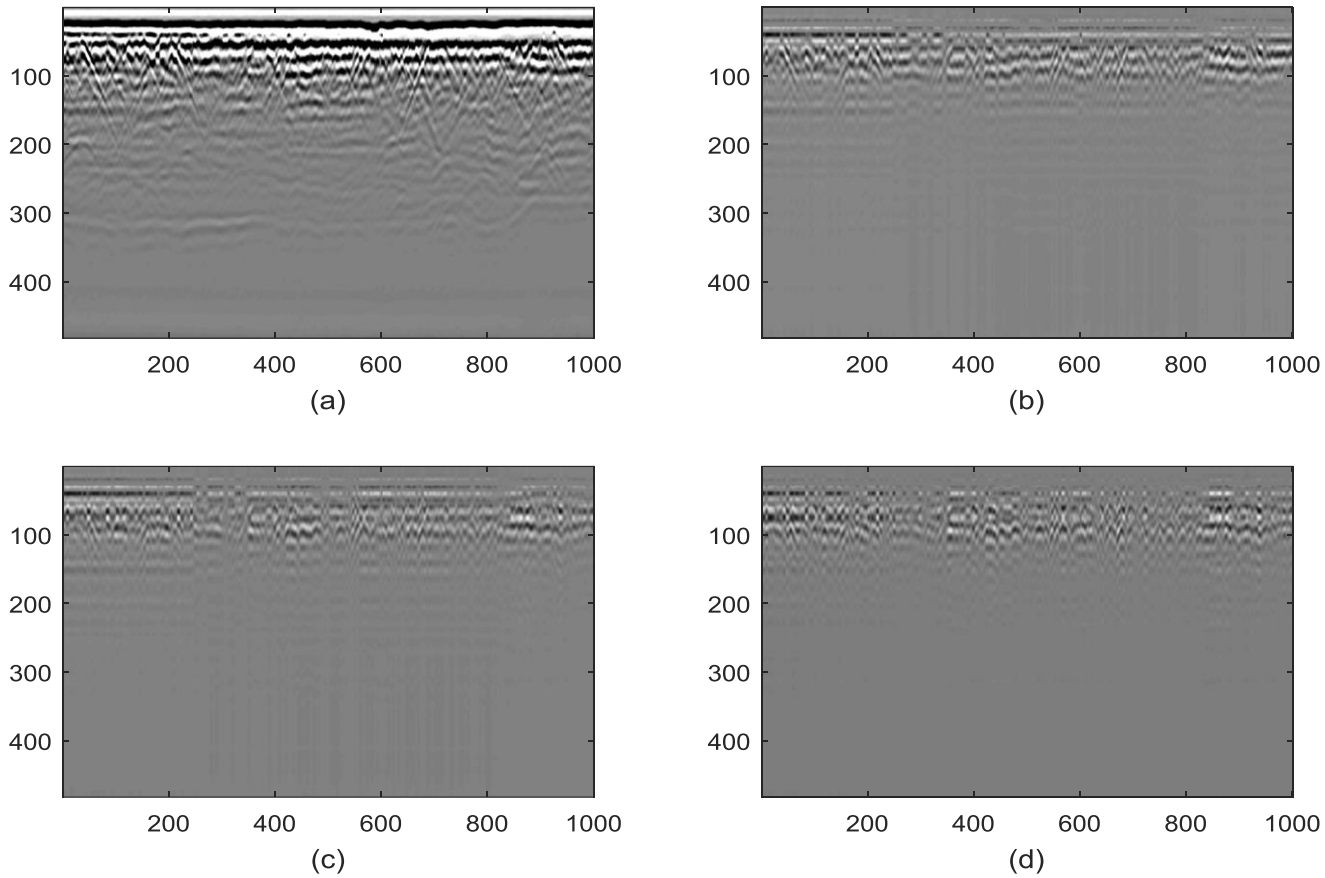


Figure 4. 7: (a) The original B-scan. The B-scan reconstructed with (b) one singular b-scans, (c) two singular b-scans, and (d) three singular b-scans, as estimate for the clutter. The true reflections are more visible, when admitting the first three singular B-scans as estimate of clutter.

4.2.2.2. Wavelet denoising

After an attempt to provide a solution to the non-parametric denoising problem of GPR radargrams, use of the Wavelet transform`s properties is illustrated and introduced here. Mallat [192] solved the denoising problems in general by discussing the Multi-Resolution Analysis (MRA), which is directly linked to the perfect reconstruction of mirror filters bank structure. [192].

Based on MRA, a denoising scheme (please refer to figure 4.8) is built to remove some artefacts that could involve further complications of hyperbola extraction and an analysis

process. A discrete wavelet transform (DWT) is applied for denoising the profile, while a special reconstruction schema is conducted aiming to reduce the background noise effect. In fact, this cross correlation is performed between two reconstructed components in order to enhance the hyperbolas. The two components referred above correspond to one being normally reconstructed while the other reconstructed one is obtained ignoring horizontal and diagonal details, which is the key of the enhancement aspect.

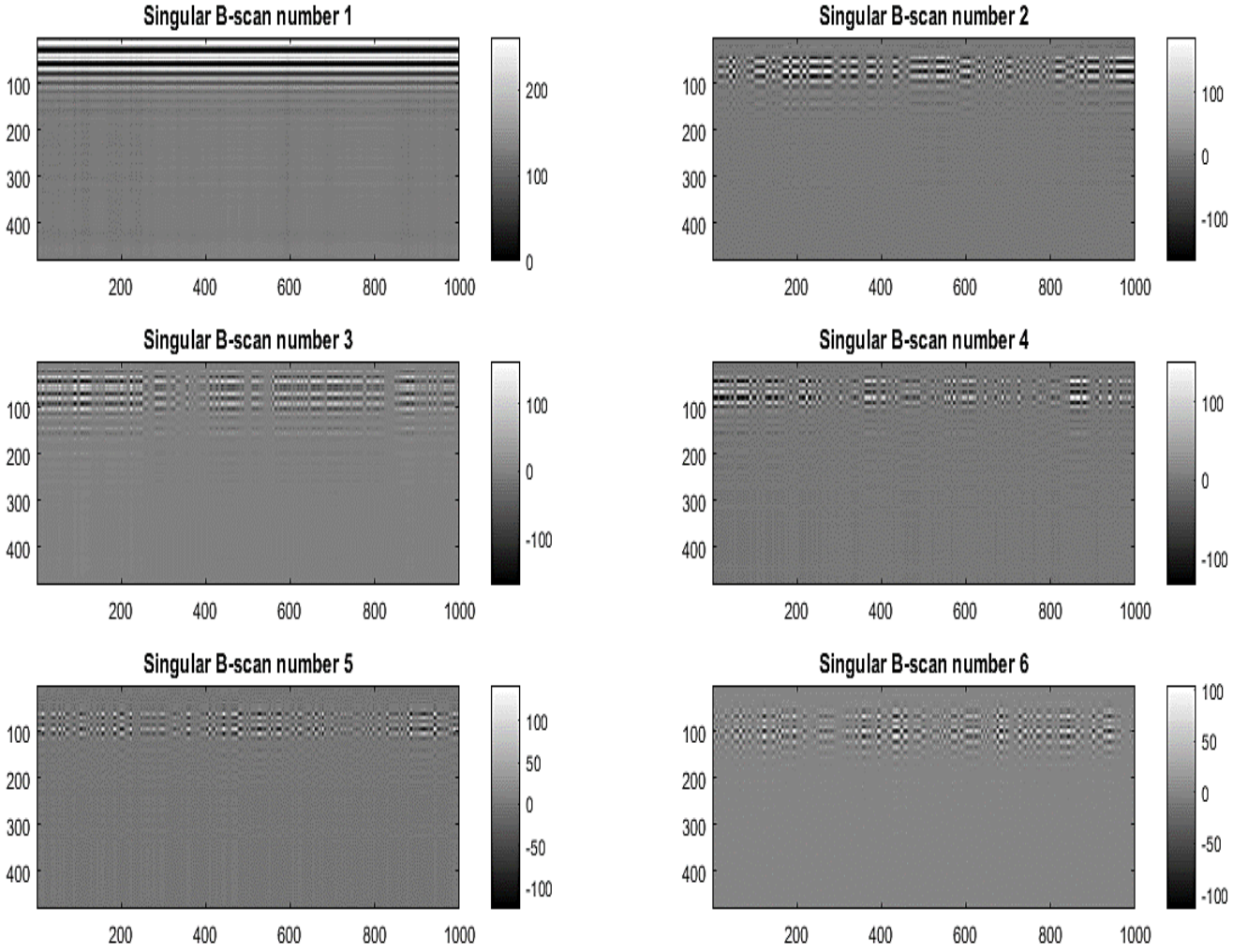


Figure 4. 8: The first six singular B-scans corresponding to the decomposition of an experimental B-scan acquired in the field (figure 4.7.a).

4.2.2.2.1. Wavelet pyramidal algorithm

DWT is computed by the Mallat pyramidal algorithm [192], which proposes to decompose a signal at a level m by a simple convolution operation with a low-pass scaling filter h_1 to have

an approximation signal at level m+1, and a high-pass wavelet filter g_1 to have a detailed signal at level m + 1 (please refer to figure 4.9).

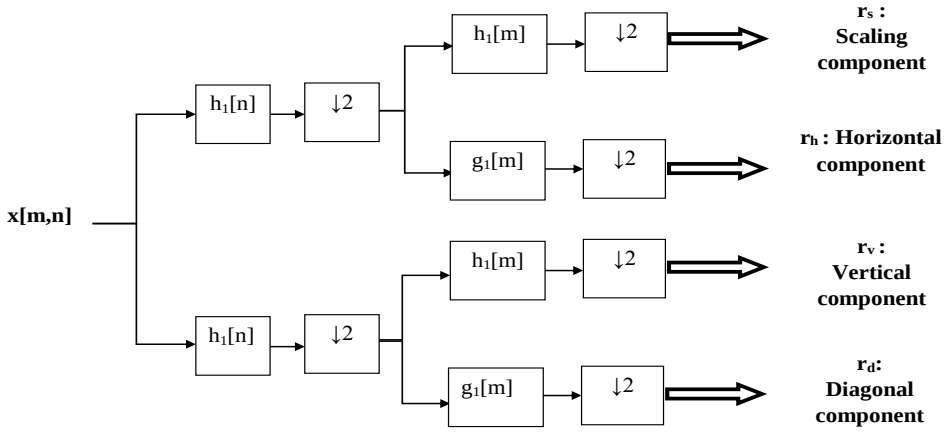


Figure 4. 9: DWT decomposition schema

The 2D wavelet transform is formulated by extending the 1D pyramidal algorithm. Thus, a 1D wavelet decomposition is performed in each direction separately. The 2D DWT of a randomly picked pixel x [m_1,n_1] from the image is given by the formulas:

$$\begin{cases} r_s(m_1,n_1) = 2\sum_{m,n} h_1(m)h(n)x(2m_1-m,2n_1-n) \\ r_h(m_1,n_1) = 2\sum_{m,n} g_1(m)h(n)x(2m_1-m,2n_1-n) \\ r_v(m_1,n_1) = 2\sum_{m,n} h_1(m)g(n)x(2m_1-m,2n_1-n) \\ r_d(m_1,n_1) = 2\sum_{m,n} g_1(m)g(n)x(2m_1-m,2n_1-n) \end{cases} \quad (4.5)$$

Where m_1 and n_1 denote the x-axis and y-axis in the image, respectively, and m and n are the corresponding discrete values. The set of filters {h, g} correspond to an arbitrary set of scaling and wavelet functions.

There are four components in each scale:

- scaling component highlighting the fine fluctuations in the image;
- horizontal component which undertakes the changes along the horizontal axis;
- vertical component which captures the changes along the vertical axis;
- diagonal component.

The choice of the mother wavelet or the basis for the aforementioned denoising procedure, although arbitrary, it is critically important since it has a direct effect on the performance of the technique. There is no specific mathematical rule known to resolve this task, but it is highly recommended to use the wavelet that could be compared to a GPR signal in the main form of

the envelope. According to the study conducted in [193], 'Symmlet order 6'(Sym6) and 'Debauchees' order 6' (Db6) are the recommended ones to analyse GPR data. Moreover, several tests were conducted on a set of synthetic radargrams, in addition to a database of GPR images. Accordingly, it has been concluded that three wavelet types could be used to denoise the set of data deployed in this study (synthetic and experimental ones): Sym6, Db6 and Haar wavelets.

An example of decomposition by these wavelets is presented in Figure 4.10.

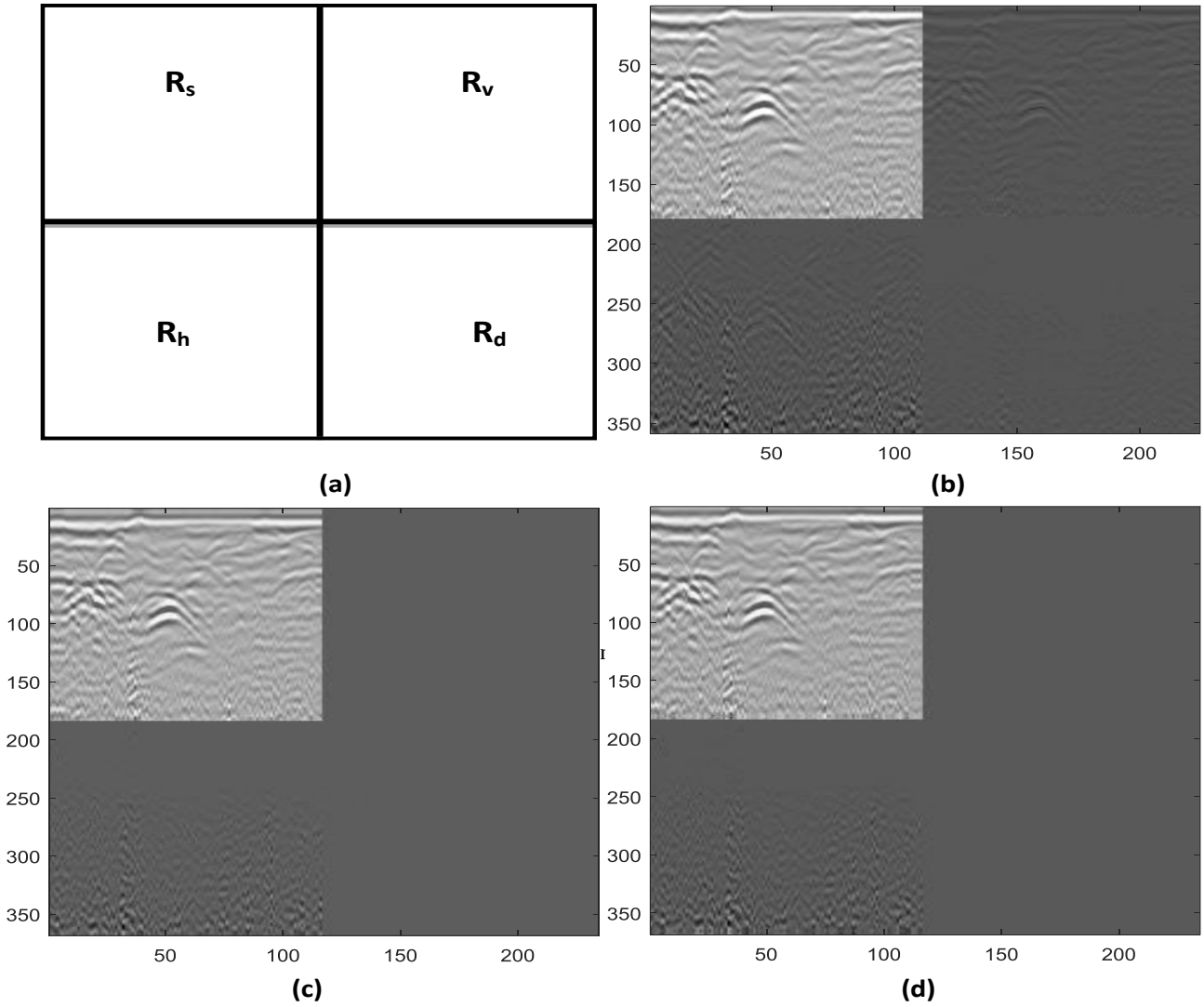


Figure 4. 10: An example of a wavelet radargram decomposing with: (b) 'Haar', (c) 'Debauchees order 6', (d) 'Symmlet order 6' mother wavelet. The scheme (a) represents the position of each component in the images: the scaled (Rs), vertical (Rv), horizontal (Rh) and diagonal (Rd) components.

It is visually interpreted that every wavelet acts differently than the others. This demonstrates the capability of separating some new affine fluctuations and transitions in the

radargram that were not previously detected. However, for this current example, it is clear that the Haar wavelet is the most convenient one to denoise this image.

4.2.2.2.2. Thresholding

To denoise radargrams by wavelet transform, only detail coefficients (vertical and horizontal components) are thresholded since they contain the structures that are largely affected by noise. The simplest thresholding technique is the 'hard' thresholding, where the threshold values of the detail coefficients are found according to the following formula:

$$T_{th}(i) = \begin{cases} T(i) & if\ |T(i)| > th \\ 0 & if\ |T(i)| \leq th \end{cases} \quad (4.6)$$

T(i) is the detail coefficient number i, and th the selected threshold.

The chosen threshold is universal one, defined as follow:

$$th = \sigma\sqrt{2log(N)} \quad (4.7)$$

N is the length of the signal and σ the noise level, which is given by the equation:

$$\sigma = \frac{median\ of\ detail\ coeffecients}{0.6745} \quad (4.8)$$

4.2.2.2.3. Reconstruction

The new thresholded detail coefficients $\{x_h', x_v', x_d\}$, are used for the signal reconstruction process instead of the detail coefficients $\{x_h, x_v, x_d\}$. The reconstruction is performed by an inverse discrete wavelet transform (idwt).

Two profile types are reconstructed: one normal Rr_s, with the coefficients $\{x_s', x_h', x_v', x_d\}$, and a second one R_{rb}, ignoring horizontal and diagonal details. A cross-correlation is performed between the two images (please refer to figure 4.11 and 4.12).

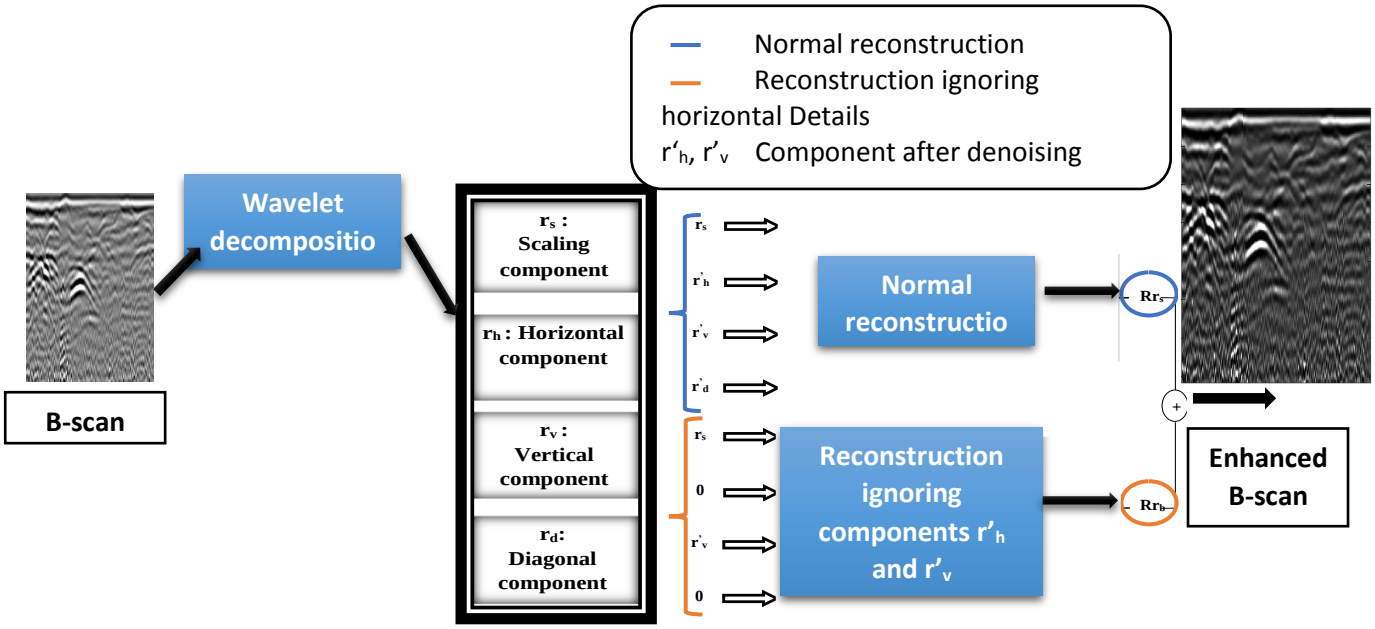


Figure 4. 11: Wavelet based denoising schema

In order to assess the performance of this schema (figure 4.12), measurements of SNR and PSNR, for the same image of the example seen in figure 4.10, are collected in table 4.1 for both the normal reconstruction, which uses the idwt and the currently proposed schema. The tests are done with three types of wavelets: Haar, Db6 and Sym6. SNR (signal to noise ratio) and PSNR (Peak signal to noise ratio) are metrics to measure quality of images, they are defined by equation 4.9.

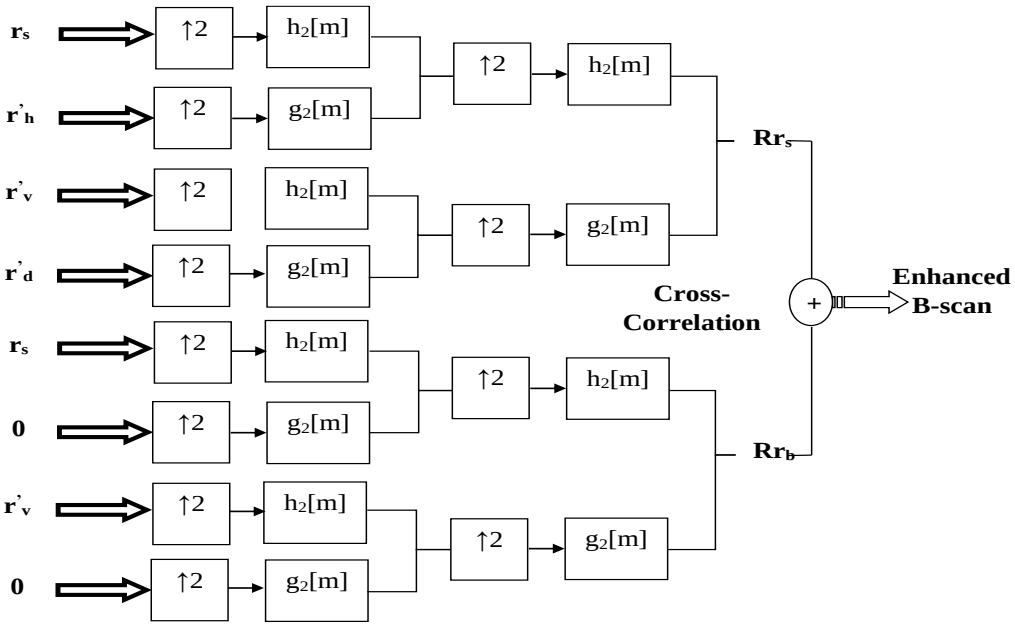


Figure 4. 12: idwt reconstruction schema, h2 and g2 represent conjugate mirror filters bank

Given a noise-free $sz_1 \times sz_2$ image I and its noisy approximation A:

$$\begin{cases} SNR = \frac{\sum_{m=0}^{sz_1}\sum_{n=0}^{sz_2}(I(m,n))^2}{\sum_{m=0}^{sz_1}\sum_{n=0}^{sz_2}(A(m,n))^2} \\ PSNR = 10\,log_{10}\frac{(max(I(i,j)))^2}{MSE} \\ MSE = \frac{1}{sz_1 sz_2}\sum_{m=0}^{sz_1}\sum_{n=0}^{sz_2}(I(m,n)-A(m,n))^2 \\ i = 1,\dots,sz_1 \; and \; j = 1,\dots,sz_2 \end{cases} \quad (4.9)$$

Where MSE is the mean square error.

Table 4. 1 : SNR and PSNR measures for B-scan of example of figure 4.10

Wavelet	Metric/Measure in dB	Normal reconstruction (idwt)	Cross-correlation reconstruction
Haar	**SNR**	0	44.02
	PSNR	24.27	19.37
Db6	**SNR**	1.55e-09	44.06
	PSNR	18.67	19.38
Sym6	**SNR**	-1.17e-07	44.06
	PSNR	15.90	19.38

It is visually clear from table 4.1 that the proposed schema provides better quality images in comparison with the normal idwt schema. This is due to the fact that the proposed schema gives the higher SNR and PSNR values for the three wavelet types.

Although the Haar wavelet is able to extract more affine details of the images, the results show that Db6 and Sym6 provide better results for the cross correlation based schema (i.e. higher SNR and PSNR values).

4.2.2.3. Corners detection

Corners in an image are determined by locating points that have locally maximum cornered measures. The determinant of the inertia matrix IM (4.10) can be used as a cornered measure.

$$IM(x,y) = \begin{bmatrix} \overline{I_x(x,y)\,I_x(x,y)} & \overline{I_x(x,y)\,I_y(x,y)} \\ \overline{I_y(x,y)\,I_x(x,y)} & \overline{I_y(x,y)\,I_y(x,y)} \end{bmatrix} \quad (4.10)$$

$\overline{I_x(x,y)}$ and $\overline{I_y(x,y)}$ are average gradients of image with respect to x and y in a small window centered at (x, y) [194].

Note that the eigenvalues of matrix IM are indicators of the strength of gradients to each other in normal directions inside the centred window. In order to detect corners, a 'minimum eigenvalue algorithm' is used [194].

An eigenvalue of the IM matrix is given by equation (4.11):

$$\begin{cases} \delta = \frac{-B \pm \sqrt{B^2 - 4C}}{2} \\ B = -\left(\bar{I_x}^2 + \bar{I_y}^2\right) \quad and \quad IM = \bar{I_x}^2 \bar{I_y}^2 - \left(\bar{I_x}\,\bar{I_y}\right)^2 \end{cases} \tag{4.11}$$

4.2.2.4. Hough transform

An edge detection step is performed before applying Hough transform. The one dimensional (1D) Hough transform algorithm is summarized in table 4.2. The results given in this paragraph are for the same radargrams analysed in figure 4.10.

Table 4. 2: One dimensional Hough Transform

Step	Algorithm
Step 1	The input are the corners points which represent parameters (α, γ)
Step 2	Create an empty accumulator array with the estimated beta parameters for the simulated configuration; depending on antenna frequency, it is trivial to deduce the maximum depth of penetration of the wave and other related parameters.
Step 3	For every pixel (I_{px}, I_{py}) and a couple (α, γ), the corresponding beta parameter is computed with equation: $\beta = \frac{I_{py}^2 - \alpha}{(I_{px} - \gamma)^2}$ The corresponding accumulator entry is incremented by one.
Step 4	Find accumulator peaks, which corresponds to hyperbola apexes.

5.2.3. Results comparison

Firstly, tests were done with an original Hough transform (please refer to chapter 3, section 3.2), after clutter removal and an image denoising process. The number of accumulator cells is defined by the number of edge pixels. The third parameter is varied in a given interval, which can be set by the user depending on the survey configuration and expected results. Every edge pixel is then transformed within the hyperbola parametric equation into the Hough plane for every value of the third parameter β. Only if the calculated values for the first and the second parameters (α, γ) are real and greater than zero, the corresponding cell in the accumulator array is incremented by one. When this procedure is accomplished, for every possible parameter combination, the accumulator cell with the maximum count is searched and the coordinates of the apex and the velocity of propagation of the wave are given. Figure 4.13 presents details

about hyperbola apexes detected by original Hough algorithm. The position of apexes is depicted by red crosses in figure 4.13.a. and they are designated in red color in figure 4.13.b representing the peaks of the accumulator. It is apparent from the illustration (figure 4.13.a) that it generates a lot of random detections, hence, the misclassification rate is high when using the original Hough transform on these high resolution images.

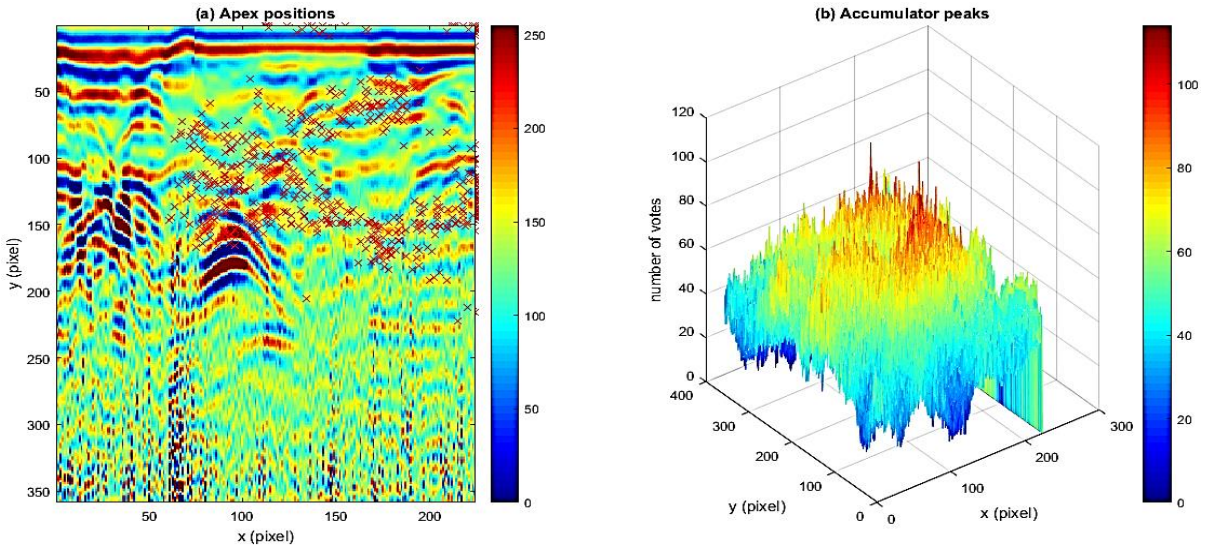


Figure 4. 13: Original Hough Transform: (a) apexes of hyperbolas are depicted in red cross motifs, (b) Hough accumulator plot

Secondly, tests were conducted within the 1D Hough schema. The size of the accumulator array was chosen depending on the frequency of the survey, 30 MHz in this case. The estimated temporal discretization step is $\Delta t = 8.3799 * 10^{-11}$, the maximum velocity $V_{max} = 3 * 10^{-8}$ (velocity of air), and the maximum value of alpha parameter gathered experimentally $\alpha = 3.4695 * 10^{4}$. Hence, the maximum value estimated for the beta parameter is $\beta = 290$. Thus, an accumulator array with 290 cells is created and initialized to zeros (figure 4.14.b). For every acquired corner point (α, γ) in every pixel in the picture, the parameter β is calculated and the corresponding entry of the accumulator is incremented by one. Once the voting process is finished, accumulator peaks correspond to hyperbola apexes. The results gathered with the 1D Hough transform are drawn in figure 4.14.a. It is evident that the problem of acquiring random detections is solved. This achievement was only possible because of the usage of the 1D Hough transform. The previously detected apexes outside the picture borders are omitted in figure 4.14.a.

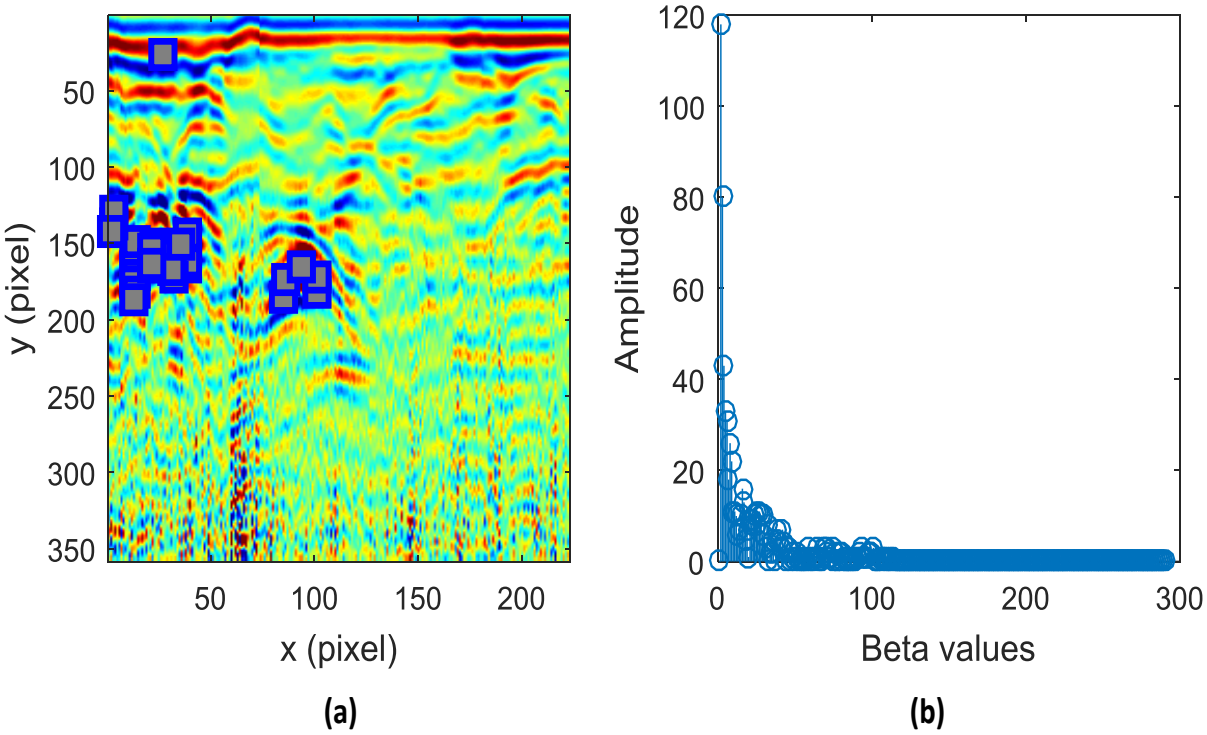


Figure 4. 14: Modified Hough Transform: (a) apexes of hyperbolas depicted in blue rectangular motifs, (b) 1D Hough accumulator plot

Table 4.3 regroups the detection results obtained by the two Hough algorithms over the same image, while the true hyperbolas location are given in figure 4.15.

Table 4. 3: Detection results collected for the two tested algorithms for the same radargram analysed on figure 3.11

Algorithm	**Machine Configuration**	**Time of execution**	**Number of:**			
			Marked hyperbolas	**False detections**	**Mean (multiple detections)**	**True detections**
Hough Transform	Intel Core i5-4200U CPU @ 1.60 GHz 2.30 GHz/RAM 6 GHz	37.50	5	~ >30	~ >10	2
Proposed 1D Hough schema		1.67		19	3	4

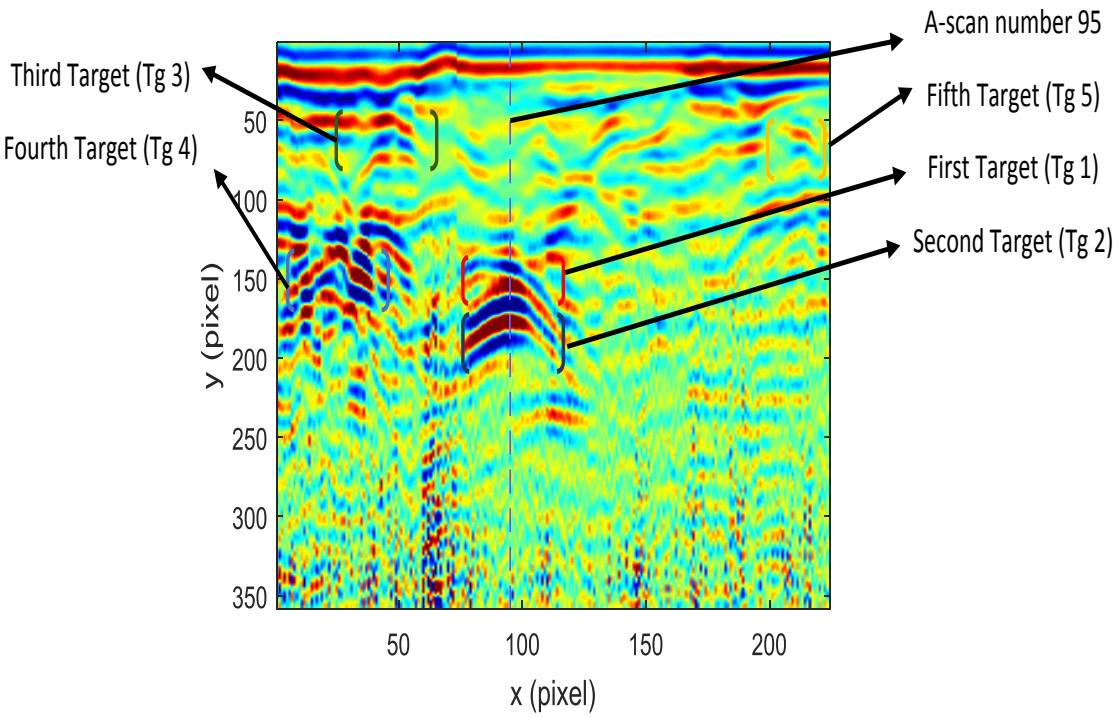


Figure 4. 15: True hyperbola positions. The line represents the A-scan to be analysed in the next paragraph

The numbers clearly show that the 1D Hough algorithm gives better results, since the corresponding number of detected hyperbolas is higher, the number of false detections is lower, and the average number of multiple detections for the same target is acceptable in comparison with original Hough.

Nevertheless, the problem of multiple detections recorded for the same hyperbolas persists, namely the peaks splitting issue. It is commonly reported that this problem is directly related to the high resolution of images.

In the next chapter of this thesis, an adequate solution for this problem is introduced, which consists of narrowing down the hyperbolas positions to certain areas and then applying this algorithm to these regions.

The velocity calculations are given for the pointed hyperbola apexes of figure 4.14.a, and the corresponding penetration depth, or the target position estimate, are regrouped in table 4.4.

Table 4. 4: Detection results collected for the two tested algorithms for the same radargram analysed on figure 3.11

Point number	Velocity (x10^8 m/s)	Depth of target (m)
P1	2.6327	1.7618
P2	2.6327	1.8163
P3	2.6327	1.8804
P4	2.6327	1.9332
P5	2.6327	1.7092
P6	2.6327	1.8328
P7	2.6327	1.6656
P8	2.6327	1.7888
P9	2.6327	1.6991
P10	1.0326	0.8035
P11	1.2411	0.9466
P12	2.6327	1.9199
P13	2.6327	1.6476
P14	2.6327	1.8999
P15	2.6327	1.6690
P16	2.6327	1.8898
P17	2.6327	2.0055
P18	2.6327	1.7419
P19	2.6327	1.5846
P20	2.6327	1.8158
P21	2.6327	1.6991
P22	2.6327	1.8491
P23	2.6327	1.8222
P24	2.6327	1.7980
P25	2.6327	1.6595
P26	2.6327	1.5138
P27	2.6327	1.4167
P28	2.1496	1.6776
P29	1.8616	0.2024
P30	2.6327	1.5660

The points highlighted in yellow (P2, P17) represent a clear example of the peaks splitting issue, occurring when the peak is divided into two or three peaks. It is easy to distinguish the real apexes. Although, the problem is more complicated when multiple peaks representing the same points and with a huge number of targets (>10) exists. The points highlighted in blue represent the true detections corresponding to velocity and depth profiles. The velocity calculated at this stage represents an average value between all the media that the wave transits through it. In order to reconstruct the true velocity of the wave, a time frequency representation of A-scan signals is proposed to reconstruct the time of propagation within the target. This study aims to give a clear idea about the nature of the target, depending on the permittivity of the medium.

4.3. GPR Time Frequency Representation

4.3.1. Aim of research

After using the Hough pattern recognition technique, the fitted parameters of hyperbolic signatures are used to estimate location and size of the related target objects. However, this approach provides an average propagation velocity of the signal in the medium. To avoid this, the frequency representation of the 1D A-scans enables the user to visualize the time propagation of the wave within the medium. Fourier transform of such non-stationary signal does not explicitly evaluate the evolution of frequency contents with time. This is visually drawn in figure 4.16.a where the signal appears like a monotone function, and the transition parts of the slice are invisible.

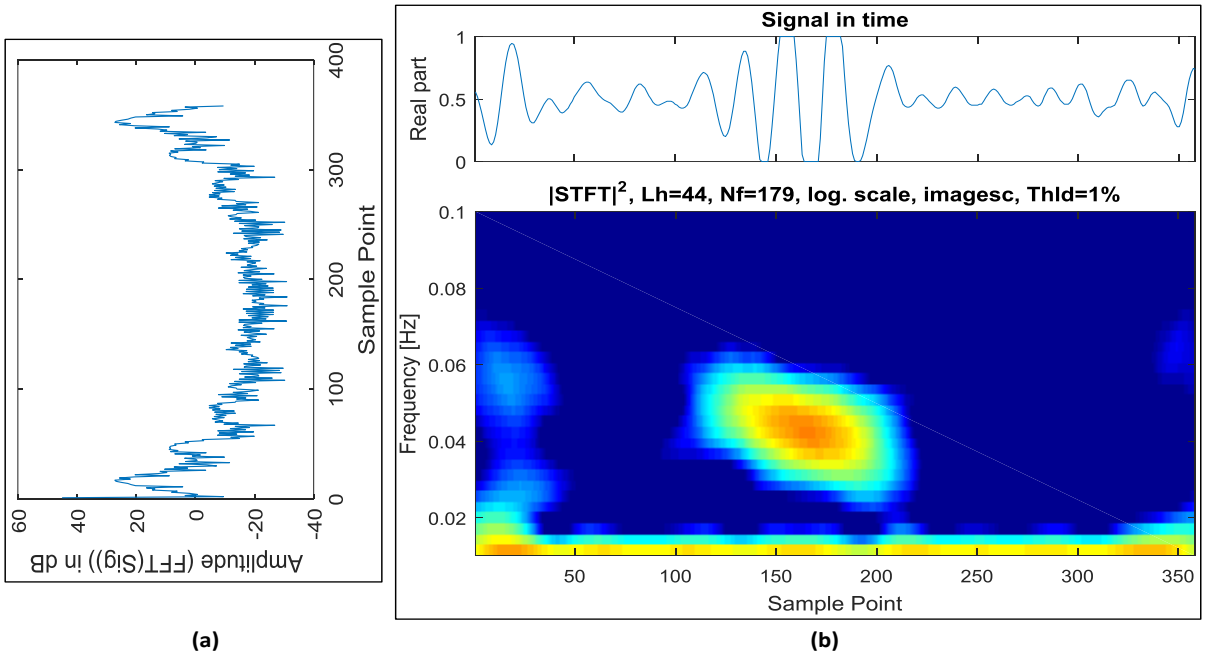


Figure 4. 16: (a) Fourier transform of A-scan number 95, (b) Short time Fourier transform of the A-scan signal given in the upper figure

This shortcoming could be compensated by applying Short Time Fourier transform (STFT), which gives an adequate representation of the signal in a time–frequency plan. In STFT, pre-fixed segments of the signal are correlated with a window function concentrated in both time and frequency domains. The distribution is given in figure 4.16.b. A low-aliasing short Hamming window function was chosen for this study. A window resolution, one-fourth of the

length of the signal (358/4~=89), was optimized after a trial and error process. The process alleviates and reconciles the arrangements regarding aliases of the spectral component, due to the usage of a small resolution time window; the unclear targets due to larger resolution window, and the influence of other neighbouring frequency components.

In STFT, narrow windows yield good time resolution and poor frequency resolution and vice versa. The window size used in examples of figure 4.16.b gives better frequency resolution but poor time resolution.

While the STFT gives an adequate compromise between time and frequency information, the drawback is that it uses a fixed size time window for all frequencies and GPR signals require a more flexible approach.

In the next paragraphs, a detailed study is conducted to assess the capability of different techniques to give a high-resolution representation of GPR signal in a time-frequency plan, despite the imposed constraints. The time frequency analysis is done with a Flandrin Toolbox [195].

4.3.2. Application for case of non-quadratics approaches

4.3.2.1. Traditional time-frequency distribution

The first approach to discuss at this level, which proposes a direct solution to the drawbacks presented by STFT, is the Scalogram. The Scalogram uses a more flexible window, but an important point within this approach is the choice of the mother wavelet. For seismic applications, the Morlet wavelet is often used. Since a GPR signal is very similar to the seismic one, a Morlet wavelet is selected as mother wavelet. The half-length of the Morlet analysing wavelet at coarsest scale is set to 4, after several tests for optimizing this parameter, it is deduced that using this value we obtain higher resolution with minimal interferences.

The occurrence of energy variations at specific frequencies could be observed in the Scalogram of figure 4.17.

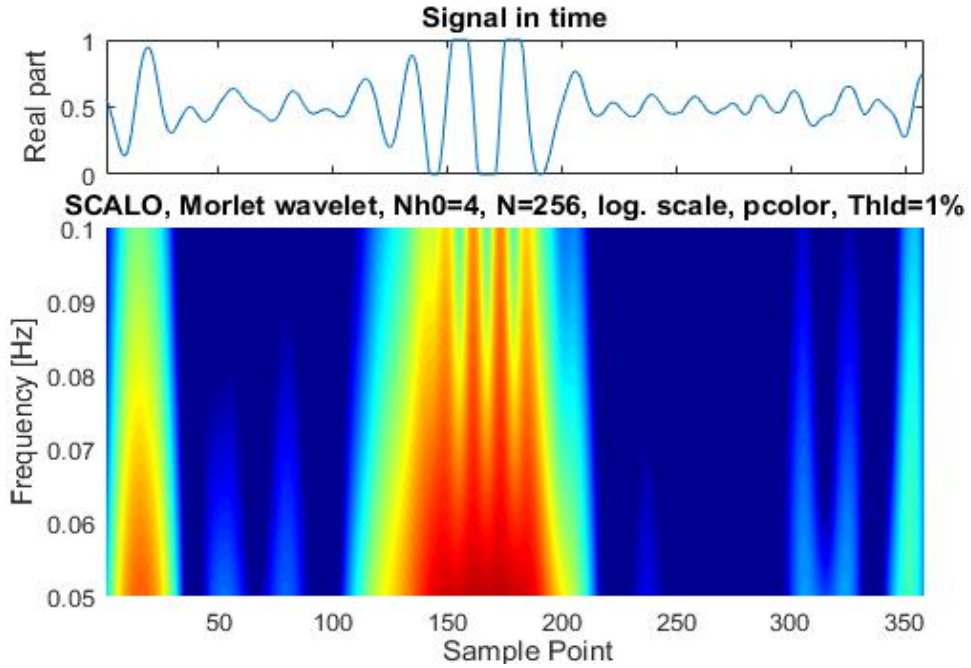


Figure 4. 17: Scalogram of A-scan number 95: mother wavelet is a Morlet wavelet with a half-length=4

According to these variations, a threshold is set to get a time-frequency distribution with minimal overlapping frequencies. It is visually interpreted that this distribution offers higher frequency resolution, but it still does not reach the desired expectations. There are some undetected transitions of signals. Besides, the higher energetics events, situated in the middle of the slice, interfere and appear to the reader as a single event.

Another variant of STFT is the Stockwell transform (S-transform) given in figure 4.18, which gives a very poor time-frequency resolution compared with the Scalogram.

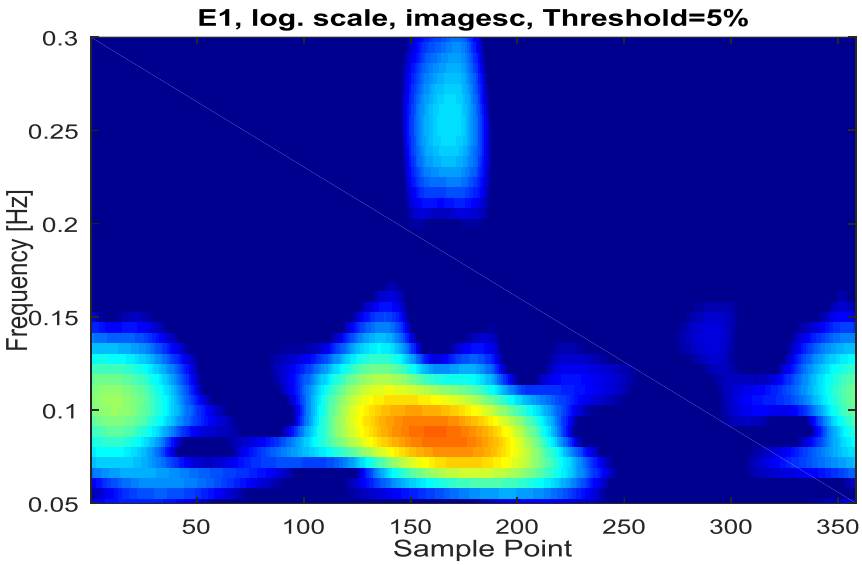


Figure 4. 18 : S-transform of A-scan number 95

4.3.2.2. Hilbert Huang based distributions

Hilbert Huang is a new technique offering a visualization of the signal in a time-frequency plan based on the empirical mode decomposition approach. It also had proved the capability of providing higher resolution for analysing different types of signals [163, 196].

Figure 4.19.a represents the IMF components received from the EMD technique. The last signal is the residue of decomposition. However, the Hilbert Huang spectrum is given in figure 4.19.b.

The noise presented in the picture, in the form of colored random points, is named the ripple phenomenon which often occurs when one obtained IMF cannot satisfy the strict mono-component conditions. This phenomenon completely misleads the analysis in this case.

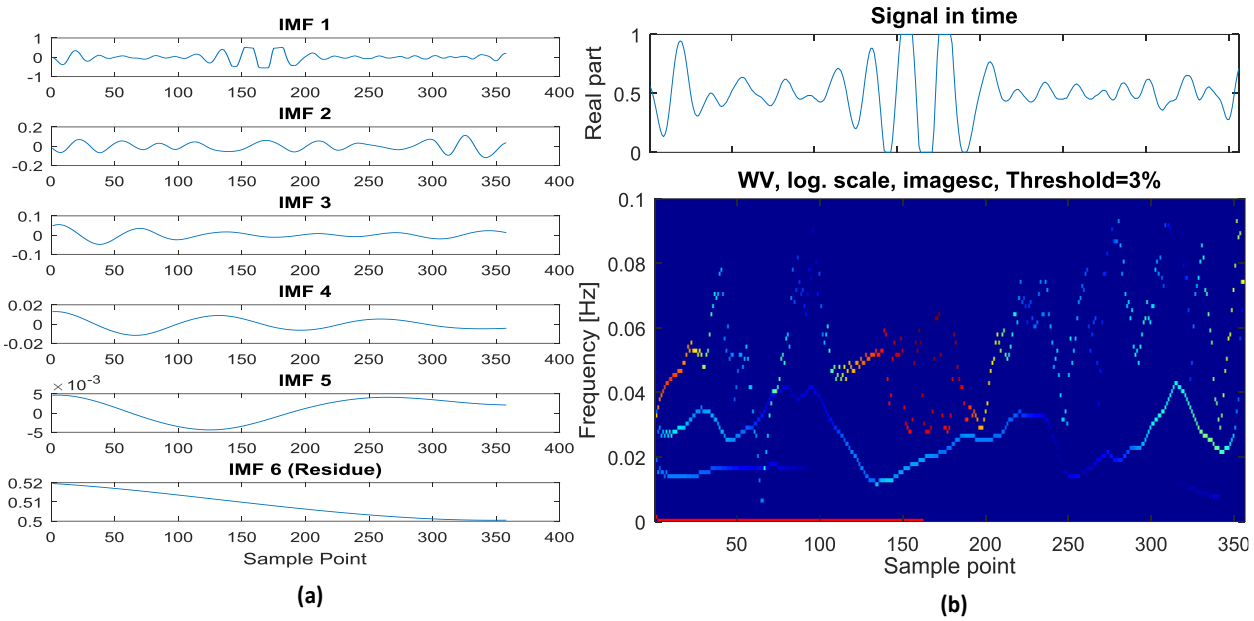


Figure 4. 19: (a) Hilbert Huang Spectrum (b) IMFs results of EMD decomposition

In general, it is theoretically stated that HHT suffers from a number of shortcomings:

- The EMD decomposition technique generates undesirable IMFs at low-frequency region.
- The first gathered IMF cover a too wide frequency range that the property of mono-component is not achieved.
- The EMD operation is not able to separate signals that contain low-energy components.

Through the literature, an improved version of HHT [196], which performs a filtration of inappropriate IMF on several steps, is proposed. This improved version uses the wavelet packet transform (WPT) to separate the signal into a set of narrow band slices. Thus, the low energy components are decomposed into different bands, and it becomes easier to identify low energy components. Afterwards, an EMD process is applied to decompose these narrow band slices as the original EMD does. The set of IMFs are in narrower bands and satisfy the condition of the mono-component. A shifting process is then conducted to select vital IMFs from unrelated ones. This can be acquired by calculating the correlation coefficients of the IMFs with the inspected raw signal. The HHT, with WPT as a pre-decomposition technique and IMF selection method, is referred to as an improved HHT. The algorithm is given in table 4.5.

Table 4. 5 : Improved HHT algorithm [196]

Step	Process
WPT decomposition	▪ The original signal sig is decomposed into a number of narrow bands equals to 2^n: $\{B_1, B_2, \ldots, B_N\}$ ▪ Cross-correlation between the original signal and the narrow band slice is calculated: $\lambda_1 = \max_{i=1,\ldots,N}(corr(sig, B_i))$
EMD process	▪ If $corr(sig, B_i) > \frac{\lambda_1}{\eta}$, the narrowband is retained, otherwise it is rejected. ▪ EMD is performed for every retained narrow band B_i, i=:1,…, N.
IMF selection	▪ Cross-correlation between the IMF and the narrow band slice is calculated: $\lambda_2 = \max_{i=1,\ldots,N, j=1:5}(corr(IMF_j, B_i))$ ▪ If $corr(IMF_j, B_i) > \frac{\lambda_2}{\eta}$ the IMF is retained, otherwise it is added to the residue. ▪ The IMF of different narrow bands are concatenated in a single vector.
HHT spectrum	Hilbert Huang spectrum is calculated.

Two kinds of screening processes are considered:

- The first screening process is proposed to remove weakly correlated narrow signal, which has poor correlation with the original signal.
- The second screening process is employed to eliminate the IMFs, which displays a poor correlation with their corresponding narrow band signals generated from the WPT.

For the WPT, there are two important parameters: wavelet basis and decomposition level n. A series of candidate wavelet basis are tested within multiple images, such as dmey, db4, db2, coif4, sym3 and haar, in which the Daubauchy order 2 wavelet yields the best results. The db2 wavelet is a compactly supported biorthogonal wavelet and has a good time-frequency localization ability.

A WPT, with n =7, is chosen considering the real-time performance of the practical system and discriminability of different states. A number of experiments show that the last retained IMFs are not sensitive to the ratio factor η, so the empirical value η =10.0 is adopted [196]. The insensitivity also reveals the robustness of the proposed method sufficiently.

For optimization reasoning, the first screening process is proposed to further reduce the computing complexity. The results of the improved HHT are shown in figure 4.20. It is remarked that the ripple noise had disappeared, at a cost of the low time and frequency resolution of the non-highly energetic components. The highly energetic components are clearly distinguished, but other lower energetic events are blurred due to poor resolution issue. The next paragraph will give results of the analysis of the same A-scan signal using Wigner-Ville distributions characterized by higher resolution in both time and frequency plans.

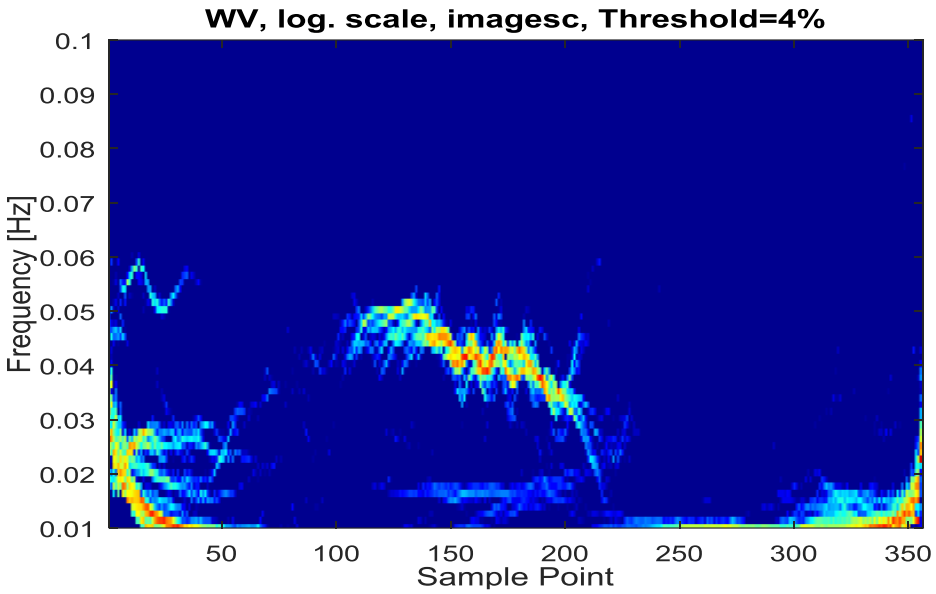


Figure 4. 20 : Improved Hilbert Huang Spectrum

4.3.3. Application for a case of quadratics approaches

The analysis is done using four types of WVD based distributions, which offer higher time-frequency resolution compared with the non-quadratic techniques: the original WVD, SPWD, SPAWD, and CWD.

In order to evaluate the performance of time frequency distributions, Renyi [197] concentration criterion is deployed. The generalized normalized Renyi entropy is defined as follows [197]:

$$R_\alpha = \frac{1}{1-\alpha}\log_{10}\left(\sum_n \sum_k \left(\frac{TF(n,k)}{\sum_n \sum_k |TF(n,k)|}\right)^\alpha\right), \alpha \geq 2 \qquad (4.12)$$

The choice of the parameter α is an important matter while using Renyi entropy. α = 1, this case will not be adopted because it refers to the specific situation of Shannon entropy. α = 2 describes the case of quadratic entropy, which is characterized by the ability to measure both useful (auto-terms) and undesirable terms (cross-terms). Furthermore, as the value of α increases, Renyi entropy continually decreases.

According to the information theory, higher quality distributions are ones having least possible entropy values. Thus, they meet the requirements of having maximum information, best energy concentration, peakiness, and higher resolution [198-200]. Hence, the optimized

constants of the kernel of the distributions, according to the lowest value of Renyi entropy, will provide better localization of the energy.

However, sometimes Renyi entropy fails to act in the desired way when the cross-terms appears. Stankovic measure [198] does not need any normalization to act accurately when oscillating terms exist, and it treats low and high concentrated components in the same way. Stanckovic measure is defined as:

$$M_p^p = \left(\sum_n \sum_k |TF(n,k)|^{1/p}\right)^p \quad (4.13)$$

Among given quadratic time frequency distributions, the most concentrated is the one with the nominal non-zero amplitudes time-frequency region, i.e. the distribution with the smallest value of M_p^p.

Since the M_p^p measure is sensitive to small amplitude values, p = 2 or p = 4 should be used in practical situations, which contradicts the original signal (time duration definition requires that $p \to \infty$). However, when extending any one dimensional existing definition of concentration/duration/spread to its two-dimensional equivalent, additional modifications may need to be introduced. Hence, in the M_p^p measure, p should not be restricted as it is in the original one dimensional duration definition case [198].

Based on the Renyi criterion, as a primarily step it is necessary to find the best parameters that yield higher concentration values. In fact, the best parameters (size and type of smoothing windows, the number of frequency bins, kernel width…) are estimated by the experiments. The selected ones are the parameters which give the lowest Renyi entropy measurements.

The time-frequency distribution for the same A-scan signal, analysed in the previous study, is presented in figure 4.21. The different optimized parameters used for each distribution are regrouped in table 4.6. Figures 4.21.b-d show a clear trend in time resolution. However, figure 4.21.a proves that WVD has a big rate of cross terms. The graph of figure 4.22, regrouping the different concentration values for all the distributions, is revealing some important observations:

- WVD is completely mislead by cross-terms, since it gives very low concentration measures.
- According to the Renyi criterion, SPAWD is the best distribution for analysing GPR signals.
- According to the Stanckovic criterion, SPWD and CWD are the highest concentrated representations (there corresponding values are very close to each other).
- Nevertheless, it is visually clear that CWD offers a better resolution in comparison with SPWD.

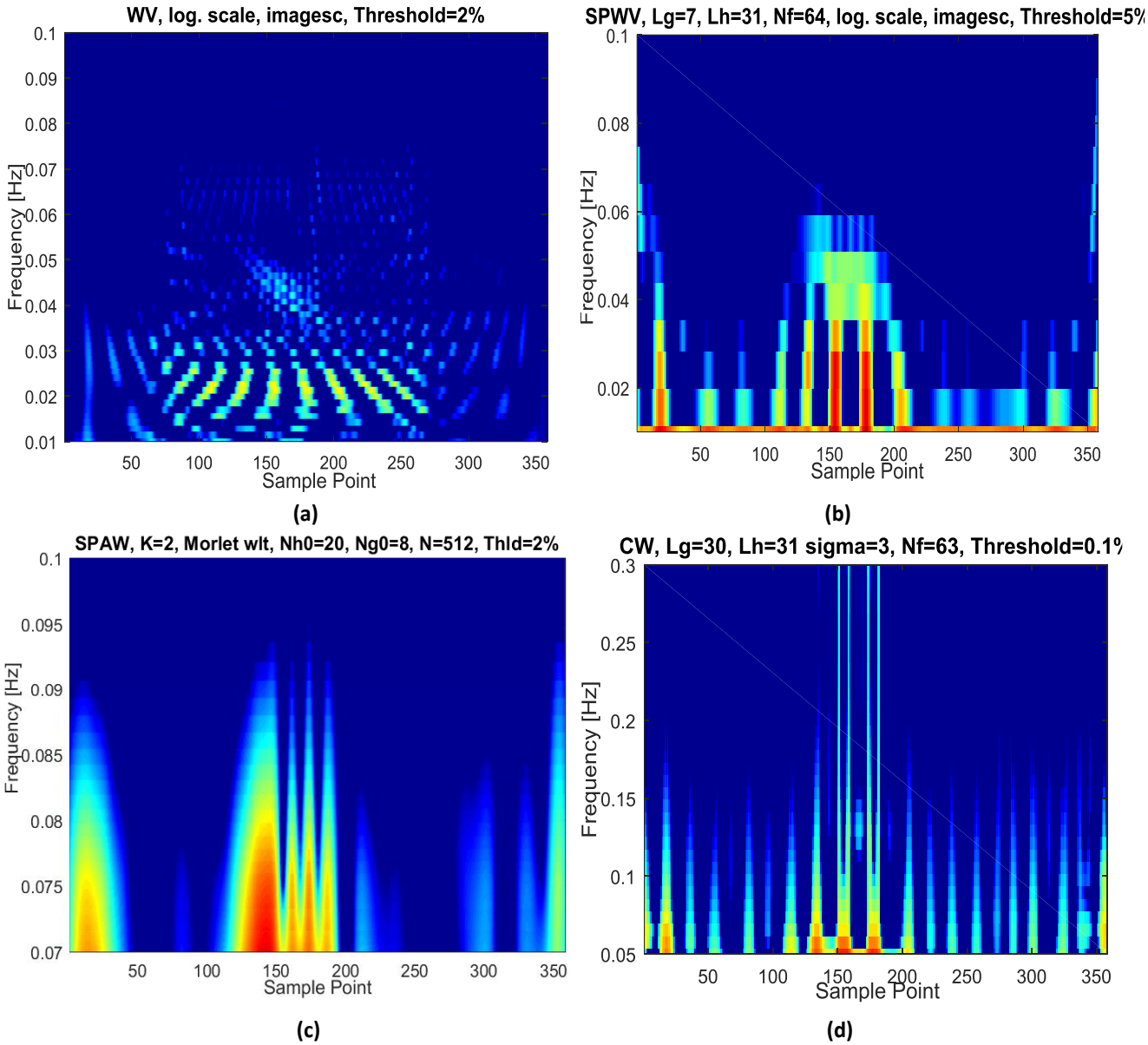


Figure 4. 21: Wigner-Ville distribution: (a) Original WVD, (b) SPWD, (c) SPAWD, (d) CWD. The frequency represents the normalized values. Nf and N corresponds to the number of frequency bins. Nh0/Lh and Ng0/Lg are the half-length of time and frequency windows.

Table 4. 6 : Optimized Parameters of the four distributions

TF Distribution/parameter	Time smoothing window	Frequency smoothing window
WVD	None	none
SPWVD	Kaiser window, length=15	Kaiser window, length=63
SPAWVD	Half length of the time smoothing window Ng0=8	A Morlet wavelet, half length of the analysing wavelet at coarsest scale Nh0=20
CWD	Kaiser window, length=61	Kaiser window, length=63, kernel width: SIGMA=3

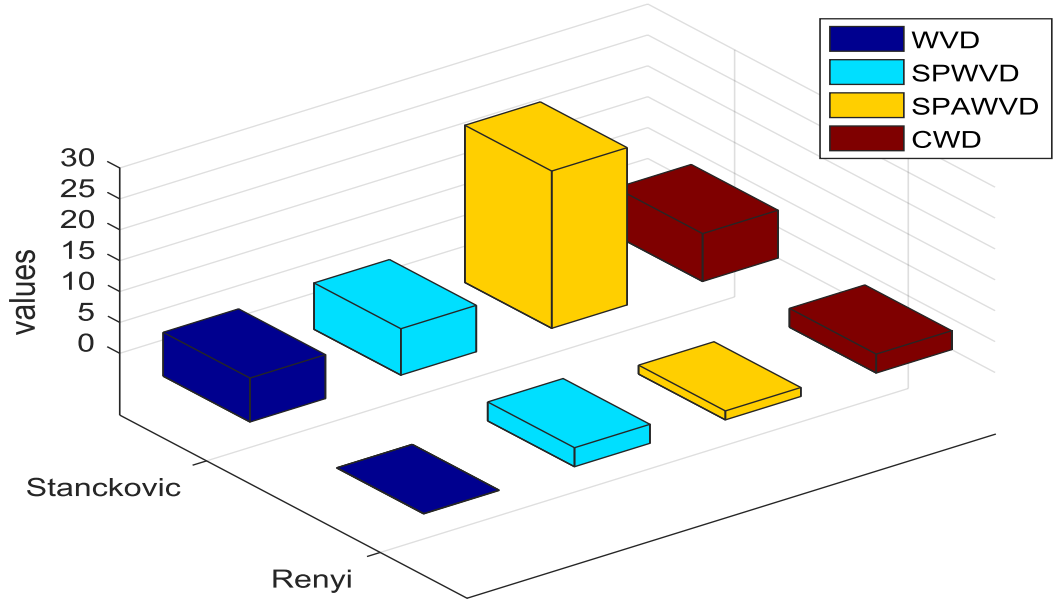


Figure 4. 22 : Renyi and Stanckovic concentration measures of the four distributions.

4.3.4. Results comparison

The calculations of the time of propagation were done for the four selected distributions: Scalogram, improved HHT, SPAWD, and CWD.

The proposed technique, to visually read the wave propagation time, was explained in figure 4.23. Regarding the low quality of certain distributions, the time interval of the incoming events was located in the radargram (figure 4.23.a) as well as the time slice (figure 4.23.b) in order to easily discard the noisy peaks. Otherwise, it will be difficult to know the adequate peaks to consider, since the two observed targets are very close to each other.

The first target (Tg 1) is pointed with the red colour and the second one (Tg 2) with the green colour. For the Scalogram (figure 4.23.e) and the improved HHT (figure 4.23.e), it was impossible to locate the time intervals that need to be read, without referring to time signal and B-scan image, and admitting some assumptions about the target.

For the SPAWD, regarding the first target one can see that it exists some overlapping peaks that come just like a single event (from 100 to 150) corresponding to target reflection. Without referring to the temporal slice, the read value will be higher.

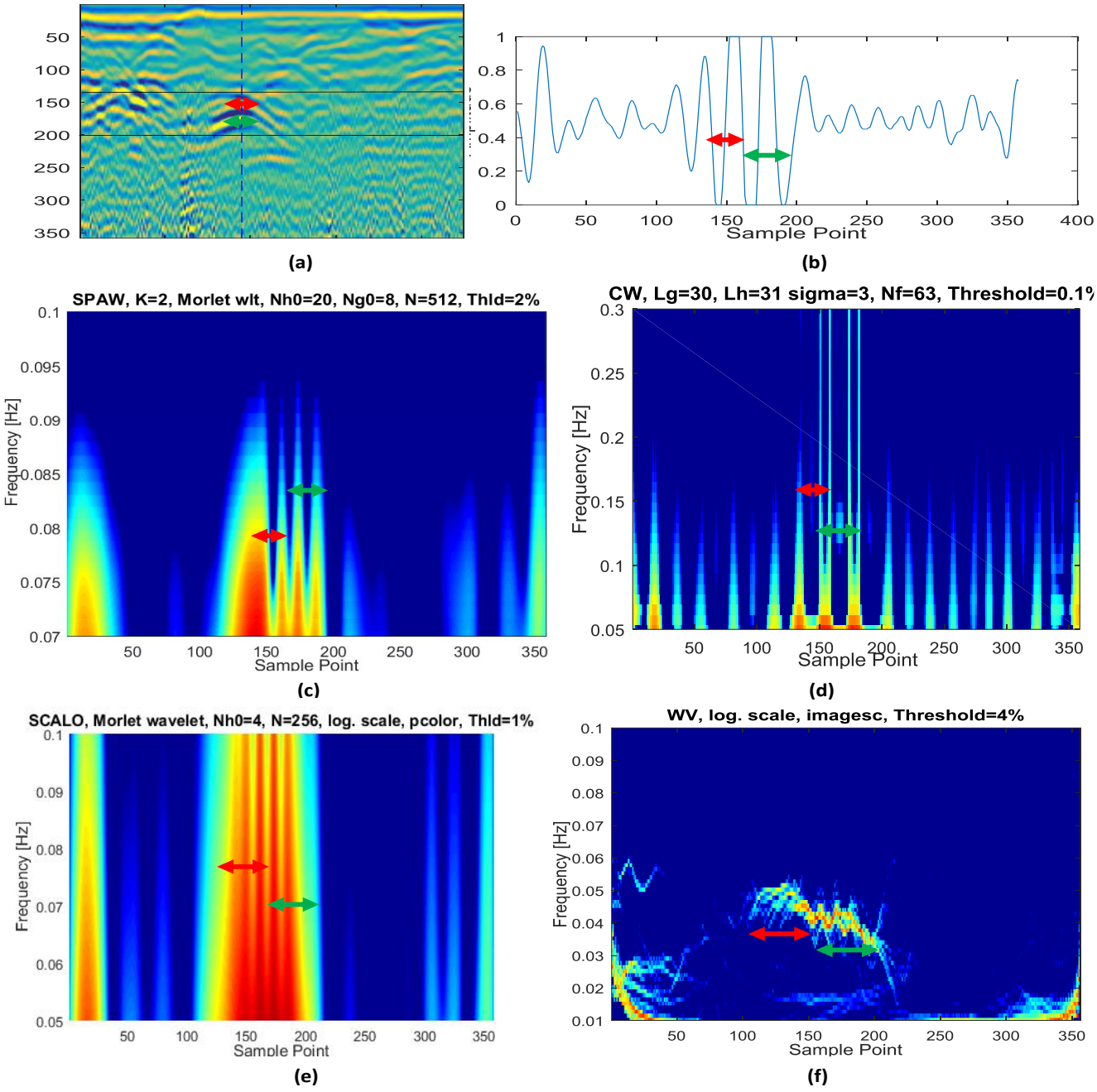


Figure 4. 23: Read intervals within each distribution for the two observed targets (a) the radargram show the considered targets, (b) time signal, (c) SPAWD, (d) CWD, (e) Scalogram, (f) Improved HHT

The proposed technique, to visually read wave propagation time within CWD was easier while considering the time evolution of the signal as well as discarding the noisy peaks. In fact, reading the time propagation using CWD was easier, since the events were clearer and the noise do not overlap the useful signal. The read values are given in table 4.7. The results given by the different time-frequency representations are not very close to each other, especially for the second target. The CWD it is the most convenient one to represent a GPR signal for this dataset.

The results given by the CWD, using this proposed approach, are the most closer one to the measured time between the two reflections for each target.

Table 4. 7 : Time-frequency approach results

Time-frequency distribution	Target number	Propagation time read (r_v)	Propagation time t_p value ($r_v$xdt) in ns
Scalogram	Tg 1	40	3.352
	Tg 2	50	4.190
Improved HHT	Tg 1	50	4.190
	Tg 2	60	5.028
SPAWD	Tg 1	40	3.352
	Tg 2	60	5.028
CWD	Tg 1	50	4.190
	Tg 2	75	6.285

The previously computed depth values are in table 4.4. Due to the relation, given by the Hough transform (mentioned in chapter 3), it must be corrected since the facts reveal that they require the calculated velocity values, which can be viewed in table 4.4. This velocity is an average value between the celerity and the velocity of surrounding medium. The depth is corrected by the formula [201]:

$$d_2 = \frac{V_2 d_1 (C - V_a)}{C(V_a - V_2)} \tag{4.12}$$

d_2 is the depth corrected value, while V_2 is the velocity of propagation within the surrounding medium estimated from the average velocity V_a. d_1 corresponds to the estimation of the position of the air-soil interface reflection out of the data.

Therefore, we can deduce the relative permittivity characterizing every target:

$$\varepsilon_r = \left(\frac{V_a t_p}{d_2}\right)^2 \tag{4.13}$$

The estimated relative permittivity of the medium of the first target is equal to $\varepsilon_{r1} = 4.5053$ and $\varepsilon_{r2} = 3.3340$ for the second target.

4.4. Conclusions

In this chapter, a modified Hough schema was proposed to overcome the drawbacks of the classical Hough algorithm. The approach had corrected the issues concerning random output and high computational time. Nonetheless, the denoising parameters should be carefully chosen in order to achieve satisfactory results.

To conclude, due to the fact that the experimental data (the one provided by Maas and Schmalzl [9]) was requested for the binary classification proposal. Hence, no exact information of the nature of targets was provided within the set of radargrams, to allow as to verify the time-frequency approach results. Nevertheless, the measures of the time between the two reflections for each target is conformed to the results given by the approach. Furthermore, it is necessary to develop an automatic procedure to detected time frequency peaks and deduce the time propagation values.

5. GPR Data Classification

5.1. Introduction

The introduced modified Hough schema presented in chapter 4 overcomes some drawbacks of the classical Hough algorithm. The approach has corrected issues related to the high dimensionality of the accumulator, such as random output problems and large computational requirements. Nonetheless, there are some persistent points relevant to the vote spreading and peaks splitting phenomena. Thus, it is recommended to narrow down the location of hyperbolas to a certain number of small windows in order to apply a fitting algorithm, like the proposed modified Hough transform, over these limited regions.

Several classifiers were designed to undertake this task while special attention was given to the test performance and to the overall accuracy of the automated and semi-automated designed systems. Additionally, these systems must take place in ongoing geophysical surveys in the field in order to provide instant interpretation of the captured data. Hence, the complexity of the classifiers must be taken into consideration as well as validation and test accuracies. This chapter proposes a less complex RBF classifier, which uses HOS cumulants as features to discriminate targets and non-targets in B-scan images. Feature selection was done in two stages: by Mutual information (MIFS) and genetic algorithms (MOGA). MOGA also performed model training and topology selection. Moreover, tests were conducted with two databases.

The outline of this chapter is as follows: Section 5.2 presents a database construction technique from the provided radargrams. To compare the performance obtained, tests were done on the same data with a deep learning technique (CNN); these results are presented in section 5.3. Section 5.4 shows a detailed analysis of the performance of different models acquired using MOGA before selecting a convenient model. Further tests were done on the same data with an SVM classifier, these results are presented in section 5.5. The performance of all these approaches are compared in section 5.6. Conclusions are drawn in section 5.7.

5.2. Database Construction

Two sets of different radargrams were employed. The first set was provided by Maas and Schmalzl [10], and the second set was publicly available in the USGS archives [1]. The first set

of data is unprocessed, while the second set is processed. Two different databases were built from these sets. Further details about the survey configurations and the extraction technique of samples in the databases are provided in the following paragraphs.

5.2.1. Data collection

5.2.1.1. Maas data

This data, provided by Maas and Schmalzl [10], belongs to four different acquisition surveys [10], employing:

1. a Geophysical Survey System "GSSI TerraSIRch System SIRs-3000" with a 50 MHz Butterfly antenna;
2. a "GSSI-5106" with 200 MHz antenna and Wu-King 30 MHz antennas.
3. a "GSSI 5103" with 400 MHz antenna.
4. an antenna type "3101D", 900 MHz frequency.

These 134 radargrams are unprocessed, and only an automatic gain function was applied to them.

5.2.1.2. Dauphin Island data

This GPR data was collected with a "GSSI TerraSIRch System SIRs-3000" with a 200 MHz antenna. Acquisition settings for the surface-surface profiles were set to 64 scans/sec, 20 scans/m, 1024 samples/sec, 16 bits/sample, dielectric constant = 15, range = 200 ns, and an automatic gain value. Surveys were conducted over various terrains. The surveys were either on foot or towed at slow speeds with a device attached behind a vehicle [1].

A series of processing operations were applied to the data [1], which was composed of 62 images:

- An infinite impulse response filter was used (low-pass = 600 MHz and high-pass = 50 MHz) to reduce external interference.
- The distance setting was used to manually calibrate the survey wheel for the terrain by laying out a 10-m-long measured line on the survey surface, which varied between asphalt, sandy dune slopes, and grass.

- Other post-processing operations included applying a static correction, subtracting the mean, removing the header gain, and applying manual automatic gain control.

5.2.2. Databases built

The database construction algorithm consists of three phases: data patching & labelling, pre-processing, and feature extraction. Figure 5.1 shows an overview of the steps of every phase.

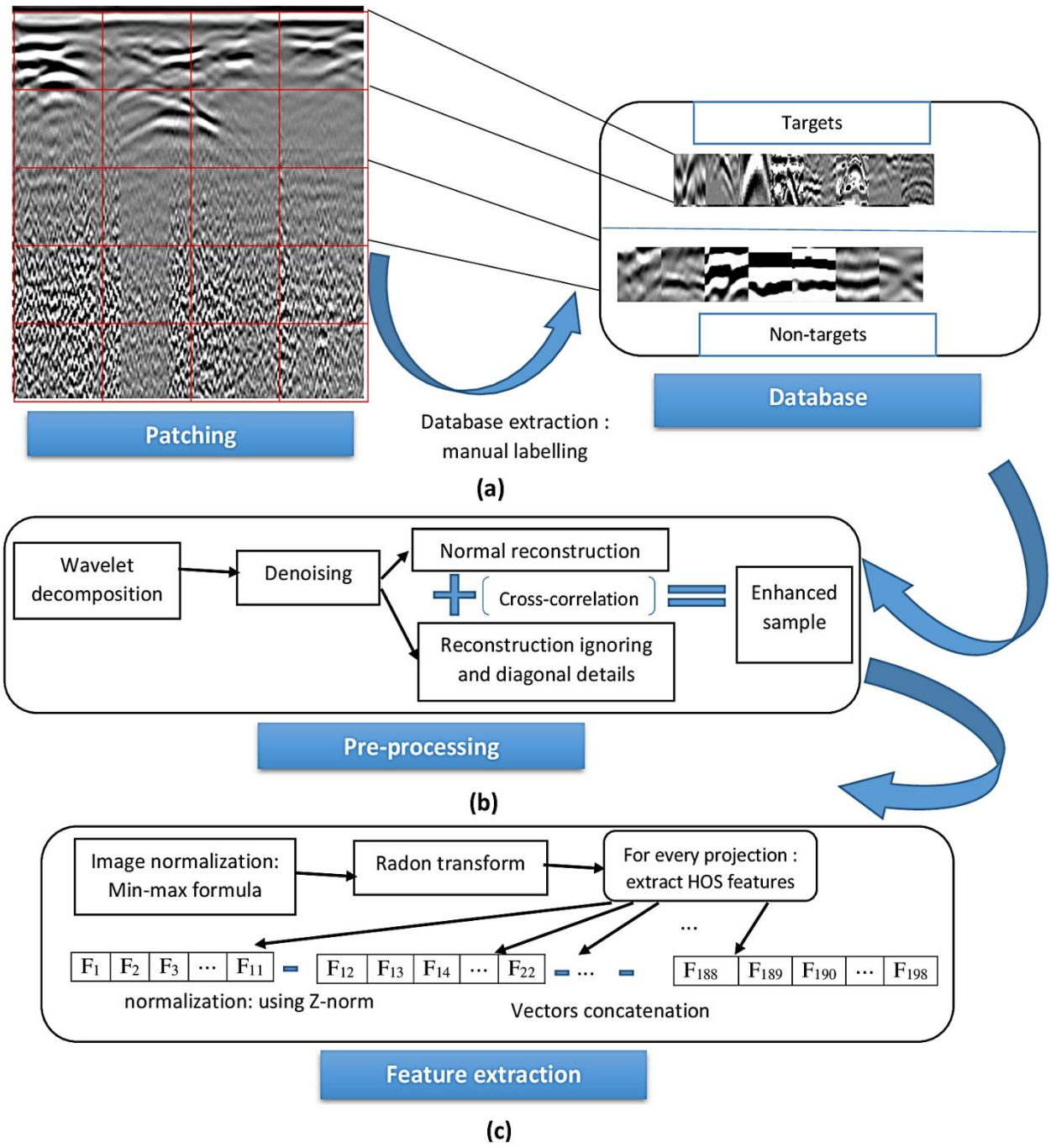


Figure 5. 1 : Database extraction procedure: a) patching & labelling data, b) pre-processing by wavelet denoising, and c) features extraction by HOS algorithm

5.2.2.1. Data patching & labeling

An automatic patching script was used to extract random samples from the set of radargrams, as shown in figure 5.1.a). A resolution of 41x41 pixels is adopted since it gives the best classification results in comparison with other tested resolutions. In fact, using a smaller

patch size does not allow for some larger signatures to be captured, and higher patch sizes impose unnecessary excess data volume.

Only clear and uncut hyperbolas were chosen, while the rest of the targets were selected by the radargrams manually. Moreover, in order to complete the database effectively, a number of non-targets were chosen from the available patches. The degree of mismatch between targets and non-targets must be as higher as possible, i.e. that the samples with labeling doubts are not included in the database. In fact, the available data for training is very limited, so if confusing structures are selected, the results could be unsatisfactory. Afterwards, these samples are manually labelled as zero or one.

Some data augmentation tricks were used to double the number of targets, which consist in simple illumination correction and background extraction operations. Nevertheless, the exactly matched samples, which could unintentionally be missed, are discarded when partitioning the data.

Finally, two unbalanced databases were built: Maas and Dauphin Island (Di). For Maas and Di respectively, the targets represent 37% and 38% of the database (but this rates are reduced to ~30%, regarding the fact that some targets that are detected similar by AproxHull are discarded from the database). The detailed statistics are presented in table 5.1. An example of samples from each database is presented in figure 5.2.

Three data partitions are created by an AproxHull algorithm [191] (AproxH partition): training, validation, and test. All the convex points were included in the training partition, and some samples were selected randomly to complete 60% of all data. The rest of the data was divided randomly between validation and test partitions, with a ratio of 20% for each set.

Further details are provided in table 5.1. Before any processing of the Maas database, all samples were normalized by a Min-Max technique:

$$I' = \frac{I - \text{mean}(I)}{\text{std}(I)} \tag{5.1}$$

I is the input sample intensities, in the format of 41×41 matrix, mean(I) and std(I) are the mean value and standard variation of the pixels intensities respectively, and I' is the normalized sample.

However, for the Dauphin Island database, normalization is not required since the data is - already calibrated and pre-processed.

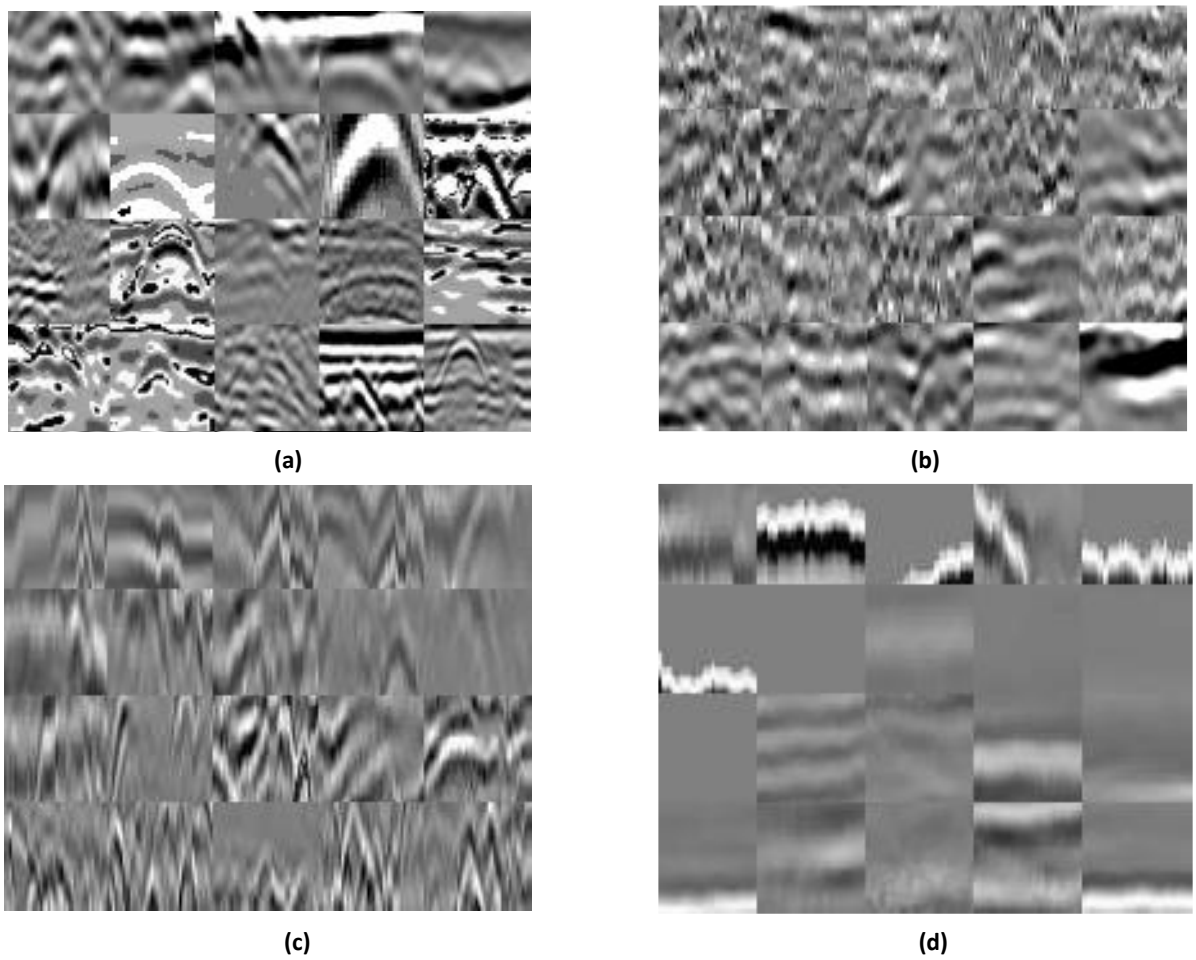


Figure 5. 2: Database samples example: a) targets from Maas database, b) non-targets from Mass database, c) targets from Di database, d) non-targets from Di database

Table 5. 1: Databases statistics

Database name	Number of			Partition size		
	Samples	Targets	Non-targets	Training	Validation	Test
Maas	3186	1202	1984	1881	628	627
Dauphin Island (Di)	3057	1138	1919	1773	592	591

5.2.2.2. Pre-processing

A denoising process, similar to the one proposed in section 5.2.2.2 of chapter 4, is executed for every sample. The global schema is shown in figure 5.1.b). A DWT is performed with a db6 as mother wavelet. Several tests were conducted with other types of wavelets (db4, Sym4, Sym6, Haar, …), but it has been proved that db6 and sym6 are the most recommended ones. The decomposition gives four components: vertical, horizontal, diagonal, and a rescaled

component. Afterwards, vertical and horizontal components are thresholded. The standard Donoho [206] threshold is calculated as:

$$T = \sigma\sqrt{2\log(n)} \tag{5.2}$$

n is the size of the signal and σ is standard deviation of noise.

Two threshold values are estimated, one for a vertical component and another one for a horizontal component. A hard thresholding approach is adopted:

$$\begin{cases} r_t = r & r > T_r \\ r_t = 0 & r < 0 \end{cases} \tag{5.3}$$

A special reconstruction process is then used, and two samples are produced: one sample, reconstructed by a simple inverse DWT algorithm, and a second one, reconstructed while discarding the diagonal and vertical components from an inverse DWT schema. A pixel-by-pixel cross-correlation operation is performed between the two samples to suppress possible background noise and enhance hyperbolas.

5.2.2.3. Feature extraction

High order statistics features, and exact cumulants are the supplied features in this study. Before a feature extraction, it is necessary to convert the samples to 1D signals. This is done through a Radon transform [207], where several projections are computed for every sample for a set of angles between the range [10°:180°] with a step of 10°. Subsequently, for every projection, 1D cumulant slices are calculated by freezing one-time lag and allowing the variation of the second time lag over an adequate interval. For this study case, the maximum time lag value is set to five, and the gathered slice has a length equal to eleven. Afterwards, these slices are normalized by a z-score formula and concatenated to have one feature vector with 198 elements (please refer to figure 5.1.c)).

The projection angle variations could be disadvantageous to the cumulants, which are sensitive to affine shape transformations [208]. To overcome this minor defect, two possible solutions could be applied:

- Normalization and padding of Radon projections by zeros [208].
- Or a normalization of samples by a min-max formulation, before Radon decomposition.

The second scheme is applied to this case since the normalization, with a preserved ratio, does not affect the integrity of data.

Several tests were performed using second, third, and fourth order HOS cumulants; from the results it was found that a third order cumulant is an adequate choice for this classification task.

An example is considered in figure 5.3, which shows a 2D cumulant contour plot of a target (please refer to figure 5.3.a) and 5.3.c)) and another non-target sample (please refer to figure 5.3.b) and 5.3.d)), computed using Swami, Mendel and Nikias [209] HOS toolbox.

All the drawn spectra show two main parts, a kernel, and an external contour.

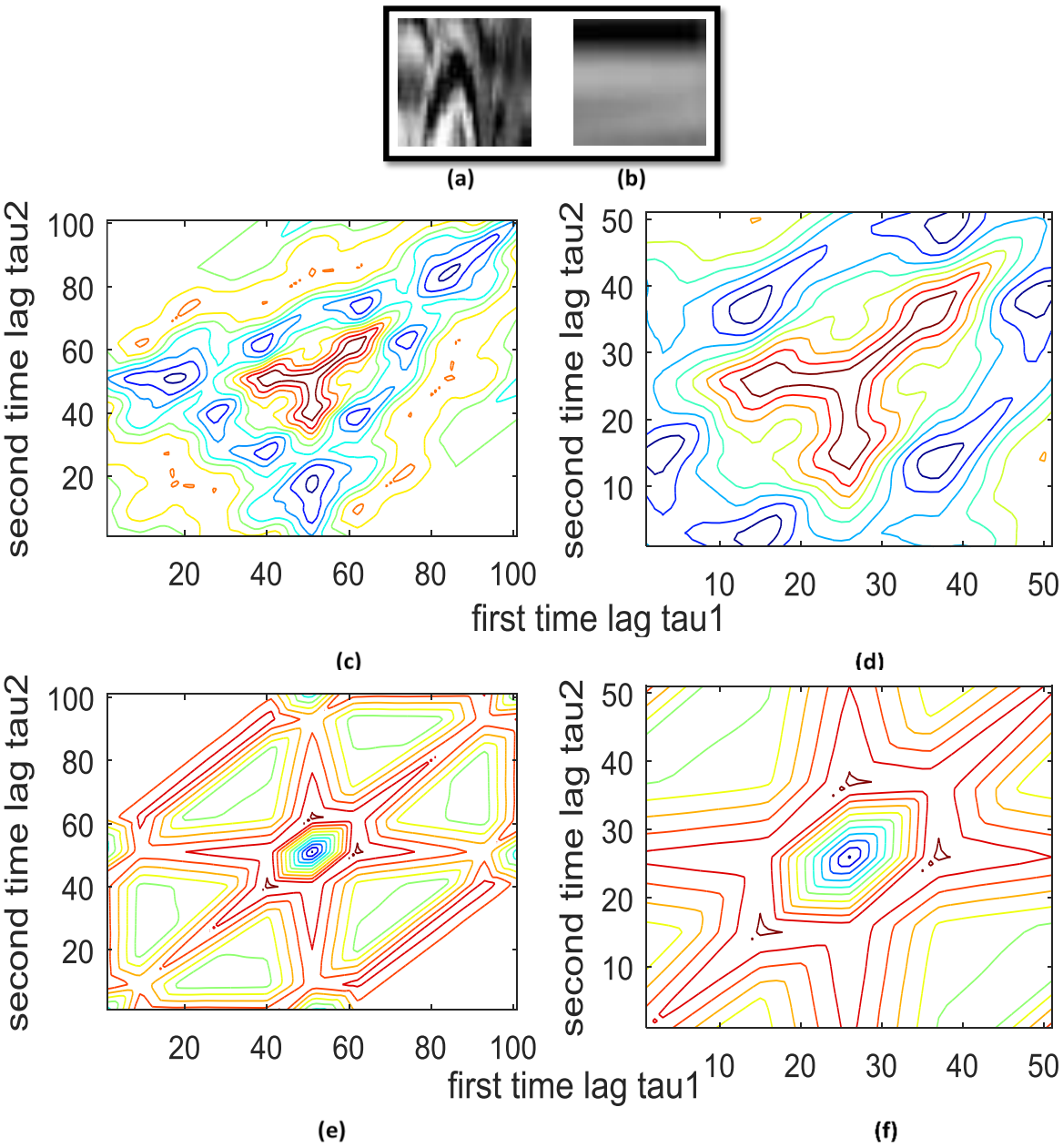


Figure 5. 3 : single projection (180°) HOS third-order HOS cumulants contour plots of two different samples with a different lag parameter values: a) the target, b) the non-target sample. HOS plot of: c)

the target with a maximum time lag equal to 50, d) the same target with a maximum time lag equal to 25, e) the non-target with a maximum time lag equal to 50, f) the same non-target with a maximum time lag equal to 25.

It was clear from the experiments, conducted using different types of targets, that the main difference between targets and non-targets was the shape of the kernel of the cumulant spectrum. Choosing a low maximum time lag allows it to have a zoom over the kernel and capture affine details of this structure. Figures 5.3.e) and 5.3.f) show this structure for a target and a non-target sample.

5.3. Convolutional Neural Network Approach

5.3.1. CNN architecture

During training, CNN input is a fixed-size 41×41 Grayscale image. The only pre-processing done is subtracting the mean Grayscale value, that is computed on the training set, from each pixel. This operation, which assures the stability of the optimization algorithm, is repeated within test sets as well.

An overview of the architecture, which englobes a convolution block for a features extraction and a fully connected block for a classification task, is shown in table 5.2.

Table 5. 2 : CNN architecture

Patches sizes: 41x41							
Conv (5x5x1,20)	**Pool('Max',2x2,2)**	**Conv (5x5x20,50)**	**Pool('Max',2x2,2)**	**Conv (4x4x50,500)**	**RELU**	**Conv (2x2x500,26)**	**Soft-max**
Convolution block					**Fully connected block**		

The convolutional block consists of a stack of three convolutional (Conv) layers where flat and non-flat filters, with a very small receptive field, are used (5x5 and 4x4) to capture the notion of left/right, up/down, and center. The number of deployed filters are 20, 50, and 500 for the three layers, respectively. The first Conv layer packs flat filters, while the second and third ones pack filters reaching a volume of 20 and 50 slices.

Normally, filters preserve the resolution of the input feature map. However, it is often useful to down-sample the output. This can be obtained by using the stride option. In this architecture,

the convolution stride is fixed to 1 pixel, meaning that the spatial padding of Conv layer input is such that the spatial resolution is preserved after convolution.

A spatial pooling is carried out by two max-pooling layers, which follow the first and second Conv. layers. Max-pooling is performed over a 2×2 pixel window, with a stride of 2 pixels (Pool('Max',2x2,2)).

A non-linearity is introduced to the data through a rectification function (RELU) in order to avoid overfitting in the training and to make the decision function more discriminative.

The fully connected block consists of two layers, a Conv layer followed by soft-max layer. The Conv layer employs twenty-six non-flat fitters, of 2x2 size, which reach a volume of 500 slices.

The network is trained by a mini-batch stochastic gradient descendent algorithm with a momentum term [210, 211].

5.3.2. Results

The data was partitioned randomly into three sets (RD partition): training, test, and validation. AproxHull accepts a 1D feature vector as input, so it is impossible to train a CNN based on a partition of data performed by this algorithm.

Several models were trained with different configurations in order to optimize the learning hyper-parameters. For the batch size, it was a bit difficult to set an optimal value regarding learning rate and general training performance. Small values yield a quick convergence at a cost of noise in the training process, while large values yield a slow learning process with accurate estimates of the gradient error. The employed values are shown in table 5.3.

Table 5. 3: CNN Hyper- Parameters

Parameter / Data	Learning rate	Number of epochs	Batch size	Momentum term
Maas	0.0001	30	20	0.9
Di	0.001	20	24	0.9

The simulation is done using the MatConvNet library [207]. The performance of the trained model, for every database, is computed for the three sets of data: Training (tr), test (tt), and validation (v). These statistics are presented in table 5.4 in terms of error rates.

Table 5. 4 : CNN performances over the three partitions of data: training, validation and test

Partition / Data	μ_{CNN}	Training (%)			Test (%)			Validation (%)			Overall (%)		
		FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_{v}	FN_{v}	ACC_{v}	FP_{ov}	FN_{ov}	ACC_{ov}
Maas	43706	10.0736	19.63	86.32	10.91	23.07	84.51	**11.17**	**22.21**	**84.68**	10.47	20.71	85.67
Di	43706	2.68	5.16	96.40	3.90	6.22	95.24	**4.37**	**6.48**	**94.84**	3.23	5.72	95.84

The relation for computing CNN model complexity is given by equation (5.4):

$$\mu_{\text{CNN}} = ((41 \times 41) + 1) \times (\text{number of layers} - 2) \quad (5.4)$$

A 5-fold cross validation strategy is deployed. Due to the fact that the network is stochastic, i.e. it is initialized by random values of weights and biases, a mean of ten values are considered as the final result for each fold. The first database (Maas) achieves a validation accuracy of 84.68%, while the Di database achieves a higher value of 94.84% since the geologists pre-treated the radargrams from where the samples were extracted.

5.4. MOGA-HOS-NN Approach

In order to obtain better results, MOGA is typically executed once with standard parameters and, based on the results obtained, additional experiments can be formulated suing tuned hyper-parameters.

5.4.1. MOGA experiments with Maas database

In the first experiment, with a Maas database (MOGA_Ms_198_1), a vector of 198 features was supplied to the program, and the number of input features (NF) was set between 1 and 198. The system is allowed to choose a number of neurons (NN) between the range [2, 20]. For such a complex problem, the number of generations (NG) was only set to 32 in order to have a global idea of the capability of the algorithm and the requirements of this dataset. The number of training trials (NT) was fixed to 10, while the corresponding maximum number of iterations for early stopping termination criterion (NI) was set to 50. The number of individuals in each generation was set to 100. The proportion of random immigrants was set to 0.1, the selective pressure to 2, and the crossover rate to 0.7. The 'nearest to the origin' strategy was used to select the best training trial. The objective was to simply minimize the five performance

criterions: False positives and negatives over training and a test sets (FP_{tr}, FN_{tr} ;FP_{tt}, FN_{tt}) as well as the complexity (μ).

The statistics of the non-dominated set of models generated by the system for this experiment are shown in table 5.5 in terms of error rates.

Table 5. 5: First experiment with MOGA using 198 features: Statistics of the non-dominated models set generated by MOGA, in terms of rates (%).

	Complexity	Training (%)			Test (%)			Validation (%)			Overall (%)		
		FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_{v}	FN_{v}	ACC_{v}	FP_{ov}	FN_{ov}	ACC_{ov}
Min	6	0	0	67.73	1.18	24.51	71.34	1.13	22.16	74.16	1.61	9.20	69.96
Mean	623.94	5.01	18.03	89.71	9.44	38.10	81.24	9.13	36.97	82.65	6.88	24.63	86.60
Max	3660	27.55	76.54	100	20.99	83.82	87.26	21.04	82.70	87.56	24.34	78.56	92.76

The complexity of RBF is computed by the relation:

$$\mu_{\mathbf{RBF}} = \text{number of neurons in hidden layer}(\text{number of features} + 1) \quad (5.5)$$

These results demonstrate the capability of the algorithm and the expected accuracy range, which can be used in subsequent MOGA executions. A mean validation accuracy of 87.56% could be obtained with a 22.16% FN mean rate on validation. These results can be ameliorated by applying restrictions on objectives of the training set (FP_{tr}, FN_{tr}). From the non-dominated set, two models were picked: one giving high accuracy and low FN rate on validation (Val model), and another one obtaining high overall accuracy (Acc model); the performance results are presented in table 5.6.

Table 5. 6: Models selected from experiment MOGA_Ms_198_1

Best/chosen Model	Model Name	Complexity	Training (%)			Test (%)			Validation (%)			Overall (%)		
			FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_{v}	FN_{v}	ACC_{v}	FP_{ov}	FN_{ov}	ACC_{ov}
ACC model	'model_2971'	1805	0.09	2.49	98.94	10.85	31.86	82.33	11.54	26.49	84.05	4.94	11.54	92.63
Val model	'model_3628'	1560	1.34	5.37	97.02	11.79	24.02	84.24	11.09	21.62	85.80	5.75	11.28	92.22

It was observed from the results that, in general, the majority of the designed models, with a reasonable accuracy, use a number of features larger than 60.

The next experience (MOGA_Ms_198_2) was formulated with the aim to reduce the complexity of the generated models since the two models selected from MOGA_Ms_198_1 (from table 5.6) have a complexity between the range [1350-2200]. For MOGA_Ms_198_2, the number of generation was fixed to 47; number of iteration for early stopping criterion to 20, number of training trials to 5, and the number of input features was set between the range [80, 198]. For the rest of parameters, they are the same, as already mentioned for MOGA_Ms_198_1, for all the experiments with the Maas database. From the results of MOGA_Ms_198_2, two other models were selected; one with higher performance on validation and another one considering the overall accuracy rate. The chosen models are 'model_2971' and 'model_3628'. Their FP, FN, and accuracy, conducted over the three partitions of data, are given in table 5.7.

Table 5. 7: The experiments run within Maas database: On the first two experiments (MOGA_198_1 and MOGA_198_2), there are no restrictions on FP and FN for training and test.

Best/chosen Model	Model Name	Complexity	Training (%)			Test (%)			Validation (%)			Overall (%)		
			FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_{v}	FN_{v}	ACC_{v}	FP_{ov}	FN_{ov}	ACC_{ov}
ACC model	'model_2398'	**2208**	0.45	3.28	98.40	10.85	25.98	84.24	10.41	28.11	84.37	4.89	11.28	92.76
Val model	'model_557'	**1356**	0.89	3.93	97.87	11.79	28.92	82.64	11.99	22.70	84.85	5.69	11.37	92.22

It is clear that the results are approximately identical to the models selected in MOGA_Ms_198_1 (please refer to table 5.8).

Table 5. 8: Models selected from experiments MOGA_Ms_198_1 and MOGA_Ms_198_2

Model name	Validation performance	
	FP_{v}	FN_{v}
'model_2398'	53	42
'model_557'	46	52
'model_2971'	51	49
'model_3628'	49	40

In order to formulate the next experiment MOGA_Ms_198_3, the four models (please refer to table 5.8) should be taken into consideration.

It is preferable to first select the models providing higher results in validation, and then try to improve the test performance by setting appropriate restrictions on the objectives of the

training set. Hence, according to the values shown in table 5.8, the FP_{tt} was chosen to be less than 53 and the FN_{tt} less than 41. The same parameters of MOGA_Ms_198_1 are kept, except for the NG set to 100, NF in the range [1, 80], and NI set to 50. A fourth experiment was conducted within this database with 198 features as well (please refer to table 5.9).

Table 5. 9: Performances over validation partition for models selected from experiment MOGA_Ms_198_1and MOGA_Ms_198_2 in terms of number of samples

Experiment	Objectives/Restrictions	Parameters
MOGA_Ms_198_1	Complexity; FP_{tr}, FN_{tr}; FP_{tt}, FN_{tt}	NG 32; NF [1, 198]; NN [2 20]; NI 50; NT 10
MOGA_Ms_198_2	Complexity<2000; FP_{tr}, FN_{tr}; FP_{tt}, FN_{tt}	NG 47; NF [80, 198]; NN [2 20]; NI 20; NT 5
MOGA_Ms_198_3	Complexity; FP_{tr}, FN_{tr}; FP_{tt}<53, FN_{tt}<41	NG 100; NF [1, 80]; NN [2 20]; NI 50; NT 5
MOGA_Ms_198_4	Complexity; FP_{tr}<150, FN_{tr}<100; FP_{tt}<53, FN_{tt}	NG 100; NF [1, 80]; NN [2 20]; NI 50; NT 5
MOGA_Ms _80_1	Complexity; FP_{tr}, FN_{tr}; FP_{tt}<54, FN_{tt}<38	NG 100; NF [1, 80]; NN [2 20]; NI 50; NT 5
MOGA_Ms _80_2	Complexity; FP_{tr}<150, FN_{tr}<100; FP_{tt}<54, FN_{tt}<38	NG 100; NF [1, 80]; NN [2 20]; NI 50; NT 5
MOGA_Ms _80_3	Complexity; FP_{tr}<150, FN_{tr}<100; FP_{tt}<54, FN_{tt}<38	NG 100; NF [80, 80]; NN [2 20]; NI 50; NT 5
MOGA_Ms _80_4	Complexity<1000; {FP_{tr}<50, FN_{tr}<50, with priority 2}; {FP_{tt}<54, FN_{tt}<37, with priority 1}	NG 100; NF [80, 80]; NN [2 20]; NI 50; NT 5
MOGA_Ms _80_5	Complexity<1000; {FP_{tr}<50, FN_{tr}<50 }, with priority 2}; FP_{tt}, {FN_{tt}<40, with priority 1}	NG 100; NF [1, 80]; NN [2 20]; NI 50; NT 5
MOGA_Ms _80_6	Complexity<1000; FP_{tr}<50, FN_{tr}<50; FP_{tt}, FN_{tt}<38	NG 100; NF [1, 80]; NN [2 20]; NI 50; NT 5

Afterwards, regarding the computational requirements of the genetic algorithm used for feature selection, it was preferable to reduce the number of features supplied to MOGA. The idea was to perform a feature selection in two stages. First, eighty features were selected from the 198 by the mutual information approach (MIFS). Here, a value of 0.9 is adopted for the regularization parameter, and the data is normalized by a z-score technique before applying the MIFS algorithm. Only those 80 selected features are then supplied to MOGA to perform the second feature selection stage by the genetic algorithm.

By supplying to MOGA another dataset with 80 features and 3186 samples (after feature reduction the data is partitioned by AproxHull tool again), six other experiments were carried out but with different parameters and new restrictions on objectives. The formulation of these problems (objectives and restrictions on objectives), in addition to the simulation parameters are listed in table 5.8. In MOGA_Ms_80_1, restrictions were set on FP_{tt} and FN_{tt} according to the results of MOGA_Ms_198_4.

Moreover, in MOGA_Ms_80_3 and MOGA_Ms_80_4, the system was forced to use all the feature sets to measure the quantity of information held in this dataset in comparison with the initial dataset that is within 198 features. In experiments MOGA_Ms_80_4 and MOGA_Ms_80_5 the objectives associated the training set have higher priority over other objectives because it is required to have higher results on validation and specifically low false negative (FN_v) rates for this application.

From every conducted experiment, the model that gives higher results over the validation partition is picked. The results are summarized in table 5.10. The desired models must satisfy a number of requirements: low FN_v, higher validation accuracy, and higher test accuracy as well. 'model_7744' is the model that obtains the best results within the dataset of 198 features, while 'model_6942' is the one showing best results over the experiments carried out with the dataset of 80 features. The model 'model_6942', designed within the dataset of 80 features, is less complex and gives a higher performance on the validation and test sets.

Table 5. 10: The performances of the selected models achieving higher accuracy ACC_v and low FN_v from every experiment conducted with the Maas database. The models highlighted in blue are the ones selected for every dataset

Experiment	Model Name	Complexity	Training (%)			Test (%)			Validation (%)		
			FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_v	FN_v	ACC_v
MOGA_Ms_198_1	'model_557'	1356	0.89	3.93	97.87	11.79	28.92	82.64	11.99	22.70	84.85
MOGA_Ms_198_2	'model_3628'	1560	1.34	5.37	97.02	11.79	24.02	84.24	11.09	21.62	85.80
MOGA_Ms_198_3	'model_7744'	696	1.70	7.21	96.07	10.85	29.41	83.12	**10.18**	**21.62**	**86.44**
MOGA_Ms_198_4	'model_3450'	567	2.24	7.99	95.43	9.67	29.41	83.92	7.24	24.86	87.56
MOGA_Ms _80_1	'model_6942'	567	2.93	10.68	93.67	7.17	37.01	85.51	**7.285**	**20.69**	**88.99**
MOGA_Ms _80_2	'model_9013'	616	1.89	7.40	95.69	6.75	35.06	86.30	8.83	20.11	88.04
MOGA_Ms _80_3	'model_484'	650	2.84	7.77	95.00	7.17	36.36	85.67	9.49	20.69	87.4
MOGA_Ms _80_4	'model_2807'	912	1.13	4.85	97.24	9.28	37.01	83.917	9.71	22.41	86.76
MOGA_Ms _80_5	'model_7902'	795	0.85	3.52	97.98	10.13	33.77	84.076	13.47	21.26	84.37
MOGA_Ms _80_6	'model_1432'	870	0.66	2.67	98.46	13.71	38.31	80.255	10.15	19.54	87.24

5.4.2. MOGA experiments with Dauphin Island database

The MOGA_Di_198_1 experiment is not the first experiment within this database. Previous experiments were organized with 198, 80 and 50 features. It was realized that for this database, it is sufficient to work only with 30 features.

Hence, for MOGA_Di_198_1, a vector of 198 features was supplied to MOGA framework and the number of input features (NF) were set between 1 and 30. The system is

allowed to choose a number of neurons (NN) ranging between [2, 10]. The number of generations (NG) was fixed at 100 and number of training trials (NT) at 10, while the corresponding maximum number of iterations for early stopping termination criterion (NI) was set to 50. The number of individuals in each generation was set to 100. The proportion of random immigrants was set to 0.1, the selective pressure to 2, and the crossover rate to 0.7. The nearest to the origin strategy was used to select the best training trial.

The [min, max, mean] statistics of [FP, FN, ACC] rates, over the three partitions as well as over the entire database, for the non-dominated set of generated models are presented in table 5.11.

Table 5. 11: MOGA_Di_198_1 experiment using 198 features: Statistics of the set of non-dominated solutions generated by MOGA, in terms of rates (%).

	Complexity	Training (%)			Test (%)			Validation (%)			Overall (%)		
		FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_{v}	FN_{v}	ACC_{v}	FP_{ov}	FN_{ov}	ACC_{ov}
Min	6	1.65	2.36	70.16	0.45	6.62	68.02	1.91	5.78	66.72	1.53	5.75	69.45
Mean	94.449	8.00	10.34	90.93	6.32	19.02	90.43	8.22	21.50	87.9	7.65	13.21	90.22
Max	300	43.22	48.39	97.8	39.77	61.59	94.58	42.48	70.52	93.41	41.84	53.36	95.36

It is clear that with the Di database, it is possible to attend a validation accuracy of 93.41% and an overall accuracy of 95.36% with a very low complexity (max 300). From the non-dominated solutions, three models were chosen; one achieving higher results on validation (low FN_v rate and high ACC_v) and two others having the best overall accuracy. Table 5.12 presents the two selected models which provide approximately the same overall performance; ‘model_2522’ and ‘mode_9863’.

Table 5. 12: Models selected from experiment MOGA_Di_198_1

Best/chosen Model	Model Name	Complexity	Training (%)			Test (%)			Validation (%)			Overall (%)		
			FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_{v}	FN_{v}	ACC_{v}	FP_{ov}	FN_{ov}	ACC_{ov}
ACC model	**'model_2522'**	**260**	3.31	3.72	96.50	3.82	13.29	93.41	3.86	12.58	93.91	6.37	3.56	95.36
	'model_9863'	300	1.65	3.35	97.57	5.25	18.50	90.88	4.54	13.24	93.23	6.99	3.18	95.36
Val model	**'model_2473'**	**230**	4.14	4.22	95.83	6.44	13.87	91.38	5.45	8.61	93.74	6.28	4.98	94.52

‘model 2522’ is chosen because it is the one with the smaller complexity. The performances over training and test sets, in terms of errors, for the chosen model, in addition to ‘model_9863’, are given in table 5.13.

Table 5. 13 : Performances over training and test sets for models selected from experiment MOGA_Di_198_1, the value represents the number of errors

Model name	Training performance		Test performance	
	FP_{tr}	FN_{tr}	FN_{tt}	FN_{tt}
'model_2522'	32	30	17	19
'model_2473'	40	34	24	13

The restrictions employed case for the next experiment are FP_{tr}<32, FN_{tr}<30, and FP_{tt}<**17**, FN_{tt}<**19**. However, it was thought inappropriate at this point to force the system with these small values, especially in terms of the validation constraints. The restrictions for MOGA_Di_198_2 are adjusted by taking into consideration ‘model_2473’: FP_{tr}<40, FN_{tr}<40. The restriction values for test partition are fixed to FP_{tt}<**30**, FN_{tt}<**30** to relax first our restriction on the test set. For MOGA_Di_198_2, the simulation parameters are the same as for MOGA_Di_198_1. The selected models from the non-dominated solutions, generated by MOGA_Di_198_2, are shown in table 5.14.

Table 5. 14 : Models selected from experiment MOGA_Di_198_2

Best/chosen Model	Model Name	Complexity	Training (%)			Test (%)			Validation (%)			Overall (%)		
			FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_{v}	FN_{v}	ACC_{v}	FP_{ov}	FN_{ov}	ACC_{ov}
ACC model	'model_5473'	290	2.17	2.36	97.74	4.77	15.61	92.06	3.64	13.91	93.73	5.93	3.12	95.80
	'model_8332'	**270**	1.65	2.36	98.03	4.53	15.03	92.40	5	14.57	92.55	5.93	3.12	95.80
Val model	**'model_9355'**	**300**	2.17	3.60	97.18	5.73	14.45	91.72	5.23	7.28	94.25	5.75	3.72	95.50

From the results, with comparison of the solutions generated by MOGA_Di_198_1, the validation performance is ameliorated while the training results are worse. Thus, to formulate another experiment, FP_{tr} and FN_{tr} of ‘model_2522’ from MOGA_Di_198_1 (please refer to table 5.12) are considered. In terms of FP_{tt} and FN_{tt}, the results of model_8332 from MOGA_Di_198_2 (please refer to table 5.15), are used as well to formulate MOGA_198_Di_3.

Table 5. 15 : Performances over training and test sets for models selected from experiment MOGA_Di_198_2, the value represents number of errors

Model name	Training performance		Test performance	
	FP_{tr}	FP_{tr}	FN_{tt}	FN_{tt}
'model_8332'	16	19	22	22
'model_9355'	21	29	23	11

Hence, MOGA_Di_3 restrictions over the training set are FP_{tr}<40, FN_{tr}<40 and validation partition are FP_{tt}<30, FN_{tt}<30. Additionally, three other experiments were conducted using the Dauphin Island dataset of 198 features. The corresponding restrictions and simulation parameters for every experiment are presented in table 5.16 (experiments MOGA_Di_198_4-6). The other parameters that are not mentioned in the table are the same as for MOGA_Di_198_1.

Subsequently, using MIFS, another dataset was created with only 50 features. The regularization parameter of the MIFS algorithm was set to 0.9 for this data as well. With this dataset, six other experiments were formulated MOGA_Di_50_1-6 (please refer to table 5.16).

Table 5. 16: The experiments run within Dauphin Island database

Experiment	Objectives/Restrictions	Parameters
MOGA_Di_198_1	Complexity; FP_{tr}, FN_{tr}; FP_{tt}, FN_{tt}	NG 100; NF [1, 30]; NN [2 10]; NI 50; NT 10
MOGA_Di_198_2	Complexity; FP_{tr}<40, FN_{tr}<40; FP_{tt}<30, FN_{tt}<30	NG 100; NF [1, 30]; NN [2 10]; NI 50; NT 10
MOGA_Di_198_3	Complexity; FP_{tr}<30, FN_{tr}<30; FP_{tt}<20, FN_{tt}<20	NG 100; NF [1, 30]; NN [2 10]; NI 50; NT 10
MOGA_Di_198_4	Complexity; FP_{tr}<40, FN_{tr}<40; FP_{tt}<20, FN_{tt}<20	NG 100; NF [1, 30]; NN [2 10]; NI 50; NT 10
MOGA_Di_198_5	Complexity; FP_{tr}, FN_{tr}; FP_{tt}<20, FN_{tt}<20	NG 100; NF [1, 30]; NN [2 10]; NI 50; NT 10
MOGA_Di_198_6	Complexity; FP_{tr}<50, FN_{tr}<50; FP_{tt}<20, FN_{tt}<20	NG 100; NF [1, 30]; NN [2 10]; NI 50; NT 10
MOGA_Di_50_1	Complexity; FP_{tr}, FN_{tr}; FP_{tt}, FN_{tt}	NG 100; NF [1, 30]; NN [2 10]; NI 50; NT 10
MOGA_Di_50_2	Complexity; FP_{tr}<40, FN_{tr}<40; FP_{tt}<20, FN_{tt}<20	NG 100; NF [1, 30]; NN [2 10]; NI 50; NT 10
MOGA_Di_50_3	Complexity; FP_{tr}<50, FN_{tr}<50; FP_{tt}<15, FN_{tt}<15	NG 100; NF [1, 30]; NN [2 10]; NI 50; NT 10
MOGA_Di_50_4	Complexity; FP_{tr}, FN_{tr}; FP_{tt}<54, FN_{tt}<38	NG 100; NF [1, 50]; NN [2 15]; NI 50; NT 10
MOGA_Di_50_5	Complexity; FP_{tr}<57, FN_{tr}<43; FP_{tt}<54, FN_{tt}<38	NG 100; NF [1, 30]; NN [2 15]; NI 50; NT 10
MOGA_Di_50_6	Complexity; FP_{tr}<33, FN_{tr}<43; FP_{tt}<19, FN_{tt}<15	NG 100; NF [1, 30]; NN [2 15]; NI 50; NT 10

The parameters, which are not cited in the table, are set up the same as the MOGA_Di_50_1.

The first experiment was formulated without any restrictions. The number of FP and FN over training and test partitions for two selected models, one with high validation performance ('model_7437') and one with best overall accuracy ('model_7885'), are shown in table 5.17.

According to these results, the restrictions for MOGA_Di_50_2 are formulated as follow: FP_{tr}<40, FN_{tr}<40; FP_{tt}<20, FN_{tt}<20. The restriction values of FN on training and test are slightly relaxed not to restrict excessively the objective space. Further experiments (MOGA_Di_50_3-6) are formulated while taking into account the results of the previous dataset as well.

Table 5. 17: Performances over training and test sets for models selected from experiment MOGA_Di_80_1

Model name	Training performance		Test performance	
	FP_{tr}	FN_{tr}	FN_{tt}	FN_{tt}
model_7437	39	32	26	10
model_7885	26	33	22	15

For each experiment organized with this database, the model achieved the best performance over validation partition was chosen. These selected models are shown in table 5.18 with their corresponding performance rates over the three sets of data.

'model_6917' performs better than other models over validation partition for the dataset within 198 features. However, this model provides the worst results over the test set. Thus, 'model_5439' is the one selected over the models with 198 features. The selected model 'model_3434', within the dataset of 50 features, outperforms other models over the validation and test sets.

Table 5. 18: FP, FN and ACC over training test and validation partitions, in addition to overall performance, of the selected models from every experiment run within Dauphin Island database. The models highlighted in blue are the ones selected for every dataset

Experiment	Model Name	Complexity	Training (%)			Test (%)			Validation (%)		
			FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_{v}	FN_{v}	ACC_{v}
MOGA_Di_198_1	'model_2473'	230	4.14	4.22	95.83	6.44	13.87	91.38	5.45	8.61	93.74
MOGA_Di_198_2	'model_9355'	300	2.17	3.60	97.18	5.73	14.45	91.72	5.23	7.28	94.25
MOGA_Di_198_3	'model_7472'	300	2.07	3.35	97.35	6.20	16.18	90.88	3.86	7.95	95.09
MOGA_Di_198_4	'model_5075'	300	2.69	3.97	96.73	4.30	17.92	91.72	3.64	7.28	95.43
MOGA_Di_198_5	'model_5439'	156	4.45	4.84	95.37	5.01	13.29	92.57	**4.32**	**6.62**	**95.09**
MOGA_Di_198_6	'model_6917'	145	5.89	5.71	94.12	5.73	16.18	91.22	4.09	5.30	95.60
MOGA_Di_50_1	'model_7437'	260	4.25	3.74	95.99	5.78	14.4	92.40	5.89	6.66	93.90
MOGA_Di_50_2	'model_9345'	240	3.38	4.09	96.27	4.71	16.00	92.90	4.53	8.66	94.41
MOGA_Di_50_3	'model_3434'	208	4.25	6.32	94.75	5.14	16.8	92.39	**3.17**	**9.33**	**95.26**
MOGA_Di_50_4	'model_6964'	630	1.52	2.34	98.08	7.07	17.6	90.71	7.03	7.33	92.89
MOGA_Di_50_5	'model_4388'	286	3.38	3.63	96.50	5.35	16.8	92.23	6.12	6.66	93.74
MOGA_Di_50_6	'model_2457'	299	2.50	3.63	96.95	4.92	18.4	92.23	4.080	10.00	94.42

‘model_3434’ have a higher complexity compared with ‘model_5439’. Moreover, ‘model_3434’ obtains less misclassification rates over validation and test sets.

5.5. HOS-SVM Approach

Several tests were conducted using SVM with a Gaussian kernel, on the two partitions of data, the random (RD) partition used by CNN and AproxHull one used by MOGA, to compare the results. These experiments could be grouped as follows:

1) The 198 features vector is supplied to SVM without any previous feature selection;

2) The 80/50 features selected by MIFS are supplied to SVM;

3) The selected set of features used by the best model designed by MOGA are tested with SVM.

The performance of SVM crucially depends on the chosen penalty and spread parameters of the kernel. A grid search approach is used to optimize these two parameters. The search is done within this set of values: {0.003, 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, 100, 300}. The tests were conducted within 121 combinations. The best parameters were chosen for every experiment, depending on the validation partition performance (low FN rate and higher accuracy) and are shown in table 5.19. The features are normalized by a z-score technique before supplying them to the classifier. Nb_F make reference to the number of features supplied to MOGA, i.e. 198 or 80/50 (Maas/Di) features.

Table 5. 19: SVM Gaussian kernel optimized penalty and spread parameters for the run experiments

	Random partition used by CNN			**AproxHull partition used by MOGA**		
Parameter / Data	**Nb_F**	**Penalty term**	**Spread Parameter**	**Nb_F**	**Penalty term**	**Spread parameter**
Maas	**198**	100	**0.01**	**198**	100	**0.01**
	80	10	**0.1**	**80**	100/300	0.1
Di	**198**	30	**0.01**	**198**	300	0.003
	50	10	**0.3**	**50**	1	0.3

The results for the first and second experiments, for the RD partition, are presented in table 5.20, while the performances collected within the AproxHull partition are given in table 5.21.

Table 5. 20 : SVM performances for the random partitioned data used by CNN

Partition / Data	Nb_F	μ_{SVM}	Training (%)			Test (%)			Validation (%)		
			FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_v	FN_v	ACC_v
Maas	**198**	193089	0	0	100	12.14	21.16	84.46	10.79	21.21	85.28
	80	91712	0	0.17	99.94	10.28	19.83	**86.12**	9.225	20.80	**86.41**
Di	**198**	103950	0.36	1.23	99.31	4.90	10.20	93.13	4.38	9.76	93.62
	50	27660	0.24	1.02	99.47	5.26	10.64	**92.73**	5.21	9.93	**93.03**

Table 5. 21: SVM performances for the data partition performed by AproxHull algorithm and used by MOGA

Partition / Data	Nb_F	μ_{SVM}	Training (%)			Test (%)			Validation (%)		
			FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_v	FN_v	ACC_v
Maas	**198**	155628	0	1.97	99.20	8.96	30.88	83.92	10.18	25.95	85.17
	80	77840	0	0.61	99.73	7.81	32.47	**86.15**	6.84	24.71	**88.20**
Di	**198**	75240	1.03	2.60	98.25	3.82	16.18	92.57	3.18	16.55	93.40
	50	40350	2.18	3.98	96.95	3.43	18.4	**93.41**	3.85	16.67	**92.89**

The two tables present the performances regarding the three sets: training, validation, and test separately.

As it is expected, results are better using the AproxHull partition, in general, since all the convex hull points are included in the training set. Moreover, within SVM, the training accuracy is always higher in comparison with any other classifier. However, the designed models are very complex.

The model complexity is computed by the formula:

$$\mu_{\mathrm{SVM}} = \text{number of support vectors} \times \text{number of features} \tag{5.6}$$

Moreover, the selected set of features, represented by the best model designed by MOGA, are tested with an SVM classifier for each database. The results are shown in table 5.22 with the corresponding optimized kernel parameters.

Table 5. 22 : SVM performances for the best model designed by MOGA

Partition / Data	Data partition	Penalty/spread	μ_{SVM}	Training (%)			Test (%)			Validation (%)		
				FP_{tr}	FN_{tr}	ACC_{tr}	FP_{tt}	FN_{tt}	ACC_{tt}	FP_v	FN_v	ACC_v
Maas	**RD**	30/0.03	54243	0.63	8.86	96.14	12.12	29.65	80.95	12.72	28.23	81.19
	AproxHull	30/0.1	56574	0	1.94	99.15	5.70	34.42	87.26	**6.18**	**26.44**	**88.20**
Di	**RD**	30/1	22200	0.26	3.02	98.71	9.96	10.99	89.66	10.73	11.43	89
	AproxHull	300/1	19425	0	0	100	0.21	0	99.83	**0**	**0**	**100**

It is clear from the results that the AproxHull partition provides higher results on testing, which is a normal output of a well-trained model. SVM obtains similar, or even slightly, better results with the features selected by MOGA, in comparison with the results given by RBF MOGA model, although with very complex models. The SVM models using 198 features yield worse results for both databases, with models of larger complexity.

A detailed comparison of the results will be done in section 5.6.

5.6. Results Comparison

In order to study the performance and demonstrate the capabilities of the MOGA algorithm, a comparison of the different results, presented in the previous sections of this chapter, are explained and justified in this section.

5.6.1. Feature usage

5.6.1.1. MOGA models

The first experiment, with the first MOGA_Ms_198_1 database, was conducted for 32 generations. In this experiment, the usage rate of every feature, with the generated non-dominated solutions, was calculated. The 63 most used features were selected. Their indices are given in section 3 of table 5.23.

Table 5. 23 : Feature selection process for Maas database: (1) Indices of features selected by MIFS, (2) Indices of features selected by the best model designed by MOGA ('model_6942', MOGA_Ms _80_1; please refer to table 4.10), (3) Indices of most used features by MOGA models in the first experiment, with 198 features. The set of features (2) is selected from the set (1) since in MOGA_Ms_80_1 80 features selected by MIFS were supplied to MOGA.

(1) Features selected by MIFS																			
3	**4**	8	**9**	13	14	15	19	20	25	**30**	**31**	36	38	42	47	48	52	**53**	58
59	63	64	68	69	**70**	74	**75**	**76**	**79**	**80**	**81**	85	86	87	90	91	**92**	96	97
98	101	102	103	107	108	109	**112**	113	114	**118**	**119**	**120**	123	124	125	129	130	135	**136**
140	141	146	151	**152**	157	162	163	168	**169**	173	174	**179**	**180**	184	185	190	**191**	**195**	196
(2) Features used by the best model designed by MOGA																			

3	8	9	13	14	15	31	36	41	42	47	48	52	53	58	59	63	64	68	69
74	75	76	79	80	81	85	87	90	91	92	96	97	101	102	108	118	119	120	123
124	125	129	130	135	141	146	151	152	157	162	163	168	169	173	174	180	184	185	190
191	195	196																	
(3) Most used features by MOGA models in the first experiment, with 198 features																			
1	**3**	**4**	**9**	11	17	29	**30**	**31**	32	40	45	46	**50**	51	53	62	65	66	67
70	71	72	**75**	**76**	**79**	**80**	**81**	82	83	**92**	95	100	104	**112**	**118**	**119**	**120**	122	132
133	**136**	143	145	147	150	**152**	154	155	161	167	**169**	170	**179**	**180**	183	188	189	**191**	192
193	**195**	198																	

There is a correspondence of 38.1% between the most selected features by non-dominated models of MOGA_Ms_198_1 and the features selected by the best model 'model_6942' of MOGA_Ms _80_1 (the common features are highlighted in bold), i.e. 38.1% features are already identified in the first experiment MOGA_Ms_198_1. This point demonstrates the efficiency of this algorithm, in regards to feature selection and discrimination, since MOGA_Ms_198_1 was conducted only for 32 generations.

5.6.1.2. SVM models

An experience with the entire set of features, selected by MIFS, that was supplied to SVM was conducted as well. The selected model draws a validation accuracy rate of 86.41% with an FN_v rate of 20.8% within a RD partition used by CNN, an FN_v rate of 24.71%, and a validation accuracy of 88.20% within the AproxHull partition used by MOGA (please refer to table 5.24).

Nevertheless, MOGA_Ms_198_3 and MOGA_Ms_198_4 were forced to use all the 80 features selected by MIFS in the experiments. MOGA_198_3 had restrictions as did MOGA_Ms_198_4 as well, while MOGA_Ms_198_4 had restrictions over the training set that was prioritized over others. The selected model from MOGA_Ms_198_3 gives slightly higher results, 87.4% validation accuracy, and 20.69% FN validation rate.

The results provided by MOGA are less than the one provided by SVM (88.20%>87.4%) in term of validation accuracy, but the FNv is very high 24.71% with SVM model with AproxHull partition (please refer to table 5.24).

Table 5. 24 : Experiments run over Maas database using a set of 80 features with different feature selection- algorithms, with MOGA and SVM classifiers

Classifier	**Nb_F**	**Feature selection algorithm**	**Data Partition**	**Complexity**	**Validation (%)**		
					FP_v	**FN_v**	**ACC_v**

SVM	80	MIFS	RD one (used by CNN)	91712	9.225	20.80	86.41
SVM		MIFS	AproxHull	77840	6.84	**24.71**	**88.20**
MOGA		MOGA, features of 'model_484' from experiment 'MOGA_Ms_80_3'	AproxHull	650	9.49	**20.69**	**87.4**

Nonetheless, the features selected by the best models given by MOGA for the two databases, 'model_6942' from MOGA_Ms _80_1 (FNv rate of 20.69% and ACCv of 88.9%) and 'model_3434' from MOGA_Di _50_3 (FNv rate of 9.33% and ACCv of 95.26%), are supplied to SVM (please refer to table 5.25). Within the same data partition used by MOGA, the model selected for the Maas database obtained an ACCv of 88.20% with an FNv rate of 26.44%, while the model selected for the Maas Di obtained ACCv of 100% with an FNv rate of 0% (please refer to table 5.25). SVM gives worse results with a huge complexity for the Maas database, while for the Di database, SVM provides higher results since the training process is perfect with SVM but the complexity is huge. Otherwise, the results with random partition are worse with SVM since the training process with the Approximull partition is better, which was already clarified and pointed out previously.

Table 5. 25 : Experiments run with SVM classifier using the same set of features selected by the best model designed by MOGA for the Maas and Di databases

Database	Classifier	Nb_F	Feature selection algorithm	Data Partition	Complexity	Validation (%)		
						FP_v	FN_v	ACC_v
Maas	**SVM**	63	MIFS+MOGA, features of 'model_6942'	AproxHull	56574	**6.18**	**26.44**	**88.20**
	MOGA				567	**7.285**	**20.69**	**88.99**
Di	**SVM**	25	MIFS+MOGA, features of 'model_3434'		19425	0	0	100
	MOGA				208	**3.17**	**9.33**	**95.26**

5.6.2. MOGA and CNN & SVM models comparison

The CNN is computationally consuming on the training phase. It`s corresponding complexity values reflect this and it is completely different from SVM. A CNN could take, on average for the problem discussed in this thesis, 30-45 min on training while an SVM takes less than 1 min.

For Maas database, the CNN provides a validation accuracy of 84.68% with an FN_v rate of 22.21% (please refer to table 5.26), while the selected model designed by MOGA 'model_6942' draws a validation accuracy of 88.99% with an FN_v rate of 20.69%. The SVM model, with the same features selected by MOGA, gives a validation accuracy of 88.20% and an FN_v rate of 26.44%.

However, for the Di database, CNN provides a validation accuracy of 94.84% with an FN_v rate of 6.48% (please refer to table 5.26), and the selected model designed by MOGA 'model_3434' provides a validation accuracy of 95.26% with an FN_v rate of 9.33%. The designed SVM model, within the features selected by 'model_3434', yields an accuracy of 100% with an FN_v rate of 0%.

Table 5. 26 : MOGA and SVM & CNN best model`s results comparison

Database	Classifier	Complexity	Validation (%)		
			FP_v	FN_v	ACC_v
Maas	CNN	43706	11.17	22.21	84.68
	MOGA (Model_6942)	567	7.285	20.69	88.99
	SVM	56574	6.18	26.44	88.20
Di	CNN	43706	4.37	6.48	94.84
	MOGA (Model_3434)	208	3.17	9.33	95.26
	SVM	19425	0	0	100

Regarding the results of the Mass database, MOGA 'model_6942' outperforms SVM and CNN, which have a very high complexity (please refer to table 5.27).

Table 5. 27 : Summary of model details and corresponding complexities: 'model_6942' from MOGA_Ms_80_1, and 'model_3434' from MOGA_Di_50_3. SVM refer to the model designed using the same features selected by the best model of MOGA ('model_6942' for Maas data and'model_3434' for Di data) with AproxHull partition

Data	Model	Complexity	Number of hidden neurons/support vectors	Number of features
Maas	Model_6942	567	9	63
	CNN	43706	6	41x41
	SVM	56574	898	63
Di	Model_3434	208	8	25
	CNN	43706	6	41x41
	SVM	19425	777	25

Otherwise, for the Dauphin Island database, MOGA 'model_3434' outperforms CNN. Nevertheless, SVM ,in this case, gives higher results on validation since the training phase was better with the cost of a larger complexity (please refer to table 5.27).

In table 5.27, complexity calculations resume for every model and for each database. The designed models with MOGA have very low complexities.

5.6.3. Literature comparison

The bibliographic study, conducted in section 2.5 of chapter one, is resumed in terms of performance of every proposed model in table 5.25.

'model_6942' outperforms the results presented by the authors, Maas and Schmalzl [10], which uses the same data. Maas and Schmalzl [10], report an average FP_v rate of 25.78% and an FN_v rate 34.93 over three different test folders since they use multiple cascades for model classification, which is more complex. The estimated complexity of the model, calculated based on details authors had provided in their manuscript, is 100800 that exceeds the complexity of SVM model.

Table 5. 28: Bibliographic study: models performance and details

Authors	Accuracy	Features type	Classifier	Additional comments
Our approach	**Maas data: overall FP 4.94, FN 15.71, accuracy 91.103**	**HOS cumulant features**	**MOGA design of an RBF Network**	**number of hidden neurons ϵ [10,20], and features vector size <80**
Besaw and Stimac [132]	overall Pd[1]=60% at FAR[2]=0.18 FA/m2 i.e. FP 18% and FN=40%	semicircles or squares features	Deep CNN	_____
Sakaguchi, Morton, Collins and Torrione [133]	AUC[3]=0.983	semicircles or squares features	CNN+SVM	with 11 convolutional filters
Hamdi and Frigui [14]	Pd=95% at FAR=10%	Edge Histogram Descriptor or Gabor features	eCHMM + neural network	very complex model
Torrione, Morton, Sakaguchi and Collins [67]	overall Pd=93% at NAN FA/m2	Histogram of oriented gradient (HOG)	random forest	large complexity of the model: ~70,004
Maas and Schmalzl [10]	FP_v=25.78% and FN_v=34.93% (average over the three testing sets used)	Haar-like features	Viola Jones algorithm	complex model

[1] Pd: Probability of detection or true positive rate
[2] FAR: False alarm rate or false positive rate
[3] AUC: Area under curve of ROC

Moysey, Knight and Jol [13]	overall accuracy >93%	covariance, Fourier-Mellin, Radon transforms, Principle components analysis	neural network	25 neurons in hidden layer
Hui-Lin, Wei-Ping and Yu-Hao [12]	min classification error 4.78%	central moments	DistAI constructive neural network	low complexity of the model: 56
Gamba and Lossani [11]	average overall accuracy 90%	images with 20x20 dimension	two-layer feedforward network with 400 input nodes and one output	- estimated complexity: 800 - presented tests are done only on two radargrams

Unfortunately, it is a bit difficult to compare the gathered results using MOGA with the rest of the models presented in table 5.25 because the data used is different and the criterions used to access the performance are different as well. Nonetheless, according to the details provided by the authors, these approaches propose complex models, expect for that of Hui-Lin, Wei-Ping and Yu-Hao [12]. These authors proposed a less complex model, using a constructive neural network, and did not provide details about results. They pretended to have a min classification error of 4.78%. Nevertheless, the evolutionary algorithm is more efficient since they did not get stuck on the local minima and maxima. However, Moysey, Knight and Jol [13] report an overall accuracy >93% with a neural network of 25 neurons in a hidden layer. Our approach attends an overall accuracy>93% of a model with a less number of neurons.

5.7. Conclusions

In this chapter, a new approach was proposed to design low complex RBF models for classifying GPR data. The approach englobes three stages: feature extraction, feature selection, and classifier design and optimization. The used features are HOS cumulants, while feature selection is performed in two stages: the first optionally stage is conducted with a mutual information technique, and a second mandatory stage is conducted with genetic algorithms. The tests were performed within two databases: Maas and Dauphin Island. The designed models obtain higher, or similar results than other classification techniques, and present a performance similar to the best results found in the literature, albeit using models with much smaller complexity.

The performance of the designed models, for hyperbola detection or pre-screening on radargrams, is investigated in chapter 6.

6. Automatic system for GPR targets discrimination and analysis

6.1. Introduction

The idea exposed through this thesis is the design of an automatic anomalies/targets software detector that could be made for application in ongoing surveys operations in the field. From a technical point of view, software detection of anomalies in radargrams is more difficult compared with other prospection techniques, like electrical tomography, because:

- Various types of media surrounding the objects cause multiple disturbances.
- Presence of multiple occurrences of incomplete or noisy hyperbolic reflections, which are due to the conditions of acquisition and inhomogeneities of the surrounding media.
- Interference of neighboring hyperbolic reflections.

The search, through the original large radargram images, is very consuming in terms of time and computational resources and sensitive to noise as well as neighboring hyperbolic interference segments. The idea is to narrow down the position of the hyperbolas to certain limited windows, and then apply the fitting technique to extract hyperbola parameters. The analysis of the original radargrams is done in several steps:

1. Pre-processing and noise reduction of radargrams.
2. Identification of small two-dimensional sections from dense radargrams, using a sliding search window, and storage of the extracted patches.
3. Targets discrimination using the proposed classification approach.
4. Fitting hyperbolas over the small selected regions, and extraction of target dimensions and depth.

The outline of this chapter is as follows: In Section 6.2, the automatic system is presented, and the technique of classifiers usage is explained. The choice of some parameters for an adequate usage of the discriminator classifiers are justified. Results obtained by the proposed approach are presented in Section 6.3, as well as a comparison of the results with a 1D directly applied to the radargrams, without a prescreening phase. The results are compared with the other approaches found in literature achievements in Section 6.4.

6.2. The automatic system

The classifiers designed in chapter 5, for each database, must be tested over the large radargrams. In order to perform this task, a sliding window approach is adopted. This approach uses a window of a size defined by the user depending of the targets types (depending on the size of targets signature could be handling a variable area on the images), which will be used by the classifier to detect if a target is present within this window, or not. Not every possible window within the radargram is used; instead jumping steps will be used, which are the number of pixels in which the sliding window is shifted in the x and y directions This technique is time consuming and generates overlapping detections with the wrong choice of scale of the search window and jumping steps. In this section, a detailed study is conducted to choose an adequate scale of the search window and jumping steps. Before performing a window search over the radargrams, a clutter removal step is executed using SVD decomposition (previously presented in chapter 4). The windowed regions are pre-processed before a feature extraction for the Maas dataset is conducted. In chapter 3, the best models selected earlier are used to discriminate either the windowed regions containing a hyperbola or not. The algorithm of the developed script/program is summarized in table 5.1.

Table 6. 1 : Automatic system

Step	Process
Hyperbolas marking	The clearest hyperbolas are marked in green
Clutter removal	A clutter removal operation is executed with a SVD decomposition
Sliding window	A sliding windows approach is implemented, with a corresponding window size and jumping steps, selected adequately after multiple trials. It is preferable to choose adequate values while trying to avoid multiple detection in the same area.
Data storage	The patches are selected by the sliding window procedure and stored in a matrix.
Data normalization	After the sliding window process is terminated, the stored data is normalized by a min-max formula.
Data pre-processing	The patches are pre-processed by the same schema previously introduced in chapter 4, while ignoring the clutter reduction stage.
Feature selection	HOS features extraction is performed for every stored patch, by the same technique detailed previously in chapter 4.
Data normalization	The features are normalized by a z-norm technique
Data classification	The stored patches are classified as target/non-target.
Results displaying	The patches classified as target are depicted in red in radargrams

6.2.1. Choice of search window dimension and jumper parameters

Let's use the radargram from figure 6.1.a) as a first example from the Maas dataset (the same treated in chapter 3). The developed script consists of a sliding window approach to localize the hyperbolas followed by target discrimination using the previously chosen classifiers and application of 1D Hough transform over the narrowed regions. Before running this script over the 155 radargrams, a previous step of marking hyperbolas is necessary. As non-specialists persons, it is difficult to mark all the hyperbolas in the images. Hence, the marked hyperbolas in this point are only the clearest ones.

The first image contains five marked targets (the clearest ones) depicted in green in figure 6.1.a). In this case, the resolution of search windows (w_s) is fixed to 30x30 pixels, while the jumping step in x-direction is set to 10 pixels and the one in y-direction is set to 10 pixels as well. The patches in red are the results of the detection program. The newly discovered hyperbolas (or new targets) are highlighted in blue. If only one red patch overlaps or intersects (with at least 30%) with any green region, then the patch should be classified as a true positive, and that green region should be considered as been identified. The case where the green patch overlaps with more than a single red patch, is the case of 'multiple detections of the same target'. From figure 6.1a), it is visually demonstrated that the program had detected two of the five marked hyperbolas, while it had discovered five other new targets. However, figure 6.1b) presents the same radargram with a different configuration that uses a larger window with the same jumping step values.

For this example (figure 6.1), counts of true detection, false detections, and new non-marked detected hyperbolas were done. Here a confusing situation is presented, two cases are exposed. In the first case the new targets are counted as true detections, while in the second case they are classified as false targets.

The count results are given in table 6.2. From the results, for the first case where the new targets are considered as false detections, it is clear that there is a large number of false detections which is wrong. In fact, these new targets, are patches that the user may have doubts to label, and this way they should not be classified as false detections.

It is remarked that with 40x40 windows, the number of false detection is less, as well as the number of multiple detections of the same target (the fact of having multiple overlapping windows over the same regions).

Another example, from the same dataset, it is considered in figure 6.2. In this case, a larger window size with sixty pixels was required to capture wider hyperbolas (figure 6.2.a)).

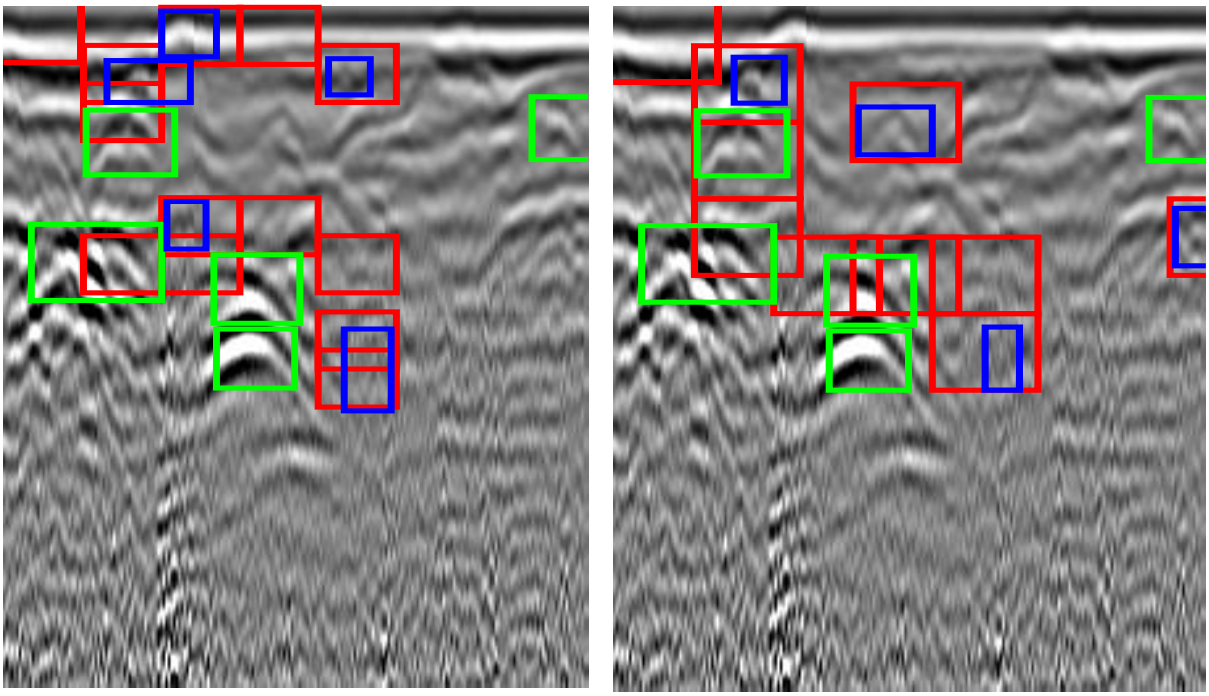


Figure 6. 1: The results of target discrimination developed program over a radargram from Maas dataset, the used configuration for every picture is: a) w_s=30, i_s=30, and j_s=20, b) w_s=40, i_s=30, and j_s=20

Table 6. 2 : Statistics of detected hyperbolas of figure 6.1

Case/Considering the new targets	Window's dimension	Overall Hyperbolas in the image	Number of		
			Overall true detections	False detections	New detections
Yes	**30**	**10**	**8**	**4**	**5**
	40	**9**	**7**	**2**	**4**
No	30	5	3	10	----
	40	5	3	6	----

Within a small search window, 40 pixels (figure 6.2.b)), the algorithm was unable to detect wider hyperbola structures. For figure 6.2., three from seven marked hyperbolas were found by the program, while, for figure 6.2.b), there is only one recorded detection. Nonetheless, with larger sizes of the search window, two major issues should be considered:

- Larger adequate jumping steps are required to avoid redundant detections.
- A selected window could contain more than one hyperbola.

Depending on the set of data, that could contain multiple types of hyperbolas, sometimes a wider search window is required. In these circumstances, (number of hyperbolas inside a window is superior to two), a second search stage should be carried out inside the selected windows to isolate multiple detections.

The choice of the search window scale is dependent on dimensions of the hyperbolas. Nevertheless, with larger scales, larger jumping steps are necessary to avoid multiple detections.

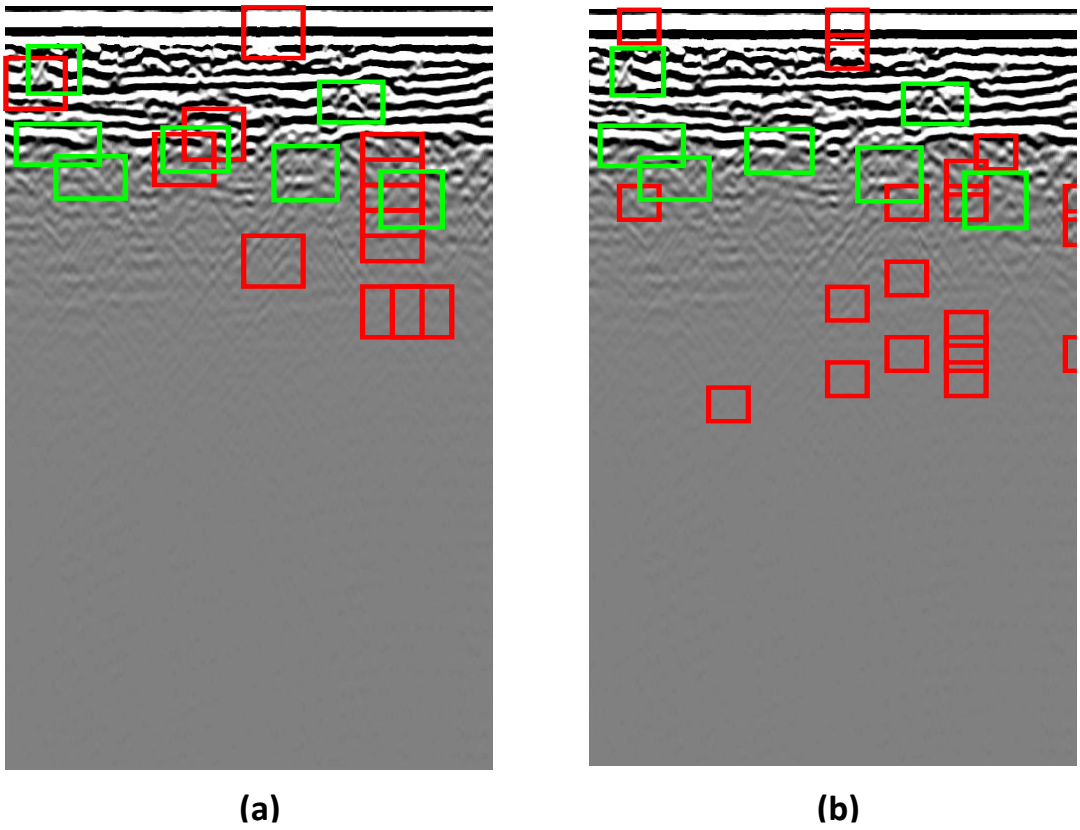
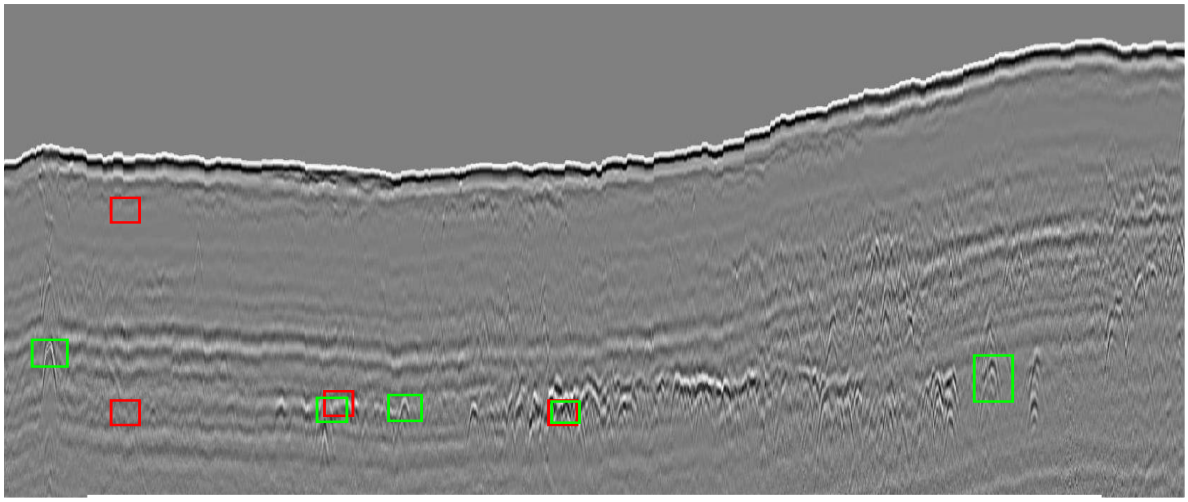


Figure 6. 2 : The results of target discrimination developed program over another radargram from Maas dataset, the used configuration is: a)w_s=60, i_s=30, and j_s=30, b) w_s=40, i_s=30, and j_s=30.

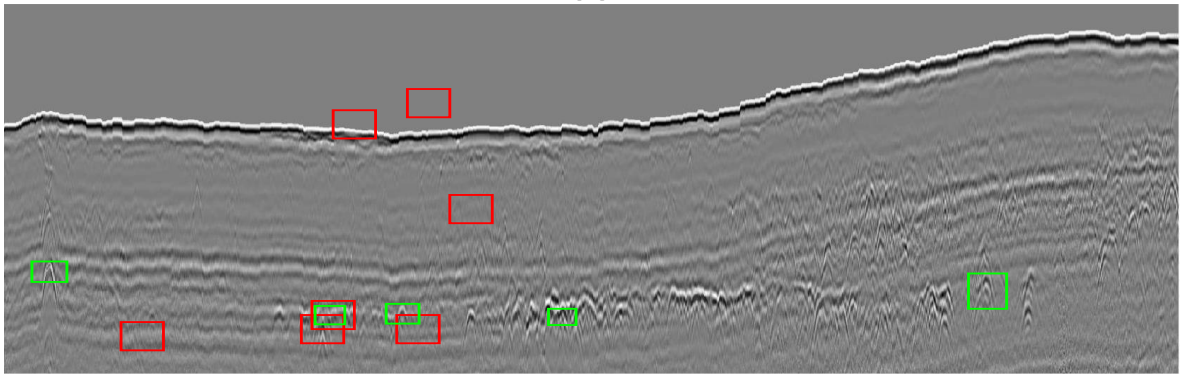
An example from the Dauphin Island dataset is given in figure 6.3 in order to assess the performance of the second designed classifier. In the case, using a search window scale ws=40 (please refer to figure 5.3.a)), two from five-marked hyperbolas were detected, while one false detection was recorded. However, the program detected one unmarked hyperbola (new target) as well. Otherwise, by using a search window scale w_s=60 (please refer to figure 6.3.b)), multiple detections of the same targets as well as more false alarms are recorded.

6.2.2. Choice of jumping steps

Within a bigger search window, it is more logical to augment the jumping steps to avoid multiple detections. However, with very dense radargrams, augmenting the jumping steps means missing some areas from the scan. Let's take the same example from figure 6.2 again.



(a)



(b)

Figure 6. 3: The results of target discrimination developed program over a radargram from Dauphin Island dataset, the used configuration is: a) w_s=40, i_s=15, and j_s=15, b) w_s=60, i_s=15, and j_s=15.

In the previous paragraph, it had been stated that a resolution of the search window of 60x60 pixels is the best. Let's adopt this resolution with larger jumping steps than the one already tested (please refer to figure 6.4).

In figure 6.4.a), the used jumping steps in x-direction and y-direction are both equal to forty pixels. It is visually clear, in this figure, that four of the seven marked hyperbolas were detected by the program. While with smaller steps (figure 6.2.b)), only two of the seven marked hyperbolas are detected. A third configuration was tested, using the same jumping steps as

figure 6.4.a), but with a larger resolution for the search window (w_s=80). The detection statistics for this case, and for the two others discussed cases, are displayed in table 6.3. For the last case, no false detection was recorded, and three new hyperbolas were detected. These numbers reflect the fact that using a wider window and a larger step, for this example, yields better detection rates. Therefore, it is a kind of dilemma.

There are no mathematical formulas to choose these parameters. Only several trials could give a more clear idea about the adequate configuration to set up with. Nonetheless, it is necessary to choose adequate parameters so that the program could localize all the hyperbolas, while avoiding multiple detections. The new detections (or non-clear detected hyperbolas) are not highlighted in the figures to avoid congestion.

The presented results here, and in the previous section, represent a few trials done within these three examples. Otherwise, more trails were conducted to identify the best parameters for every example, taking in consideration to have at least all the marked hyperbolas detected.

Another point that it was not mentioned before is the rank of the SVD clutter filtering schema. It should be chosen adequately because it affects the results of the detecting program.

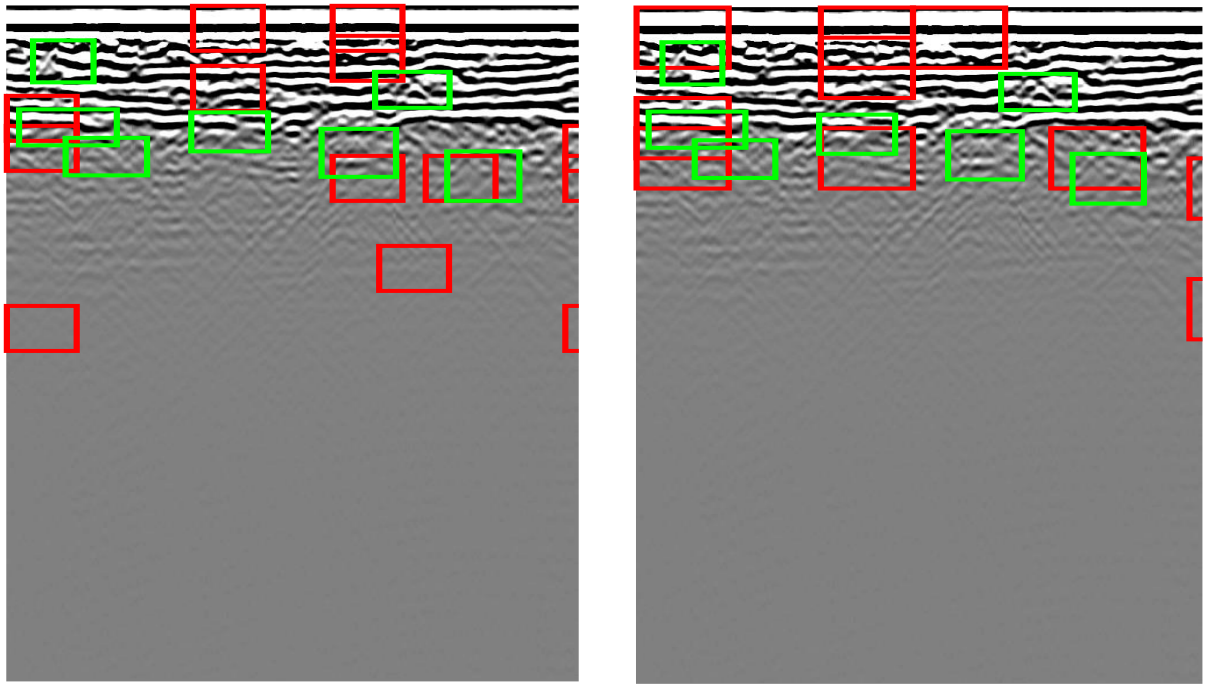


Figure 6. 4: The results of target discrimination deve loped program over a radargram from Maas dataset (the same example of figure 6.2), the used configuration is: a) ws=60, is=40, and js=40, b) ws=80, is=40, and js=40

Table 6. 3 : statistics of detected hyperbolas of figure 6.2.b) and figure 6.4

Window's resolution	Jumping steps	Marked Hyperbolas	Number of		
			Overall true detections	False detections	New detections
60	i_s=30, j_s=30	7	9	4	6
60	i_s=40, j_s=40	7	7	3	3
80	i_s=40, j_s=40	7	8	0	3

6.3. Hough Application

The modified Hough schema, previously explained in chapter 4, is applied over every selected region of the radargrams. Let's take again the example from figure 6.1 (previously undertook in chapter 4, figure 4.10), with a search window of 40x40 pixels, steps of 30, and 20 pixels in x and y directions respectively (slightly modified to avoid the multiple detections previously recorded in the middle of the picture).

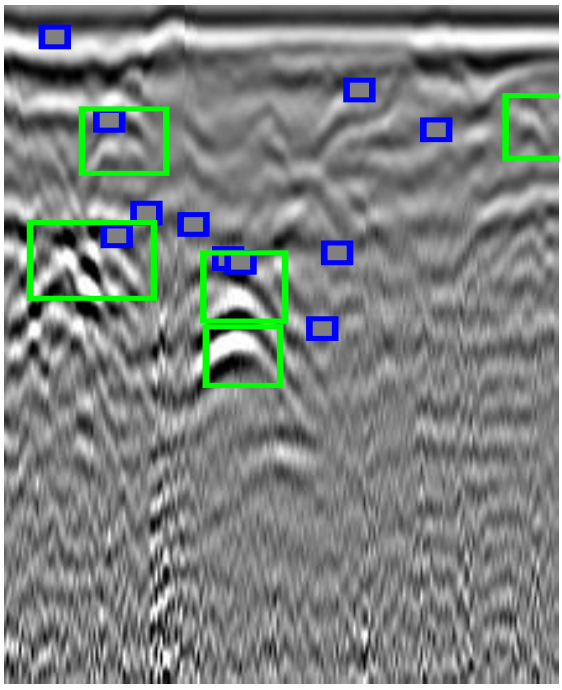


Figure 6. 5: The results after 1D Hough application over the selected patches over a radargram from Maas dataset (the same example of figure 6.1), the used configuration is w_s=50, i_s=12, and j_s=15. The green rectangles represent the previously marked hyperbolas.

Figure 6.5 demonstrates the positions of the apexes from the detected hyperbolas. The program localized three of the five marked hyperbolas. However, the picture contains fourteen overall hyperbolas that are visually clearer when the radargrams are pre-processed. The new

targets (9 hyperbolas) was detected using our program, in several experiments, within different configurations. Once one of these hyperbolas was detected in three consecutive experiments, it was marked as new target. Finally, we concluded that this picture, actually, contains fourteen overall hyperbolas.

According to the statistics presented in table 6.4, the program draws a detection rate of 60% (3 of the 5 marked ones), a misclassification rate [4]of 40% (2 missed hyperbolas of the 5 clearer ones) for this radargram, and a false positive rate of 20% (1 false detection). From a global point of view, the detection rate is equal to 64.28%, the misclassification rate to 35.71% (5 missed hyperbolas of the 14 overall existing hyperbolas in the picture), and the false positive rate is equal to 7.14% (1 false detection). The new detected hyperbolas should not be classified as non-target.

Table 6. 4 : Detection results collected for the two tested algorithms for the same radargram analysed on figure 3.11 of chapter 3 (the same as figure 5.1), the redundant detections are counted as a single detection. The non-marked hyperbolas are true detection but they ae not very clear in the original pictures.

Algorithm	**Machine Configuration**	**Time of execution (sec)**	**Overall hyperbolas**				
				Marked hyperbolas	**False detections**	**Mean (multiple detections)**	**True detections**
1D Hough schema	Intel Core i5-4200U CPU @ 1.60 GHz 2.30 GHz/RAM 6 GHz	1.67	14	5	19	3	4
Classification+ 1D Hough schema		0.127			1	1	9

Nevertheless, the 1D Hough algorithm applied directly over the radargram gives a number of false detection of 19, in comparison with the currently tested technique that gives only one false detection.

However, for the second example of figure 6.6, from the Maas dataset, the program draws a detection rate of 71.43% (5 of the 7 marked ones), a misclassification rate of 28.57% (2 missed hyperbolas of the 7 clearer ones) for this radargram, and a false positive rate of 42.85% (3 false detection). From a global point of view, the detection rate is equal to 80%, the misclassification rate to 20% (2 missed hyperbolas of the 10 overall existing hyperbolas in the picture), and the false positive rate is equal to 30% (3 false detections).

[4] Misclassification rate: (FP+FN)/total, equivalent to (1-Accuracy). It is also known as "error rate".

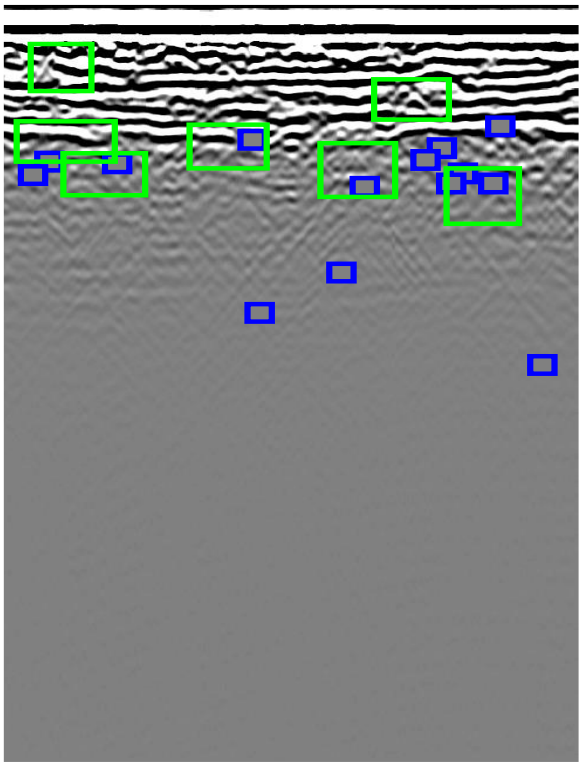


Figure 6. 6: The results after a 1D Hough application over the selected patches over a radargram from Maas dataset (the same example of figure 6.2), the used configuration is w_s=*80*, i_s=35, and j_s=*15*. The green rectangles represents the previously marked hyperbolas.

6.4. Results Comparison

The test here were done on two phases:

1) Tests were first performed on a test set of small patches.

2) Then the tests are performed over big radargrams, where the performances were calculated manually.

The comparison of tests done on the first phase was previously presented in chapter 5.

Moreover, through the previously presented example, it has been demonstrated that the procedure of discrimination, before application of the 1D Hough transform, yields better performance.

However, there are only two approaches [9, 10] that show the results in the final stage after Hough application.

Maas and Schmalzl [9], had realized tests on three different sets of radargrams. They had reported:

- A detection rate of 65.44% and a false positive rate of 5.95% over the first set.

- A detection rate of 67.74% and a false positive rate of 6.69% over the second set.
- A detection rate of 75.17% and a false positive rate of 6.55% over the third set.

However, Gamba and Lossani [10], had reported the performances drawn only on two examples. They reported:

- A detection rate of 75% and a misclassification rate of 14% over the first radargram.
- A detection rate of 89% and a misclassification rate of 19% over the second radargram.

For the approach proposed in this thesis, we report an average value, over all the examples presented here, as well as other examples non presented here: A detection rate of 80%, a false positive rate of 15%, and a misclassification rate of 20%.

It is impossible to perform a fair comparison of the results because the simulation conditions are different and the authors do not give many details that could allow us to compare the results. Nevertheless, the presented statistics demonstrate that the gathered results are in the same range, or slightly higher for some examples, than the existing approaches.

6.5. Conclusions

The developed program is capable of detecting some invisible hyperbolas that are not marked since they are not very clear in the radargrams.

The two previously designed classifiers perform well and achieve promising results. The tests of these classifiers were first performed over a validation and test sets of small patches, and then over the large radargrams for the purpose of discrimination before applying the 1D Hough transform.

The proposed approach, i.e. the designed classifiers combined with 1D Hough transform, obtains results similar or better than the already existing ones. The majority of the problems pointed in the previous chapters, commonly known with the traditional Hough algorithm like peak splitting and random detections, are fixed with this approach.

7. Final comments and future work

7.1. Conclusions

Ground penetrating radar is an active research technology, related with a vital application, which allows the identification of undersoil structures and different object characterizations. Executing a data inversion process is a hard and a remarkably compelling task. The use of GPR signatures, enables to determine the location and material information about the buried targeted objects. The existence of a computational intelligent application that is capable of assisting geologists in the analysis of radargram images is, therefore, important. Moreover, it would greatly improve the quality of an inversion process and instant discrimination of targets that could be performed in the field.

In this thesis, several data inversion approaches were proposed as alternatives to the Hough transform which is typically non-applicable with the higher resolution radargrams.

In chapter 2, a review study was conducted to assess the performance of other inversion schemas compared with the Hough fitting technique. It was stated that electromagnetic field-based inversions of GPR signals are not exploited in real-time data analysis systems because they are computationally suitable for small scatters, while they canot be applied for large-scale estimates. These facts do not allow real-time applications. However, a fitting technique fused with classification approaches achieves better results. Nevertheless, the classifiers that already exist are characterized by a higher complexity architecture, which does not allow their use in real-time applications.

Chapter 3 presents a review of the signal processing used algorithms, as well as the basic concepts of data-driven modelling techniques which were employed for developing the proposed support system.

In chapter 4, a 1D Hough schema with a correction stage to accurately identify the nature of targets from their electromagnetic properties is presented. The velocity values represent the mean value between several mediums. Consequently, the value of the electrical permittivity will be slightly modified. A user-based approach was proposed, which relies on a read value of propagation time from a time-frequency representation of the corresponding A-scan. It is preferable to automate such an approach in order to avoid the incertitude committed by the user. Under such an approach, penalization could also occur.

In chapter 5, an RBF neural network-based discrimination system for automatic identification of GPR targets through an analysis of radargram images in accordance with the previously discussed schema (in chapter 4) was proposed.

A MOGA design framework is employed to find the best possible RBF neural network structure and its corresponding parameters. Several experiments were conducted in MOGA, using first the complete set set of 198 features. Due to the complexity of the problem using this number of features, a previous selection stage, using a mutual information approach, is then performed. Tests were performed over two different databases. One (Maas database) containing raw data that was not previously corrected or pre-treated , and the other containing very clean images, not widely cluttered - Dauphin Island database. For the Maas database, the best results are obtained by a model from an ensemble of preferable models from experiment 'MOGA_Ms _80_1', where the number of features supplied to MOGA are only 80 features selected from the initial set by MIFS. Accuracy values over validation set of 88.99% and FN_v rate of 20.69% were achieved by model_6942' which employs 63 features and 9 neurons on the hidden layer.

For Dauphin Island data, an accuracy over validation set of 95.26% and FN_v rate of 9.33% were achieved by model_3434' from experiment 'MOGA_Di_50_3'. The number of features supplied to MOGA, in this case, was 50 and 'model_3434' achieved these performances with only 25 features and 8 neurons in the hidden layer. .

In general, when comparing the MOGA classification results with SVM and CNN, we were able to conclude that, despite the huge complexity of the SVM and CNN models, the accuracy of the selected models, as well as of the average of the ensemble of preferable models, is superior.

For the Dauphin Island database, the SVM model achieves slightly better results but with a huge complexity. We believe that with more experiments we could obtain better results than the ones achieved by 'model_3434'.

The proposed technique was also compared with other similar (although with not the same specifications and not using the same data) published approaches, achieving results in the same level or slightly better with a huge model complexity reduction. With the same data (Maas and Schmalzl [10]), using the MOGA approach, obtains higher results than the Viola Jones cascades.

The tests done over the radargrams were presented in chapter 6. In general, the classifier gives good results within an adequate window size and jumping steps. Nevertheless, some

targets are sometimes missed because, with the use of non-processed data, it is difficult to choose denoising parameters (SVD rank).

7.2. Future Work

The procedures may always be improved by adding new capabilities and trying to automate some tasks in order to have a complete GPR data inversion tool. Below, we have identified three future research directions that help to improve the quality of the already available developed procedures.

7.2.1. Automatization of the procedure of correction of velocity and permittivity reconstruction

An automatic procedure for tracking and detection of informative peaks and discarding noisy ones will be preferable, because with a manual value, reading the incertitude error will be high. No specific ideas about the used algorithm appear in the literature so it could simply be an image processing technique or a complete signal processing algorithm. Further research should be conducted on this point.

7.2.2. Ameliorating results of the classifier trained on Dauphin Island database by an active learning approach

This active learning procedure, which was previously tested in MOGA in a biomedical application [212], had drawn satisfactory results. On the contrary of the used passive learning procedure, active learning consists of choosing the most informative samples from the set available for the training to include in the training set. The learner then actively changes the data samples used in training, incorporating badly classified patterns, as the learning proceeds. With this mechanism, the complexity of the problem can be reduced, and the time used by the MOGA algorithm, which is necessary in order to solve the problem of optimization, is decreased.

It was mentioned before that the results for this database could be improved since SVM had achieved slightly higher results, which was an unusual outcome. Adapting this learning mechanism in an iterative use of MOGA could ameliorate the results.

7.2.3. Training a universal classifier that could be used to classify wider types of data

The quality of detection, resulting from such an application, depends strongly on the quality of the available data for training. Moreover, every GPR system is characterized by different types of signatures that vary also depending on the field of application. The developed classifiers are used only within the same data acquired with the same system for each case (Maas and Di datasets). However, it possible to create a general hyperbola-detection with good detection results. In order to achieve this goal, a larger amount of data with different systems currently available in the market is a necessary requirement.

Reference

[1] A.S. Forde, C.G. Smith, B.J. Reynolds, Archive of ground penetrating radar data collected during USGS field activity 13BIM01—Dauphin Island, Alabama, April 2013, in: Data Series, Reston, VA, 2016.
[2] M. Loke, Electrical imaging surveys for environmental and engineering studies, A practical guide to, 2 (1999).
[3] D.H. Ballard, Generalizing the Hough transform to detect arbitrary shapes, Pattern Recognition, 13 (1981) 111-122.
[4] K.S. Desale, R. Ade, Genetic algorithm based feature selection approach for effective intrusion detection system, in: Computer Communication and Informatics (ICCCI), 2015 International Conference on, IEEE, 2015, pp. 1-6.
[5] C.M. Fonseca, P.J. Fleming, Multiobjective genetic algorithms made easy: selection sharing and mating restriction, in: First International Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications, 1995, pp. 45-52.
[6] R. Battiti, Using mutual information for selecting features in supervised neural net learning, IEEE Transactions on Neural Networks, 5 (1994) 537-550.
[7] H.R. Khosravani, A.E. Ruano, P.M. Ferreira, A simple algorithm for convex hull determination in high dimensions, in: Intelligent Signal Processing (WISP), 2013 IEEE 8th International Symposium on, IEEE, 2013, pp. 109-114.
[8] H.R. Khosravani, A.E. Ruano, P.M. Ferreira, A convex hull-based data selection method for data driven models, Applied Soft Computing, 47 (2016) 515-533.
[9] P. Viola, M.J. Jones, Robust Real-Time Face Detection, International Journal of Computer Vision, 57 (2004) 137-154.
[10] C. Maas, J. Schmalzl, Using pattern recognition to automatically localize reflection hyperbolas in data from ground penetrating radar, Computers & Geosciences, 58 (2013) 116-125.
[11] P. Gamba, S. Lossani, Neural detection of pipe signatures in ground penetrating radar images, IEEE Transactions on Geoscience and Remote Sensing, 38 (2000) 790-797.
[12] Z. Hui-Lin, W. Wei-Ping, W. Yu-Hao, Constructive neural network for landmine classification using ultra wideband GPR, in: 2008 International Conference on Machine Learning and Cybernetics, 2008, pp. 1197-1201.
[13] S. Moysey, R.J. Knight, H.M. Jol, Texture-based classification of ground-penetrating radar images, Geophysics, 71 (2006) K111-K118.
[14] A. Hamdi, H. Frigui, Ensemble hidden Markov models with application to landmine detection, Eurasip Journal on Advances in Signal Processing, (2015) 15.
[15] A.v.d. Merwe, I.J. Gupta, A novel signal processing technique for clutter reduction in GPR measurements of small, shallow land mines, IEEE Transactions on Geoscience and Remote Sensing, 38 (2000) 2627-2637.
[16] N. Milisavljevic, Analyse et fusion par la theorie des fonctions de croyances de donnees multi-sensorielles pour la detection de mines anti-personnelles, in, 2001, pp. 302 p.
[17] M. Bertero, Introduction to inverse problems in imaging, Institute of Physics Publishing, Bristol, England, 1998.
[18] C. Weng Cho, Preliminary Background, in: Waves and Fields in Inhomogenous Media, Wiley-IEEE Press, 1995, pp. 1-44.
[19] R. Persico, J. Sala, Single and Double Sequence of Investigation Domains in 2-D Linear Inversions Applied to GPR Data, IEEE Geoscience and Remote Sensing Letters, 11 (2014) 1215-1219.
[20] F. Soldovieri, G. Prisco, R. Persico, A strategy for the determination of the dielectric permittivity of a lossy soil exploiting GPR surface measurements and a cooperative target, Journal of Applied Geophysics, 67 (2009) 288-295.

[21] F. Soldovieri, A. Brancaccio, G. Prisco, G. Leone, R. Pierri, A Kirchhoff-Based Shape Reconstruction Algorithm for the Multimonostatic Configuration: The Realistic Case of Buried Pipes, IEEE Transactions on Geoscience and Remote Sensing, 46 (2008) 3031-3038.
[22] R. Persico, F. Soldovieri, Two dimensional inverse scattering from buried magnetic anomalies, 2008, 51 (2008).
[23] T. Tran-Duc, N. Linh-Trung, M.N. Do, Modified Distorted Born Iterative Method for Ultrasound Tomography by Random Sampling, in: 2012 International Symposium on Communications and Information Technologies (ISCIT), 2012, pp. 1065-1068.
[24] R.E. Kleinman, P.M.v.d. Berg, A modified gradient method for two- dimensional problems in tomography, Journal of Computational and Applied Mathematics, 42 (1992) 17-35.
[25] P. Rocca, G. Oliveri, A. Massa, Differential Evolution as Applied to Electromagnetics, IEEE Antennas and Propagation Magazine, 53 (2011) 38-49.
[26] J. Gazdag, Wave equation migration with the phase-shift method, GEOPHYSICS, 43 (1978) 1342-1351.
[27] C. Ozdemir, S. Demirci, E. Yigit, A. Kavak, A hyperbolic summation method to focus B-scan ground penetrating radar images: An experimental study with a stepped frequency system, Microwave and Optical Technology Letters, 49 (2007) 671-676.
[28] M.L. Moran, R.J. Greenfield, S.A. Arcone, A.J. Delaney, Multidimensional GPR array processing using Kirchhoff migration, Journal of Applied Geophysics, 43 (2000) 281-295.
[29] S. Jiayu, L. Qing Huo, P. Torrione, L. Collins, Two-dimensional and three-dimensional NUFFT migration method for landmine detection using ground-penetrating Radar, IEEE Transactions on Geoscience and Remote Sensing, 44 (2006) 1462-1469.
[30] S. Sharma, P. Jena, R. Kuloor, Mathematical analysis of interpolation step of Omega - K Algorithm for GPR and its implementation, in: 2011 International Conference on Communications and Signal Processing, 2011, pp. 46-50.
[31] I. Catapano, A. Randazzo, E. Slob, R. Solimene, GPR Imaging Via Qualitative and Quantitative Approaches, in: A. Benedetto, L. Pajewski (Eds.) Civil Engineering Applications of Ground Penetrating Radar, Springer International Publishing, Cham, 2015, pp. 239-280.
[32] P. Hough, Method and Means for Recognizing Complex Patterns, in, United States Patent Office, 1962.
[33] M.R.-R. Aixa, F.-M. Elsa Leticia, Image-radargram analysis based on generalized Hough transform: experimental cases, Journal of Geophysics and Engineering, 9 (2012) 558.
[34] L. Capineri, P. Grande, J.A.G. Temple, Advanced image-processing technique for real-time interpretation of ground-penetrating radar images, International Journal of Imaging Systems and Technology, 9 (1998) 51-59.
[35] W. Li, X. Cui, L. Guo, J. Chen, X. Chen, X. Cao, Tree Root Automatic Recognition in Ground Penetrating Radar Profiles Based on Randomized Hough Transform, Remote Sensing, 8 (2016) 430.
[36] L. Xu, E. Oja, P. Kultanen, A new curve detection method: Randomized Hough transform (RHT), Pattern Recognition Letters, 11 (1990) 331-338.
[37] L.M.v. Kempen, H. Sahli, J. Brooks, J.P.H. Cornelis, New results on clutter reduction and parameter estimation for land mine detection using GPR, in: 8th International Conference on Ground Penetrating Radar}=, SPIE, 2000, pp. {}.
[38] J. Wang, Y. Su, Fast Detection of Gpr Objects with Cross Correlation and Hough Transform, in, 2013.
[39] H. Harkat, S. Dosse Bennani, Ground Penetrating Radar Imaging for Buried Cavities in a Dispersive Medium: Profile Reconstruction Using a Modified Hough Transform Approach and a Time-Frequency Analysis, 2015, 5 (2015) 78-92.
[40] S.J. Ahn, W. Rauh, H.-J. Warnecke, Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola, Pattern Recognition, 34 (2001) 2283-2303.
[41] Q. Dou, L. Wei, D.R. Magee, A.G. Cohn, Real-Time Hyperbola Recognition and Fitting in GPR Data, IEEE Transactions on Geoscience and Remote Sensing, 55 (2017) 51-62.

[42] A.W. Fitzgibbon, M. Pilu, R.B. Fisher, Direct least squares fitting of ellipses, in: Proceedings of 13th International Conference on Pattern Recognition, 1996, pp. 253-257 vol.251.
[43] S. Shihab, W. Al-Nuaimy, A. Eriksen, Radius estimation for subsurface cylindrical objects detected by ground penetrating radar, in: Proceedings of the Tenth International Conference on Grounds Penetrating Radar, 2004. GPR 2004., 2004, pp. 319-322.
[44] H. Chen, A.G. Cohn, Probabilistic robust hyperbola mixture model for interpreting ground penetrating radar data, in: The 2010 International Joint Conference on Neural Networks (IJCNN), 2010, pp. 1-8.
[45] W. Gander, G.H. Golub, R. Strebel, Least-squares fitting of circles and ellipses, BIT Numerical Mathematics, 34 (1994) 558-578.
[46] Z.l. Huang, J. Zhang, Determination of Parameters of Subsurface Layers Using GPR Spectral Inversion Method, IEEE Transactions on Geoscience and Remote Sensing, 52 (2014) 7527-7533.
[47] R.G. Stockwell, L. Mansinha, R.P. Lowe, Localization of the complex spectrum: the S transform, IEEE Transactions on Signal Processing, 44 (1996) 998-1001.
[48] B. Boashash, Time frequency signal analysis: Past, present and future trends, in: C.T. Leondes (Ed.) Control and Dynamic Systems, Academic Press, 1996, pp. 1-69.
[49] L. Zhou, S. Ouyang, G. Liao, L. Jin, Width estimation of a subsurface water-filled crack using internal multiple reflections from electromagnetic scattering model of a cylinder, Journal of Electromagnetic Waves and Applications, 31 (2017) 1034-1048.
[50] R.O. Schmidt, A signal subspace approach to multiple emitter location and spectral estimation, in, 1981.
[51] R. Roy, A. Paulraj, T. Kailath, ESPRIT--A subspace rotation approach to estimation of parameters of cisoids in noise, IEEE Transactions on Acoustics, Speech, and Signal Processing, 34 (1986) 1340-1342.
[52] A. Barabell, Improving the resolution performance of eigenstructure-based direction-finding algorithms, in: ICASSP '83. IEEE International Conference on Acoustics, Speech, and Signal Processing, 1983, pp. 336-339.
[53] C.L. Bastard, V. Baltazart, Y. Wang, J. Saillard, Thin-Pavement Thickness Estimation Using GPR With High-Resolution and Superresolution Methods, IEEE Transactions on Geoscience and Remote Sensing, 45 (2007) 2511-2519.
[54] S.M. Shrestha, I. Arai, Signal Processing of Ground Penetrating Radar Using Spectral Estimation Techniques to Estimate the Position of Buried Targets, EURASIP Journal on Advances in Signal Processing, 2003 (2003) 970543.
[55] K. Dinh, T. Zayed, S. Moufti, A. Shami, A. Jabri, M. Abouhamad, T. Dawood, Clustering-Based Threshold Model for Condition Assessment of Concrete Bridge Decks with Ground-Penetrating Radar, Transportation Research Record: Journal of the Transportation Research Board, 2522 (2015) 81-89.
[56] H. Frigui, P. Gader, K. Satyanarayana, Landmine detection with ground penetrating radar using fuzzy k-nearest neighbors, in: 2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542), 2004, pp. 1745-1749 vol.1743.
[57] A. Khalifa, H. Frigui, Fusion of multiple algorithms for detecting buried objects using fuzzy inference, in: SPIE Defense + Security, SPIE, 2014, pp. 90720V.
[58] R. Janning, T. Horváth, A. Busche, L. Schmidt-Thieme, GamRec: A Clustering Method Using Geometrical Background Knowledge for GPR Data Pre-processing, in: L. Iliadis, I. Maglogiannis, H. Papadopoulos (Eds.) Artificial Intelligence Applications and Innovations: 8th IFIP WG 12.5 International Conference, AIAI 2012, Halkidiki, Greece, September 27-30, 2012, Proceedings, Part I, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 347-356.
[59] R. Janning, A. Busche, T. Horváth, L. Schmidt-Thieme, Buried pipe localization using an iterative geometric clustering on GPR data, Artificial Intelligence Review, 42 (2014) 403-425.
[60] B. Jafrasteh, N. Fathianpour, Automatic extraction of geometrical characteristics hidden in ground-penetrating radar sectional images using simultaneous perturbation artificial bee colony algorithm, Geophysical Prospecting, 65 (2017) 324-336.

[61] R. Ahmadi, N. Fathianpour, Estimating geometrical parameters of cylindrical targets detected by ground-penetrating radar using template matching algorithm, Arabian Journal of Geosciences, 10 (2017) 140.
[62] S. Florence, T. Jean-Philippe, Template-matching based detection of hyperbolas in ground-penetrating radargrams for buried utilities, Journal of Geophysics and Engineering, 13 (2016) 491.
[63] H. Frigui, O. Missaoui, P. Gader, Landmine detection using discrete hidden Markov models with Gabor features, in, 2007, pp. 65532A-65532A-65510.
[64] S. Gibb, H.M. La, Automated Rebar Detection for Ground-Penetrating Radar, in: G. Bebis, R. Boyle, B. Parvin, D. Koracin, F. Porikli, S. Skaff, A. Entezari, J. Min, D. Iwai, A. Sadagic, C. Scheidegger, T. Isenberg (Eds.) Advances in Visual Computing: 12th International Symposium, ISVC 2016, Las Vegas, NV, USA, December 12-14, 2016, Proceedings, Part I, Springer International Publishing, Cham, 2016, pp. 815-824.
[65] Y. Qin, C. Huang, Identifying underground voids using a GPR circular-end bow-tie antenna system based on a support vector machine, International Journal of Remote Sensing, 37 (2016) 876-888.
[66] N.R. Syambas, A system for identification of a buried object on GPR using a decision tree method, in: 2011 6th International Conference on Telecommunication Systems, Services, and Applications (TSSA), 2011, pp. 169-175.
[67] P.A. Torrione, K.D. Morton, R. Sakaguchi, L.M. Collins, Histograms of Oriented Gradients for Landmine Detection in Ground-Penetrating Radar Data, IEEE Transactions on Geoscience and Remote Sensing, 52 (2014) 1539-1550.
[68] J. Baili, S. Lahouar, M. Hergli, I.L. Al-Qadi, K. Besbes, GPR signal de-noising by discrete wavelet transform, NDT & E International, 42 (2009) 696-703.
[69] A. Tzanis, Signal enhancement and geometric information retrieval from 2-D GPR data with multiscale, orientation-sensitive filtering methods, First Break 32, 8 (2014) 91–98.
[70] A. Tzanis, The Curvelet Transform in the analysis of 2-D GPR data: Signal enhancement and extraction of orientation-and-scale-dependent information, Journal of Applied Geophysics, 115 (2015) 145-170.
[71] X. Wang, S. Liu, Noise suppressing and direct wave arrivals removal in GPR data based on Shearlet transform, Signal Processing, 132 (2017) 227-242.
[72] B. Oskooi, M. Julayusefi, A. Goudarzi, GPR noise reduction based on wavelet thresholdings, Arabian Journal of Geosciences, 8 (2015) 2937-2951.
[73] B.M. Battista, A.D. Addison, C.C. Knapp, Empirical Mode Decomposition Operator for Dewowing GPR Data, Journal of Environmental and Engineering Geophysics, 14 (2009) 163-169.
[74] C.-S. Chen, Y. Jeng, A data-driven multidimensional signal-noise decomposition approach for GPR data processing, Computers & Geosciences, 85 (2015) 164-174.
[75] J. Li, C. Liu, Z. Zeng, L. Chen, GPR Signal Denoising and Target Extraction With the CEEMD Method, IEEE Geoscience and Remote Sensing Letters, 12 (2015) 1615-1619.
[76] S. Chen, D. Donoho, M. Saunders, Atomic Decomposition by Basis Pursuit, SIAM Review, 43 (2001) 129-159.
[77] F. Abujarad, A. Omar, GPR Data Processing Using the Component-Separation Methods PCA and ICA, in: Proceedings of the 2006 IEEE International Workshop on Imagining Systems and Techniques (IST 2006), 2006, pp. 60-64.
[78] B. Karlsen, J. Larsen, H.B.D. Sorensen, K.B. Jakobsen, Comparison of PCA and ICA based clutter reduction in GPR systems for anti-personal landmine detection, in: Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing (Cat. No.01TH8563), 2001, pp. 146-149.
[79] E. Temlioglu, I. Erer, Clutter Removal in Ground-Penetrating Radar Images Using Morphological Component Analysis, IEEE Geoscience and Remote Sensing Letters, 13 (2016) 1802-1806.
[80] F. Abujarad, Independent factor analysis for clutter reduction in GPR data for landmine detection, in: Proceedings of the 15th International Conference on Ground Penetrating Radar, 2014, pp. 989-992.

[81] F. Abujarad, A. Jostingmeier, A.S. Omar, Clutter removal for landmine using different signal processing techniques, in: Proceedings of the Tenth International Conference on Grounds Penetrating Radar, 2004. GPR 2004., 2004, pp. 697-700.
[82] L. Yuan, F. Guang-You, GPR clutter reduction and buried target detection by improved Kalman filter technique, in: 2005 International Conference on Machine Learning and Cybernetics, 2005, pp. 5432-5436 Vol. 5439.
[83] T.J. Li, L.J. Kong, Z.O. Zhou, Symmetry filtering method for GPR clutter reduction, in: International Conference on Microwave and Millimeter Wave Technology Proceedings, ICMMT, 2008, pp. 1515-1517.
[84] D.J. Daniels, Ground Penetrating Radar, Institution of Engineering and Technology, 2004.
[85] D. Potin, E. Duflos, P. Vanheeghe, Landmines Ground-Penetrating Radar Signal Enhancement by Digital Filtering, IEEE Transactions on Geoscience and Remote Sensing, 44 (2006) 2393-2406.
[86] L.A. Cirillo, C.B. Brown, A.M. Zoubir, Polynomial Phase Signal based Detection of Buried Landmines using Ground Penetrating Radar, in: Proc. 11th IEEE Workshop on Statistical Signal Processing(SSP), 2001, pp. 166-169.
[87] L.v. Kempen, H. Sahli, Signal processing techniques for clutter parameters estimation and clutter removal in GPR data for landmine detection, in: Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing (Cat. No.01TH8563), 2001, pp. 158-161.
[88] D. Deiana, L. Anitori, Detection and classification of landmines using AR modeling of GPR data, in: Proceedings of the XIII Internarional Conference on Ground Penetrating Radar, 2010, pp. 1-5.
[89] D.P. Nabelek, K.C. Ho, Detection of shallow buried objects using an autoregressive model on the ground penetrating radar signal, in: J.T. Broach, J.C. Isaacs (Eds.) Detection and Sensing of Mines, Explosive Objects, and Obscured Targets Xviii, 2013.
[90] L. Rabiner, B.-H. Juang, Fundamentals of speech recognition, Prentice-Hall, Inc., 1993.
[91] K.D. Morton, P.A. Torrione, L. Collins, Nonparametric Bayesian Time-Series Modeling and Clustering of Time-Domain Ground Penetrating Radar Landmine Responses, in: R.S. Harmon, K.H. Holloway, J.T. Broach (Eds.) Detection and Sensing of Mines, Explosive Objects, and Obscured Targets Xv, 2010.
[92] T.S. Ferguson, A Bayesian Analysis of Some Nonparametric Problems, Ann. Statist., 1 (1973) 209-230.
[93] J. Sethuraman, A CONSTRUCTIVE DEFINITION OF DIRICHLET PRIORS, Statistica Sinica, 4 (1994) 639-650.
[94] H. Ishwaran, L.F. James, Gibbs Sampling Methods for Stick-Breaking Priors, Journal of the American Statistical Association, 96 (2001) 161-173.
[95] J. Paisley, L. Carin, Hidden Markov Models With Stick-Breaking Priors, IEEE Transactions on Signal Processing, 57 (2009) 3905-3917.
[96] P. Torrione, K.D. Morton, R. Sakaguchi, L.M. Collins, Histogram of gradient features for buried threat detection in ground penetrating radar data, in: 2012 IEEE International Geoscience and Remote Sensing Symposium, 2012, pp. 3182-3185.
[97] D. Reichman, L.M. Collins, J.M. Malof, Improvements to the Histogram of Oriented Gradient (HOG) prescreener for buried threat detection in ground penetrating radar data, in: SPIE Defense + Security, SPIE, 2017, pp. 101820U.
[98] D. Reichman, J.M. Malof, L.M. Collins, Algorithm Development for Deeply Buried Threat Detection in GPR Data, in: S.S. Bishop, J.C. Isaacs (Eds.) Detection and Sensing of Mines, Explosive Objects, and Obscured Targets Xxi, 2016.
[99] H. Frigui, P. Gader, Detection and discrimination of land mines in ground-penetrating radar based on edge histogram descriptors, in: J.T. Broach, R.S. Harmon, J.H. Holloway (Eds.) Detection and Remediation Technologies for Mines and Minelike Targets Xi, Pts 1 and 2, 2006.
[100] H. Frigui, P. Gader, Detection and Discrimination of Land Mines in Ground-Penetrating Radar Based on Edge Histogram Descriptors and a Possibilistic Nearest Neighbor Classifier, IEEE Transactions on Fuzzy Systems, 17 (2009) 185-199.

[101] H. Frigui, A. Fadeev, A. Karem, P. Gader, Adaptive edge histogram descriptor for landmine detection using GPR, in, 2009.
[102] S. Harris, B. Alvey, D.K.C. Ho, A. Zare, LBP features for hand-held ground penetrating radar, in: Proc. SPIE 10182, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXII, 101820Z, 2017.
[103] R. Sakaguchi, Using Image Processing Methods to Improve the Detection of Buried Explosive Threats in GPR Data, in, 2016.
[104] K. Pearson, LIII. On lines and planes of closest fit to systems of points in space, Philosophical Magazine, 2 (1901) 559-572.
[105] S. Park, K. Kim, K.H. Ko, Multi-Feature Based Multiple Landmine Detection Using Ground Penetration Radar, Společnost pro radioelektronické inženýrství, 23 (2014) 642-651.
[106] J.B. Rodriguez, M.F. Pantoja, X.L. Travassos, D.A.G. Vieira, R.R. Saldanha, A prediction algorithm for data analysis in GPR-based surveys, Neurocomputing, 168 (2015) 464-474.
[107] A. Catakli, H. Mahdi, H.A. Shukri, Texture analysis of GPR data as a tool for depicting soil mineralogy, in, 2011, pp. 1-8.
[108] W. Zhao, E. Forte, M. Pipan, Texture Attribute Analysis of GPR Data for Archaeological Prospection, Pure and Applied Geophysics, 173 (2016) 2737-2751.
[109] R. Haralick, K. Shanmugam, I. Dinstein, Texture Features for Image Classification, IEEE Transactions on Systems, Man, and Cybernetics, 3 (1973).
[110] P. Torrione, L.M. Collins, Texture features for antitank landmine detection using ground penetrating radar, IEEE Transactions on geoscience and remote sensing, 45 (2007) 2374-2382.
[111] P. Torrione, L. Collins, Application of texture feature classification methods to landmine/clutter discrimination in off-lane GPR data, in: Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International, IEEE, 2004, pp. 1621-1624.
[112] M.-H. Horng, Y.-N. Sun, X.-Z. Lin, Texture feature coding method for classification of liver sonography, Computerized medical imaging and graphics, 26 (2002) 33-42.
[113] C. Gilmore, S. Peters, J. LoVetri, D. McNeill, GPR target detection using a neural network classifier of image moments as invariant features, in: Antenna Technology and Applied Electromagnetics and URSI Conference, 2004. ANTEM 2004. 10th International Symposium on, IEEE, 2004, pp. 1-4.
[114] P. Klęsk, M. Kapruziak, B. Olech, Statistical moments calculated via integral images in application to landmine detection from Ground Penetrating Radar 3D scans, Pattern Analysis and Applications, (2017).
[115] M.-K. Hu, Visual pattern recognition by moment invariants, IRE transactions on information theory, 8 (1962) 179-187.
[116] W. Zhou, G. Wang, X. Chen, M. Chen, M. Tian, Classification of ground penetrating radar echo signals using wavelet packet and RBF, in: Radar, 2006 IEEE Conference on, IEEE, 2006, pp. 5 pp.
[117] M.S. El-Mahallawy, M. Hashim, Material classification of underground utilities from GPR images using DCT-based SVM approach, IEEE Geoscience and Remote Sensing Letters, 10 (2013) 1542-1546.
[118] Q. Lu, J. Pu, Z. Liu, Feature extraction and automatic material classification of underground objects from ground penetrating radar data, Journal of Electrical and Computer Engineering, 2014 (2014) 28.
[119] W. Shao, A. Bouzerdoum, S. Phung, L. Su, B. Indraratna, C. Rujikiatkamjorn, Automatic classification of GPR signals, in: Ground Penetrating Radar (GPR), 2010 13th International Conference on, IEEE, 2010, pp. 1-6.
[120] W. Al-Nuaimy, Y. Huang, M. Nakhkash, M.T.C. Fang, V.T. Nguyen, A. Eriksen, Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition, Journal of Applied Geophysics, 43 (2000) 157-165.
[121] K. Ho, L. Carin, P.D. Gader, J.N. Wilson, An investigation of using the spectral characteristics from ground penetrating radar for landmine/clutter discrimination, IEEE transactions on Geoscience and Remote Sensing, 46 (2008) 1177-1191.

[122] A.N. Balan, M.R. Azimi-Sadjadi, Detection and classification of buried dielectric anomalies by means of the bispectrum method and neural networks, IEEE Transactions on Instrumentation and Measurement, 44 (1995) 998-1002.
[123] A.D. Strange, J.C. Ralston, V. Chandran, Near-surface Interface Detection for Coal Mining Applications using Bispectral Features and GPR, Subsurface Sensing Technologies and Applications, 6 (2005) 125-149.
[124] U.S. Khan, W. Al-Nuaimy, F.E. Abd El-Samie, Detection of landmines and underground utilities from acoustic and GPR images with a cepstral approach, Journal of Visual Communication and Image Representation, 21 (2010) 731-740.
[125] K. Tokuda, T. Kobayashi, T. Masuko, S. Imai, Mel-generalized cepstral analysis - a unified approach to speech spectral estimation, in: ICSLP, 1994.
[126] S. Furui, Cepstral analysis technique for automatic speaker verification, IEEE Transactions on Acoustics, Speech, and Signal Processing, 29 (1981) 254-272.
[127] T.G. Savelyev, L. van Kempen, H. Sahli, GPR anti-personnel mine detection: improved deconvolution and time-frequency feature extraction, in: A.L. Gyekenyesi, P.J. Shull (Eds.) Nondestructive Evaluation and Health Monitoring of Aerospace Materials and Composites Ii, Spie-Int Soc Optical Engineering, Bellingham, 2003, pp. 232-241.
[128] W. Shao, A. Bouzerdoum, S.L. Phung, Signal classification for ground penetrating Radar using sparse kernel feature selection, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7 (2014) 4670-4680.
[129] N.R. Mudigonda, R. Kacelenga, M. Edwards, Holographic neural networks versus conventional neural networks: a comparative evaluation for the classification of landmine targets in ground-penetrating radar images, in, 2004, pp. 996-1007.
[130] K. Etemad, R. Chellappa, Separability-based multiscale basis selection and feature extraction for signal and image classification, IEEE Transactions on Image Processing, 7 (1998) 1453-1465.
[131] L.E. Besaw, P.J. Stimac, Deep Learning Algorithms for Detecting Explosive Hazards in Ground Penetrating Radar Data, in: S.S. Bishop, J.C. Isaacs (Eds.) Detection and Sensing of Mines, Explosive Objects, and Obscured Targets Xix, Spie-Int Soc Optical Engineering, Bellingham, 2014.
[132] L.E. Besaw, P.J. Stimac, Deep Convolutional Neural Networks for Classifying GPR B-Scans, in: R.V.T.U.S.A. Appl Res Associates (Ed.) Conference on Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XX, Spie-Int Soc Optical Engineering, Baltimore, MD, 2015.
[133] R.T. Sakaguchi, K.D. Morton, L.M. Collins, P.A. Torrione, Recognizing subsurface target responses in ground penetrating radar data using convolutional neural networks, in, 2015, pp. 94541A-94541A-94510.
[134] P.D. Gader, M. Mystkowski, Z. Yunxin, Landmine detection with ground penetrating radar using hidden Markov models, IEEE Transactions on Geoscience and Remote Sensing, 39 (2001) 1231-1244.
[135] H. Frigui, K.C. Ho, P. Gader, Real-Time Landmine Detection with Ground-Penetrating Radar Using Discriminative and Adaptive Hidden Markov Models, EURASIP Journal on Advances in Signal Processing, 2005 (2005) 419248.
[136] A. Filippidis, L.C. Jain, P. Lozo, Degree of familiarity ART2 in knowledge-based landmine detection, IEEE transactions on neural networks, 10 (1999) 186-193.
[137] L. O. Odhiambo, R. S. Freeland, R. E. Yoder, Soil Characterization Using Textural Features Extracted from GPR Data, in, ASAE, St. Joseph, MI, 2004.
[138] M.R. Shaw, S.G. Millard, T.C.K. Molyneaux, M.J. Taylor, J.H. Bungey, Location of steel reinforcement in concrete using ground penetrating radar and neural networks, NDT & E International, 38 (2005) 203-212.
[139] D.W. Patterson, Artificial Neural Networks: Theory and Applications, Prentice Hall, 1996.
[140] P. Van Vuuren, Landmine detection by means of ground penetrating radar: a model-based approach, SOUTH AFRICAN INSTITUTE OF ELECTRICAL ENGINEERS 105 (2014) 90-103.
[141] F. Girosi, T. Poggio, Networks and the best approximation property, Biological cybernetics, 63 (1990) 169-176.

[142] S. Lameri, F. Lombardi, P. Bestagini, M. Lualdi, S. Tubaro, Landmine Detection from GPR Data Using Convolutional Neural Networks, in: 25th European Signal Processing Conference (EUSIPCO), Kos island, Greece, 2017.
[143] D. Reichman, L.M. Collins, J.M. Malof, Some good practices for applying convolutional neural networks to buried threat detection in Ground Penetrating Radar, in: Advanced Ground Penetrating Radar (IWAGPR), 2017 9th International Workshop on, IEEE, 2017, pp. 1-5.
[144] L.E. Besaw, Detecting buried explosive hazards with handheld GPR and deep learning, in: SPIE Defense+ Security, International Society for Optics and Photonics, 2016, pp. 98230N-98230N-98211.
[145] A. Krizhevsky, G. Hinton, Learning multiple layers of features from tiny images, (2009).
[146] F.A.A. Queiroz, D.A.G. Vieira, X.L. Travassos, Analyzing the Relevant Features of GPR Scattered Waves in Time- and Frequency-Domain, Research in Nondestructive Evaluation, 24 (2013) 105-123.
[147] B.P. Rohman, M. Nishimoto, Multi-scaled power spectrum based features for landmine detection using Ground Penetrating Radar, in: Signals and Systems (ICSigSys), 2017 International Conference on, IEEE, 2017, pp. 83-86.
[148] F. Queiroz, D.A. Vieira, X. Travassos, M. Pantoja, Feature extraction and selection in ground penetrating radar with experimental data set of inclusions in concrete blocks, in: Machine Learning and Applications (ICMLA), 2012 11th International Conference on, IEEE, 2012, pp. 48-53.
[149] I. Kononenko, Estimating attributes: analysis and extensions of RELIEF, in: European conference on machine learning, Springer, 1994, pp. 171-182.
[150] R. Gilad-Bachrach, A. Navot, N. Tishby, Margin based feature selection-theory and algorithms, in: Proceedings of the twenty-first international conference on Machine learning, ACM, 2004, pp. 43.
[151] H. Harkat, A. Ruano, M. Graça, S. Dosse Bennani, GPR hyperbolas classification using a multi-objective genetic approach, in: 9th European Symposium on Computational Intelligence and Mathematics, Faro, Portugal, 2017.
[152] N. Boubaki, Cavity detection using two geophysical methods : Ground-Penetrating Radar and Electrical Resistivity Tomography, in, Université Paris Sud - Paris XI, 2013.
[153] J.C. Maxwell, A Treatise on Electricity and Magnetism, Cambridge University Press, Cambridge, 2010.
[154] M. Nabighian, Electromagnetic Methods in Applied Geophysics, 1988.
[155] A.P. Annan, GPR Methods for Hydrogeological Studies, in: Y. Rubin, S.S. Hubbard (Eds.) Hydrogeophysics, Springer Netherlands, Dordrecht, 2005, pp. 185-213.
[156] A.P. Annan, Chapter 1 - Electromagnetic Principles of Ground Penetrating Radar A2 - Jol, Harry M, in: Ground Penetrating Radar Theory and Applications, Elsevier, Amsterdam, 2009, pp. 1-40.
[157] A. Taflove, S.C. Hagness, Computational electrodynamics: the finite-difference time-domain method, 3rd ed., Artech House, Norwood, 2005.
[158] J.-P. Berenger, A perfectly matched layer for the absorption of electromagnetic waves, Journal of Computational Physics, 114 (1994) 185-200.
[159] R. Rejiba, Fayçal, C. Camerlynck, Modélisation 3D de la propagation des ondes radar en milieu hétérogène, atténuant et dispersif par la méthode des différences finies dans le domaine temporel, in, 2002, pp. 81-98.
[160] F. Rejiba, Modélisation de la propagation des ondes électromagnétiques en milieux hétérogènes - Application au Radar Sol, in, University Paris VI, 2002.
[161] R.M. Joseph, S.C. Hagness, A. Taflove, Direct time integration of Maxwell's equations in linear dispersive media with absorption for scattering and propagation of femtosecond electromagnetic pulses, Opt. Lett., 16 (1991) 1412-1414.
[162] P.M. Merlin, D.J. Farber, A Parallel Mechanism for Detecting Curves in Pictures, IEEE Trans. Comput., 24 (1975) 96-98.
[163] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, IEEE Transactions on Information Theory, 36 (1990) 961-1005.
[164] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series

analysis, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 454 (1998) 903-995.
[165] N.E. Huang, M.-L. Wu, W. Qu, S.R. Long, S.S.P. Shen, Applications of Hilbert–Huang transform to non-stationary financial time series analysis, Applied Stochastic Models in Business and Industry, 19 (2003) 245-268.
[166] B. Boashash, Chapter 3 - Theory and Design of High-Resolution Quadratic TFDs0 A2 - Boashash, Boualem, in: Time-Frequency Signal Analysis and Processing (Second Edition), Academic Press, Oxford, 2016, pp. 103-137.
[167] J. Krug, S. Zaunseder, M. Rabenau, R. Poll, H. Sager, Comparative Analysis of different Wigner-Ville Distribution Implementations for the ECG-based Detection of Obstructive Sleep Apnea, in: World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany, Springer, 2009, pp. 906-909.
[168] F. Hlawatsch, T.G. Manickam, R.L. Urbanke, W. Jones, Smoothed pseudo-Wigner distribution, Choi-Williams distribution, and cone-kernel representation: Ambiguity-domain analysis and experimental comparison, Signal Processing, 43 (1995) 149-168.
[169] A. Prochazka, N. Kingsbury, P.J.W. Payner, J. Uhlir, Signal Analysis and Prediction, 2013.
[170] P. Gonçalves, R.G. Baraniuk, Pseudo affine Wigner distributions: Definition and kernel formulation, IEEE Transactions on Signal Processing, 46 (1998) 1505-1516.
[171] P. Flandrin, O. Rioul, Affine smoothing of the Wigner-Ville distribution, in: Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on, IEEE, 1990, pp. 2455-2458.
[172] J.M. Mendel, Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications, Proceedings of the IEEE, 79 (1991) 278-305.
[173] C.L. Nikias, J.M. Mendel, Signal processing with higher-order spectra, IEEE Signal Processing Magazine, 10 (1993) 10-37.
[174] W. Li, Using genetic algorithm for network intrusion detection, Proceedings of the United States Department of Energy Cyber Security Group, 1 (2004) 1-8.
[175] C.E. Shannon, A mathematical theory of communication, The Bell System Technical Journal, 27 (1948) 379-423.
[176] D.S. Broomhead, D. Lowe, Radial basis functions, multi-variable functional interpolation and adaptive networks, in, Royal Signals and Radar Establishment Malvern (United Kingdom), 1988.
[177] S. Haykin, Neural Networks: A Comprehensive Foundation , McMillan Publ, Co., New York, (1998).
[178] Y. Wang, M. Chang, H. Chen, M.Q. Wang, Application of RBF Neural Network in Intelligent Fault Diagnosis System, in: Proceedings of International Conference on Soft Computing Techniques and Engineering Application, Springer, 2014, pp. 561-566.
[179] K.B. Kim, C.K. Kim, Performance Improvement of RBF Network Using ART2 Algorithm and Fuzzy Logic System, in, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 853-860.
[180] P.M. Ferreira, A.E. Ruano, Exploiting the separability of linear and nonlinear parameters in radial basis function networks, in: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373), 2000, pp. 321-326.
[181] S. Chen, C.F. Cowan, P.M. Grant, Orthogonal least squares learning algorithm for radial basis function networks, IEEE Transactions on neural networks, 2 (1991) 302-309.
[182] C. Chinrungrueng, C.H. Sequin, Optimal adaptive k-means algorithm with dynamic adjustment of learning rate, IEEE Transactions on neural networks, 6 (1995) 157-169.
[183] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, 1999.
[184] K. Levenberg, A method for the solution of certain non-linear problems in least squares, Quarterly of Applied Mathematics, 2 (1944) 164-168.
[185] D.W. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society for Industrial and Applied Mathematics, 11 (1963) 431-441.
[186] A.E. Ruano, P.M. Ferreira, C.M. Fonseca, An overview of nonlinear identification and control with neural networks, in: Intelligent Control Systems using Computational Intelligence Techniques, Institution of Engineering and Technology, 2005, pp. 37-88.

[187] A.E. Ruano, Intelligent control systems using computational intelligence techniques, Iet, 2005.
[188] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, (2014).
[189] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning, 20 (1995) 273-297.
[190] C.M. Fonseca, P.J. Fleming, Multiobjective optimization and multiple constraint handling with evolutionary algorithms. I. A unified formulation, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 28 (1998) 26-37.
[191] A. Ruano, H.R. Khosravani, P.M. Ferreira, A Randomized Approximation Convex Hull Algorithm for High Dimensions, IFAC-PapersOnLine, 48 (2015) 123-128.
[192] E. Hajimani, INTELLIGENT SUPPORT SYSTEM FOR CVA DIAGNOSIS BY CEREBRAL COMPUTERIZED TOMOGRAPHY, in: Faculdade de Cîencias e Tecnologia, Algarve University, 2016, pp. 249.
[193] H. Harkat, S.B. Dosse, A. Slimani, Time-Frequency Analysis of GPR Signal for Cavities Detection Application, in: Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015, Springer, 2016, pp. 107-117.
[194] H. Harkat, S. Dosse Bennani, A. Slimani, Time-Frequency Analysis of GPR Signal Using a Modified Hilbert Huang Approach, Revue Méditerranéenne des Télécommunications, 5 (2015).
[195] F. Perira, Analyse spatio-temporelle du champ géomagnétique et des processus d'accélération solaires observés en émission radio, in, Université d'Orléans, 2004.
[196] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (1989) 674-693.
[197] J. Zheng, S.-p. Peng, F. Yang, A novel edge detection for buried target extraction after SVD-2D wavelet processing, Journal of Applied Geophysics, 106 (2014) 106-113.
[198] A.A. Goshtasby, 2-D and 3-D image registration: for medical, remote sensing, and industrial applications, John Wiley & Sons, 2005.
[199] F. Auger, P. Flandrin, P. Gonçalvès, O. Lemoine, Time-frequency toolbox, CNRS France-Rice University, 46 (1996).
[200] Z.K. Peng, P.W. Tse, F.L. Chu, An improved Hilbert–Huang transform and its application in vibration signal analysis, Journal of Sound and Vibration, 286 (2005) 187-205.
[201] W.J. Williams, M.L. Brown, A.O. Hero, Uncertainty, information, and time-frequency distributions, in: Advanced Signal Processing Algorithms, Architectures, and Implementations II, International Society for Optics and Photonics, 1991, pp. 144-157.
[202] L. Stanković, A measure of some time–frequency distributions concentration, Signal Processing, 81 (2001) 621-631.
[203] N.K. Verma, S. Goel, R.K. Sevakula, Study of transforms and methods for their comparison, in: Industrial and Information Systems (ICIIS), 2014 9th International Conference on, IEEE, 2014, pp. 1-8.
[204] D. Boutana, A comparative performance measure of some timefrequency distributions for Arabic plosive-vowel transition using Renyi criterion, in: Signal Processing and Its Applications, 2005. Proceedings of the Eighth International Symposium on, IEEE, 2005, pp. 587-590.
[205] L.M.v. Kempen, H. Sahli, J. Brooks, J.P.H. Cornelis, New results on clutter reduction and parameter estimation for landmine detection using GPR, in: 8th International Conference on Ground Penetrating Radar}=, SPIE, 2000, pp. {}.
[206] D.L. Donoho, De-noising by soft-thresholding, IEEE Transactions on Information Theory, 41 (1995) 613-627.
[207] S.R. Deans, The Radon Transform and Some of Its Applications, New York, 1983.
[208] B. Chen, V. Chandran, Robust Image Hashing Using Higher Order Spectral Features, in: 2010 International Conference on Digital Image Computing: Techniques and Applications, 2010, pp. 100-104.
[209] A. Swami, J.M. Mendel, C.L.M. Nikias, Hi Spec Toolbox : For Use With MATLAB: User's Guide, in: M. Natick (Ed.), Math Works, 1993.

[210] T.K. Leen, G.B. Orr, Optimal stochastic search and adaptive momentum, in: Proceedings of the 6th International Conference on Neural Information Processing Systems, Morgan Kaufmann Publishers Inc., Denver, Colorado, 1993, pp. 477-484.
[211] A. Vedaldi, K. Lenc, MatConvNet: Convolutional Neural Networks for MATLAB, in: Proceedings of the 23rd ACM international conference on Multimedia, ACM, Brisbane, Australia, 2015, pp. 689-692.
[212] E. Hajimani, M.G. Ruano, A.E. Ruano, An intelligent support system for automatic detection of cerebral vascular accidents from brain CT images, Computer Methods and Programs in Biomedicine, 146 (2017) 109-123.