WORLD MODELING IN RADAR: A REGULARIZATION BY SYNTHESIS

David Ameen Maluf

B.Eng. (American University of Beirut), 1987

M.Eng. (McGill University), 1991

Department of Electrical Engineering

McGill University

Montreal, Quebec, Canada

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

May 1995

ISBN 0-612-05750-X

Abstract

In *Radio Detection and Ranging*, the *inverse problem* is that of acquiring *knowledge* of the physical features of a body by making observations of the reflected energy and synthesising the model from the measured data. This procedure is in contrast to the *forward problem*, which consists of calculating the observable effects from a given model. The forward problem has a unique solution whereas the inverse process, being carried out on the basis of *hypotheses*, is always characterized by a lack of uniqueness.

The approach taken towards developing the *synthesis* framework is to consider a logic of argumentation of *belief functions* which is used to represent the various symbolic aspects of belief and uncertainty. The logic extends so that not just one argument, but all arguments, supporting or opposing a hypothesis are considered. That is the logic used to solve the *inverse problem*. As arguments are identified among measurements, the support they confer on a hypothesis or its negation is aggregated to provide a measure of the degree of belief in the hypotheses of interest. The aggregation operation, or the *synthesis regularization*, will depend on an *entropy* calculation to represent the *uncertainty* associated with the arguments.

Based on the theory of *Kalman filters*, sensor fusion is used to finally establish probabilistic models of the hypotheses. In conjunction with the synthesis regularization, consistent estimates will converge to a qualitative image reconstruction. The synthesis framework is compared to current solutions to the inverse problem in radio detection and ranging and applied to *Ground Penetrating Radar* image reconstruction.

Résumé

Le problème inverse dans l'imagerie à partir des radars consiste en l'acquisition de la connaissance des caractéristiques physiques du corps en observant l'énergie réfléchie et en établissant le modèle à partir des mesures calculées. Cette procédure contraste avec le problème directe qui consiste en un calcul des effets observés d'un modèle donné. Le problème directe possède une solution unique tandis que le problème inverse, est toujours caractérisé par un manque de unicité.

L'approche prise en vue de développer le cadre de la synthèse, est utilisée pour représenter les différents aspects symboliques de la croyance et le doute. Cette approche doit prendre en considération une logique d'argumentation des fonctions de croyance ainsi que l'étendue de cette logique, pour que tous les arguments, et non un seul, supportant ou opposant une hypothèse, soient considérés, c'est la logique utilisée pour résoudre le problème inverse. Comme les arguments sont identifiés parmi les capteurs et les moyens de mesure, le support qu'ils confèrent à une hypothèse ou à sa négation est aggré de procurer une mesure du degré de croyance dans les hypothèses d'intérêt. L'opération aggrégative, ou le processus de la régularisation par synthèse, dépendera d'un calcul d'entropie pour représenter le doute associé aux arguments.

La fusion des capteurs, basée sur la théorie des filtres de type Kalman, est finalement utilisée pour établir des modèles de probabilité pour les estimations. En conjonction avec la régularisation par synthèse, les prédictions des capteurs convergent vers une reconstruction d'une image qualitative. Le cadre de la synthèse est comparé aux solution courante du problème inverse avec une application sur les radars.

Acknowledgements

First of all, I must thank my thesis supervisor, Professor L. Daneshmend, for his understanding and guidance throughout this work. His encouragement has kept me going through the sleepless nights. In addition, I am very grateful to the others on my thesis committee, Professors P. Belanger and R. DeMori, who have guided me at the different stages in my research.

As I have benefited enormously from all the additional research I have been involved in and which I believe has made this work possible, I am very grateful to Professor R. DeMori for his guidance on other related issues in applied artificial intelligence and also to Dr. J. Peck. I would like to express my sincere gratitude to Dr. M. Desmarais of Centre de Recherche Informatique de Montréal (CRIM) for introducing me to the research of knowledge assessment and for allocating time and resources for me to complete this work. Special thanks go to Dr. J. Liu for his many constructive suggestions.

I am indebted to McGill University and the electrical engineering department for allowing me the opportunity of higher education and I feel privileged to have carried out this research at McGill Research Center for Intelligent Machines. Thanks go also to the mining department at McGill, especially to Mrs. N. Procyshyn, Professor Hassani, Dr. M. Momayez and Mr. A. Sadri, for their technical support.

I would like also to express my thanks to my colleagues and friends, J. Kairouz, P. Knights, A. Bedair, J. Mossop and many others for their cheerful encouragement, support and discussions.

I must express my sincere gratitude to Dr. J. Tatelman for her courage in reading and editing earlier drafts of this manuscript, when I was in the process of turning this research into an acceptable thesis.

Finally I would like to acknowledge my parents, my brother Dr. N. Maluf and my sister S. Maluf for their constant support and encouragement.

Claim of Originality

The author claims the originality of the basic ideas and research results presented in this thesis, the following being the most significant:

1. Philosophy of the *regularization synthesis* in statistical inversion theory of *a priori* data to *a posteriori* knowledge delineated by belief functions and uncertainty (Chapter 4 and Chapter 5).

2. Derivation of the exact entropy for the Dempster-Shafer theory. The proposed entropy computation is based on the principle of the theory of Dempster-Shafer (Chapter 5).

3. Derivation of the Kalman filter in its decentralized form to an autoregressive AR optimal filter (Chapter 6).

4. Derivation of the *finite to semi-finite* mapping from an uncertainty concept to a normal distribution variance (Chapter 6).

5. Introduction of the *regularization factor* as a partial momentum warp which affects the inference in the minimum entropy search. This method is devised for Dempster-Shafer as well as Bayesian representation (Chapter 5).

6. Concept of adaptive noise filtering for narrow wave scans which results in the innovation of *dynamic adaption* to a time-varying noise spectrum (Chapter 3).

7. Concept of recovering spatial distortion in electromagnetic imaging and adaptation of the *range resolution* factor from the airborne long range radar system

and derivation for ground penetrating radar application (Chapter 7).

8. The hierarchal color scheme that would suit the requirements of image encoding and concur with the basic conditions for efficient image perception (Chapter 8).

Some of these contributions have been partly reported in many preliminary forms [64] [63] [58] [59] [66] [65].

To the Maluf Family ...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> I keep six honest serving men. They taught me all I knew. Their names are What and Why and When and How and Where and Who.
>
> *R. Kipling, Just So Stories.*

Some of the earliest *world modeling* with respect to *inverse problems* was of an astronomical nature. Since the time of Aristotle (384-322 B.C.) human kind has been interested in perceiving what is beyond its "vision", but it was not until Röntgen (1845-1923) discovered X-rays, which enabled him to see inside living tissue, that "vision" beyond the naked eye entered a new era. In the following years, the theory of *Radio Detection And Ranging* in imaging has developed so rapidly that astronomy, medicine and geology are just a few of the areas where remote sensing has been found useful. Finally, whether in electromagnetic, optical or acoustic sensing, the main intention of this research and analysis is to augment our understanding of the surrounding world.

1.1 World Modeling

The objective of *world modeling in radio detection and ranging* is to determine the spatial configuration and/or physical properties of an unknown object. World modeling problems can be subdivided into a number of classes, according to the kind of information that must be retrieved in the measurements. In the present thesis, besides the methodologies, two types of *world modeling* are presented. The first stage is the *identification problem*, which amounts to establishing input-output relations from the wavelets such as deconvolution and noise estimation . The second stage is the *inverse problem*, where the identified measurements are combined to determine the models' spatial configuration. This class of inverse problem stems directly from *inverse scattering properties* of electromagnetic waves.

The *inverse problem* deals with the process of searching for *unique* solutions to the problem of the reverse mapping of dispersion and scattering phenomena. In general, the problem does not admit a unique solution and has to be approached by special cases or approximations and/or by search algorithms. In this thesis, we start to explore new methodologies as an alternative solution to the inverse problem in radio detection and ranging. As we do not claim uniqueness and there is no attempt to do so, we explore these new techniques for our world modeling *synthesis framework*.

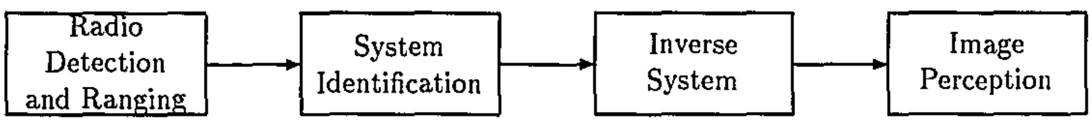


Figure 1.1: Block diagram showing the possible overall sequence. System Identification might involve any type of amplitude, time and frequency modifications and/or adjustments.

This research examines two complementary sensor fusion methods that we adapt to the inverse problem context. We initially devise the *synthesis framework* of the world modeling process based on sensor fusion methods. Then we evaluate the proposed system in comparison to the current solution for the inverse problems in radio detection and ranging with a direct application to *Ground Penetrating Radar* (GPR). The first sensor fusion method relies on *belief functions* and specifically on *Dempster-Shafer* evidential reasoning which is a generalization of Bayesian inference classification [19] [81] [32]. The second sensor fusion method is an adaptation of the recursive *Kalman filter* to its *decentralized* form [51] [39].

1.2 Motivation and Objective

The instances when the inverse problem is of importance are valid when information about the structure and composition of an object are required but cannot be ascertained from direct measurements. Rather, the measurements are recorded at a spatial location distant from the object and without affecting it in any tangible way. It is therefore apparent that the problem of inverse scattering is central to the development of techniques in remote sensing, imaging science and non-destructive analyses of materials in all their various manifestations. These appear in such diverse areas as astronomy [38], antenna synthesis [21], computed tomography in medical physics [7], profile inversion in geophysics [2] [89] and ground penetrating radar. In each of these areas of study, sensors have been developed to produce signals and images which all, in one way or another, exploit the way the radiation interacts with the property of scatterer. In the latter topic, only a few techniques have been developed to a workable state. One reason could be attributed to its recent investigation whereas most theories were tailored for its seismic counterparts.

1.3 Remote Sensing

Generally speaking, remote sensing technology describes the concept of acquiring information distantly about specific targets. We realize that ground penetrating radar introduces a paradigm shift in the understanding of remote sensing. A simple reason can be associated with the common belief of remote sensing, a characteristic known as *transparency.* In fact, the use of radar in aircraft technology can be recognized as a particular case of ground penetrating radar. Indeed, radio detection and ranging systems are not concerned with the nature of the target but merely with its presence and location. This is contrasted with the ability to penetrate the target with the intent to reconstruct in detail the morphology of the target, including its shape and composition. Although the task is difficult, ground penetrating radar is likely to be a fruitful area of research.

What mainly constitutes remote sensing is the existence of a transmitter and receiver. The basis of remote sensing can be clearly defined as the emission of a pulse of continuous wavelength (CW) associated with some modulation, such as frequency sweep (FMCW) . On the other hand, a large amount of information about the target is generated and requires sensing capabilities of the same order as the generated information. There are some remote sensing systems which rely on *passive* sources, such as in astronomy or vision, but here we will concentrate on what are known as *active* systems which in fact constitute the majority of applications of inverse theories.

In general, a property of active remote sensing systems is the fact that the source is arbitrary in all the natural dimensions – temporal, spatial, energy spectrum etc. – which introduces a control factor into the solution of the inverse problem.

1.4 Outline of Thesis

In order to fully follow world modeling in radio detection and ranging as presented in this thesis, understanding of the analytical and numerical constraints of signal processing, the inverse problem, logic reasoning and perception is essential. However, understanding of electromagnetic field computations is reduced to the equivalent basic knowledge of optics.

A chapter on the inverse scattering problem has also been included. The various known solutions to inverse problems stem from the knowledge acquired from studies of different research backgrounds that deals with scattering concepts. The work done in attempting these inverse problems aid in the understanding of the sophisticated work in world modeling issues.

The thesis itself is divided into three parts. The first part deals with the concepts and main theoretical derivation of world modeling and consists of five chapters. Section 2.1 introduces the general perspective of radio detection and ranging. The concepts of ground penetrating radar are introduced only in Section 7.1. Section 2.3 deals with the inversion methodologies for the inverse problem. Radio detection and ranging inversions are the primary concern of this thesis. The inverse problem definition was elaborated on the residual error and finite difference computations. These inversion methodologies in a sense propose the forward problem solution as an alternative approach to the inverse problem. In addition, and under the same heading as the inverse problem, the most common technique for radar inversion is also described from a theoretical as well as a practical perspective. Also in Section 2.3.3, the notion of inverse operator stacking is mentioned as the evolutionary aspect of migration techniques.

Chapter 3 introduces aspects of system identification. The theory of linear time invariant systems dominates this chapter where numerous techniques dedicated to system identification and noise filtering are discussed, in particular, the least-squares method in linear prediction and estimation. The wavelet theory and information capacity is introduced in Section 3.2. Section 3.3 discusses the methodology of approaching the system identification of electromagnetic scattering measurements. In Section 3.3, the assumption of zero offset imaging is stated as a basis of this thesis work. The scattering and measurements are performed with a pseudo-single antenna. Section 3.4 contains the equations that describe the linear input-output relations of the scattering and measurements state. A simple algorithm for estimating measurement noise corruption is presented in Section 3.4.1. In S^ction 3.5, the possibility of extending the concept of narrow scans introduced in Section 3.4.1 as the basis of an initial assumption for a sophisticated deconvolution algorithm is discussed. The least-squares method, which has been shown to be useful in many parameter estimations [60], is formulated for the general prediction radio detection and ranging output estimate in Section 3.5.1. The viability of the formulation is accomplished by deriving the equation around a simple example; the inverse filtering and echo excitation removal from radar signals.

Chapter 4 contains an introduction to the inverse problem. General difficulties one encounters in understanding the inverse problem are discussed in Section 4.2. Henceforth, the formulation of the inverse problem is proposed. However, Section 4.2 is considered relatively important to the notations and terminology which are used throughout the thesis. As this thesis introduces the definition of divergence in the inverse problem which is separated from the definition of the ill-posed inverse problem, this chapter is relatively important. Therefore in Section 4.2.1 the four categorizations of the inverse problem are proposed. Section 4.3 delineates the Backus-Gilbert approach to inverse problems. This process introduces the *regularization* term in inverse problem nomenclature. Nowadays, Backus-Gilbert is often recommended as the generic method of choice for designing and predicting the performance of experiments that require data inversion. Although one cannot obtain a complete solution based on this approach, Section 4.3 nevertheless presents a clear picture of regularization and uncertainty factors in the inverse problem. The resolution limits and resolving power which mainly validate the inverse problem solution are discussed in Section 4.3.1. The uncertainty factor resolution limits in this context are also described. The main points of this chapter are expounded in Section 4.4, where a qualitative approach to the inverse problem is presented. This new approach to the inverse problem, however, does not claim uniqueness, but incorporates techniques to stabilize divergent inverse problems (also mentioned in Section 4.2.1). Section 4.4.2 relates the *a posteriori* knowledge map and uncertainty to a knowledge mapping process. A brief discussion is presented in Section 4.4.3, hence introducing the proposed regularization synthesis.

In Chapter 5, the synthesis regularization method is proposed. An alternative descriptive title for Chapter 5 is *logic regularization*, however, the title *A World of Beliefs* was chosen due to the nature of its contents rather than its functionality. An introduction to evidential reasoning is presented in Section 5.1. The notion of representation of knowledge with uncertainty factors is presented in the same Section. This chapter contains the major theoretical contributions of this thesis. The Bayesian and belief function propagations of evidence are described in Section 5.2 where belief networks, directed cyclic/acyclic graphs and knowledge regularization synthesis form the main focus. Section 5.2.2 describes a possible notation of the Dempster-Schafer model and the frame of discernment concepts. In addition, Section 5.2.2 suggests some definitions and assumptions of the sensor data implications to the belief functions and includes the algorithm for combining belief functions. Section 5.3 describes the generic derivation of exact entropy computation for the Dempster-Shafer belief model [81]. Dempster-Shafer evidential reasoning is considered a generalization of Bayesian inference. The second contribution is in Section 5.4 and Section 5.4.1 and explains the derivation of the *regularizing factor* (rf), which is a momentum bias that largely affects the knowledge evaluation. The *regularizing factor* (rf) is usually associated with skewing the distribution of the data and results in momentum in the minimum entropy computation. This method is devised for Dempster-Shafer as well as for Bayesian knowledge representations.

Chapter 6 is devoted to the *Kalman filtering* integration method and functioning as front-end multi-sensor fusion technique. The Kalman filter is our proposed method to achieve the final synthesis and statistical mapping. In Section 6.2 we derive the finite to semi-finite mapping from an uncertainty concept to a normal distribution variance. In Section 6.3 we discuss model robustness and the need to combine the knowledge maps into an optimal estimate.

Part two of this thesis is devoted to various aspects of the experimental application of world modeling and the inverse problem to ground penetrating radar, starting with Section 7.1 which discusses ground penetrating radar from a user point of view. Besides the main functionality aspects and the hardware involved in ground penetrating radar which are well elaborated in Section 7.1.1, other factors which enter into the data interpretation are also explained. The basic electrical and electromagnetic specifications of the GSSI SIR 10 radar unit in addition to some main components such as control functions, line scanning and the board computer are described in Section 7.2; however, on-board digital filtering is presented separately in Section 7.2.1 even though it consists of part of the radar unit. Since the spectroanalysis of the transmitted signal of the radar was never measured, Section 7.2.2 discusses the hypothetical spectra. Section 7.3 becomes more conceptually technical when the reflection, refraction and scattering phenomena are explained. This Section is of importance for its proposal of a new methodology for the operation of ground penetrating radar. When both scattered and reflected signals are separated, the reflection analysis provides the data for the Impulse Reflection Knowledge Map (IRKM) which is explained in Section 7.3.1, and the scattering analysis provides the data for the Polarized Scattering Map (PSKM) which is discussed in Section 7.3.2. As spatial distortion in ground penetrating radar decreases its resolution, Section 7.3.3 reports an important characteristic of the horizontal resolution of short distance radar where the analysis of the *range resolution* (R_r) is proposed. On the other hand, to remedy the vertical spatial distortion that results from velocity change of the electromagnetic wave field in different media, Section 7.3.4 demonstrates an efficient approach to manipulating the velocity-depth-time space. Once the basis of the knowledge maps is computed as in Sections 7.3.1 and 7.3.2, the knowledge map (KM) synthesis, that is Section 7.4, systematically uses the knowledge maps (IRKM) and (PSKM) in constructing the world model.

Chapter 8 deals with data interpretation and perception as the next stage from Chapter 7 in knowledge handling. Section 8.1 introduces the general perspective of the visual perception. The parameters' description as sets of forms is presented in Section 8.2, where the hierarchical information combination is proposed. However, it is in Section 8.3 where image encoding is explained. Numerical results for the overall process (Figure 1.1) are generated and are shown. All techniques listed in this thesis are to be applied for validation purposes in their numerical performance implementation in the inversion of the measurements of ground penetrating radar with the exception of the finite difference forward modelling algorithm.

In part three and Chapter 9, general conclusions are drawn. Also, suggestions are made for further research in the inverse-scattering problem in general, and in regularization by synthesis and world modelling. The strategic review (Chapter 9) contains also a brief review of the inverse problem as well as a claim of originality stated in Section 9.3.

Part I

Synthesis: Radio Detection and Ranging

Chapter 2

General Perspective

Remote sensing is broadly defined as collecting information about a target without being in physical contact with it. Aircrafts and satellites are among many platforms where remote sensing is extensively used for remote observation. The term *remote sensing* is commonly restricted to methods that employ electromagnetic energy as the means of detecting and ranging target characteristics. In this thesis, we classify electromagnetic sensing as radio detection and ranging (RADAR), a subclass of remote sensing, in contrast to its usual classification.

Historically, the reflection of radio waves from objects was noted in the late 1800s and early 1900s. Definitive investigation of radar began in the 1920's in the United States and Great Britain for war detection purposes. In present times, radar applications are considered to be the most reliable in remote sensing.

2.1 Radio Detection And Ranging

Electromagnetic energy refers to all energy that moves with the velocity of light in a harmonic wave pattern. A harmonic pattern consists of waves that occur at equal intervals in time. The wave concept explains how electromagnetic energy propagates, but this energy can only be detected as it interacts with matter. In this interaction, electromagnetic energy behaves as though it has particle-like properties. When electromagnetic energy refracts as it propagates through different media, it is behaving as a wave (Appendix A.1).

Electromagnetic waves can be described in terms of their velocity, wavelength and frequency. All electromagnetic energy travels at the speed of light in a vacuum. Unlike velocity and wavelength, which change as electromagnetic energy is propagated through media of different characteristics (e.g. density), frequency remains constant and is therefore a fundamental reference property. Electromagnetic energy that encounters matter, whether solid, liquid or gas, is referred to as incident radiation. Interactions with matter can change the properties of the incident radiation: intensity, direction, wavelength, polarization and phase. During interactions between the electromagnetic energy and the media, mass and energy are conserved according to basic physical principles. The incident radiation can endure the following behavior:

1. **Refraction**, that is, the incident radiation passed through the medium. Transmission through media of different electrical characteristics causes a change in the velocity which is accompanied by a direction change.

2. **Reflection**, that is, returned from the medium. Conjugate responses are found such as the reflection angle and the wave polarization.

3. **Scattering**, that is, deflected in all directions.

4. **Absorption**, where the electromagnetic energy is transformed into another form of energy such as heat or any type of radiation of other wavelengths.

Hence, the incident electromagnetic energy may be reflected, refracted, scattered and absorbed, often in combination. Radio detection and ranging depends highly on the reflected waves which are called the *Albedo* [85]. The Albedo is the ratio of the energy reflected, usually onto a different medium, to the incident energy.

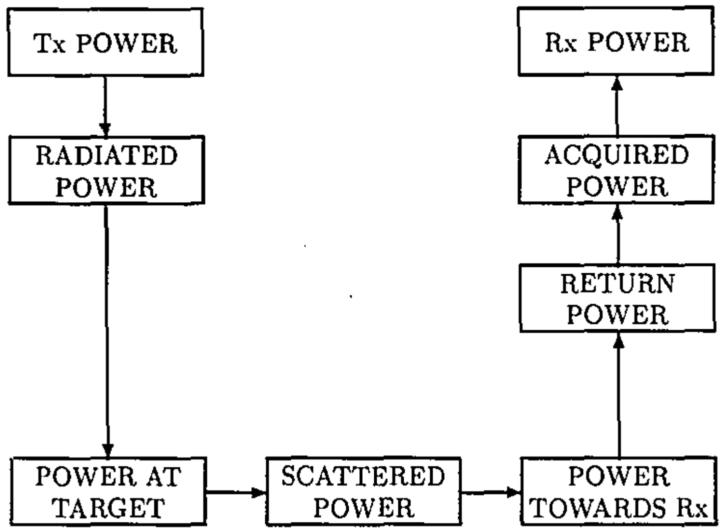


Figure 2.1: A pulse of high frequency electromagnetic energy (Tx power) is used repetitively. As this energy source waveform propagates through the environment (scattered power), some of the energy is reflected back to the origin (acquired power).

This terminology is often used in an optics context. On the other hand, scattering phenomena are continuously masked by the aspects of the reflection and refraction counterparts which are usually stronger in magnitude. Scattering results from multiple interactions of the electromagnetic energy on a particle level. The major process of scattering in radio detection and ranging is of a *non-selective scattering* type [50] where all wavelengths of the electromagnetic energy are equally scattered. Electromagnetic scattering phenomena are comparable to the illumination process at an invisible wavelength.

2.1.1 Radar Component

Having described radio detection and ranging with relation to electromagnetic concepts, in reference to Section 1.3, the radio detection and ranging system is an *active* remote sensing system as it provides its own source of energy. In fact, the system illuminates the surrounding matter with electromagnetic energy and detects *radar returns* which are the energy returning from the medium. Radio detection and ranging systems operate in the radio and microwave bands of the electromagnetic spectrum ranging from a meter to a few millimeters in wavelength. The advantage of similar active systems can be extended to the proper tuning of wavelength, hence achieving ground penetration capabilities and performing subsurface detection and ranging.

2.2 Ground Penetrating Radar

Ground penetrating radar (GPR) is a non-destructive technique similar in principal to seismic applications. An electromagnetic energy pulse propagates into the ground and the partial reflections are sampled and recorded [18]. Radar hardware design is diverse, however it shares a common technology. For purposes of the experimentation reported in this proposal, a Geophysical Survey System Incorporated (GSSI) radar unit (SIR-10) was used, with the system set to operate in the L band around a center frequency of 1 GHz. A random-letter code was assigned to different frequencies during the early stages of development to avoid mention of the wavelength regions under investigation. The transmitter and receiver antennae were located adjacent to each other, behaving as a single pseudo-antenna.

Over the past decades ground penetrating radar techniques have been applied to many practical problems as a non-destructive testing method. This has resulted in a demand for more dependable subsurface imaging, as well as improved capabilities for processing and interpreting the resulting images in an autonomous fashion. However, until recently [70], there were no devised methods that met the above requirements. Data were acquired manually and interpreted in their raw format.

Understanding of the ground penetrating radar concept is best introduced by the analogy of a flashing light beam dispersed through a semi-transparent medium. Despite the fact that well-established concepts in signal analysis and for the inverse problem have been developed for seismic applications in geophysics [78] [13] [14], such formalisms are still lacking in the ground penetrating radar domain. The focus in ground penetrating radar remains on the evolution of particular models rather than on generic methods, and, likewise, the emphasis is placed more often on empirical approaches to the field instead of on the classical foundations.

In most applications of electromagnetic imaging in subsurface exploration, a pulse of high frequency electromagnetic energy is used repetitively. As this energy source waveform propagates through the environment, some of the energy is reflected back to the origin (Figure 2.1). The reflections vary with the composition of the medium in terms of its electric properties and hence describe the *forward problem.* In Figure 2.2.a, the cross section of a cylinder has been generated by a radar simulator showing some features of the electromagnetic scattering phenomena (Figure 2.2.b). In Figure 2.2.c, an attempt to invert the process of the forward modeling is shown.

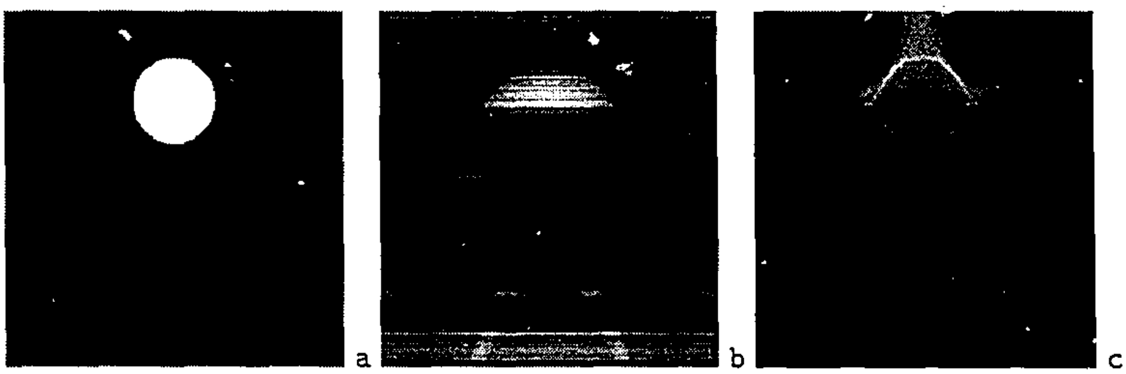


Figure 2.2: Left: A cross section of a cylinder being scanned by the ground penetrating radar beam. Middle: The output shows the scattering effect of the electromagnetic waves. Right: Image reconstruction is carried on on the basis of an inverse process.

From the imaging point of view for real world situations, all objects are three-dimensional. However, in most cases, a technique devised for solving a two-dimensional problem can be generalized for three-dimensional applications. Furthermore, in the cases of ground penetrating radar applications, two-dimensional images can be stacked together to form a three-dimensional reconstruction. It is, therefore, appropriate to consider a two-dimensional – and if possible a single-dimensional – version of a problem since, besides offering physical insight, it involves less algebraic and computational complexities.

Although the subsurface electromagnetic imaging is in its early stage and still growing in an era of advanced research, some prominent techniques like pattern recognition were applied and were pioneered for the subsurface electromagnetic imaging applications by Poulton [75] and later, by Glass [34] where the concept of artificial neural networks was introduced in the radar context. Additional models have been used to describe the behavior of ground penetrating radar in ways to facilitate the analysis of the system. Since the use of the radar acquisition system yields enormous amounts of data, it is advantageous to model the ground penetrating radar to scale as described by Smith and Scott [86]. Relationships between full-sized systems and models having scaled physical dimensions can be established. Such an approach can lead to an exponential decrease in the computational analyses that are to be performed, since geophysical characteristics and dimensionality correction ratios are preserved.

2.3 Inversion Methodologies

Inverse problems abound in science and engineering and some inversion methodologies are subsequently derived. Examples of typical inverse problems are found in: antenna synthesis [21], computed tomography in medical physics [7] and profile inversion in geophysics [2] [89]. In the latter topic, only a few techniques have been developed to a workable state. One reason could be attributed to its recent application whereas most theories were tailored for its seismic counterparts.

In formulating the solution techniques for the inverse problem as well as for the inverse-scattering problem, an important consideration has been to maintain its capability of being generalized to more complex configurations as much as possible. A consequence of this approach is that the conclusions obtained from the abstract formulation must be verified by computing numerical results for adequate image inversions. In fact, all the techniques listed in this thesis are applied for validation purposes in their numerical performance in the implementation of the inversion of the measurements of the ground penetrating radar application.

2.3.1 Forward Modeling and Residual Error

One way of solving an inverse problem is to propose a model, calculate the wave field, compare the calculated field with the measured data and subsequently modify the model (Figure 2.3.1).

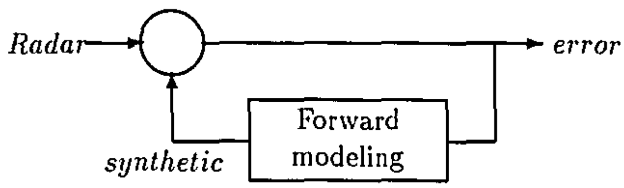


Figure 2.3: Forward modeling and residual error computation and estimation. Synthetic data are simulated according to a model which is compared to the real data. As the error is minimized, a model can be estimated.

Until Yeo [95], where a solution to the *penetrable wedge* problem was addressed, methods of computing the electromagnetic wavefield had to undergo many approximations. Initially, the research in forward modeling was dominated by mathematicians and physicists in the wave/particle properties of radiation conditions and boundary conditions. In contrast, the focus of this thesis research is to achieve a feasible, efficient and stable solution to the inverse problem without approaching forward modeling techniques.

2.3.2 Finite Difference Forward Modeling

In a *finite difference* scheme, a space-time mesh is introduced and the wave equations are replaced by a system of finite difference equations on the mesh. An appropriate numerical method is then invoked to solve for the field at every point of the mesh. The finite-difference method in the time domain (FDTD) has been adapted to model radar wavelets. The method is based on the explicit finite-difference of Maxwell's equations (Appendix A.2). The model is set up to produce the forward model wavelets when given the constraints and to result in synthetic data.

The finite-difference time domain method is a robust technique for forward modeling of electromagnetic wave radiation, propagation and scattering. As the FDTD method is computer-intensive, it becomes time consuming when used recursively in residual error computation. With the advent of the enormous increase in computer technology, the FDTD method has became one of the leading candidates for calculating scattering from radar targets. In the past three years, several articles have described the FDTD method of forward modeling in GPR [9] [87].

2.3.3 Migration Stacking Inverse Operators

There are a number of different ways of looking at this process. One which is intellectually satisfying, though not computationally useful, is to see it as a a two-dimensional deconvolution. Each point in the subsurface produces in the Euclidean data space a characteristic response of a hyperbolic nature – a two-dimensional impulse response. Since the total response in the measurement space is the linear superposition of impulse responses of all points on the subsurface, the forward modeling can be thought of as a two-dimensional convolution of hyperbolae, the impulse responses. Hence migration is the inverse of this convolution. In short, migration is the process of constructing the reflector surface from the record surface. The basic mathematical properties of migration were developed by Hagerdoorn [41].

When migration was simplified to a sequence of hyberbolae stacking, migration became ultimately the most widely accepted process whereby the antenna pattern contribution could be removed. In many respects, the process is the same as synthetic aperture processing. Migration analysis in its implemented form requires antenna pattern information as well as wave velocity for qualitative image focusing. Although the migration process does a very effective job of collapsing the hyperbolic returns back to their localized source position, the migration process is not perfect nor was it expected to be, as the problem is mainly attributed to boundary conditions.

Chapter 3

System Identification: Exploiting Wavelets

3.1 Introduction

Our approach is to base our model on mathematical equations representing only basic electromagnetic scattering phenomena, and to apply concepts of digital signal processing and wavelet filter banks [36]. In fact, as the nature of the problem has been proven to be of a linear type and validated at a later stage through a state space analysis [97], the theory of linear time invariant systems can introduce numerous techniques dedicated to system identification and noise filtering [60]. In particular, the least-squares method has demonstrated promising results in digital signal processing in geophysics [62] [78].

In general, parameter estimation and identification are usually described within a probabilistic framework. Here, we basically employ such a framework, however we attempt to allow inclusion of *a priori* information or knowledge. In Figure 1.1, the overall synthesis framework including the system identification prior to data inversion is shown. Pre-processing is included in the system identification.

3.2 Wavelet Theory and Information Capacity

The purpose of this section is to present some basic concepts and constraints of identification from an information theory point of view. *Information* may be defined as a measure of the degree of uncertainty of some process; i.e. the more uncertainty there is in a process, the less informative the process. On the other hand, uncertainty does not necessarily imply the lack of possible information but rather the lack of capacity to retrieve the information. The best analogy can be related to the science of cryptology. The encrypted text contains the same information as in its original form, though it is only in the second case that there is some potential of information retrieval.

Communication or information theory is a large subject, and it is not the intention of this thesis to explore all of its aspects. However there are a few basic principles that we are going to investigate in an informal way [24] [8]. Whether in Weiner's communication theory [93], Shannon's information theory [83] or Gabor's communication theory [29], time series signal principles and constraints in information theory are shared. The constraints in information theory have a positive impact on the knowledge extraction from a probabilistic perspective, as will be described in Section 5.3.

As there are various types of signals and, in particular, power signals, which are represented by *time series*, the concept of *wavelet* theory was introduced for time series signals with additional restrictions. For electromagnetic imaging, the concepts of wavelets can be characterized by two properties:

1. The *one-sided property*: A wavelet has zero wavelet values before its origin.

2. The *stability property*: A wavelet has finite energy; that is, mathematically speaking, $\sum ||y(n)|| < c$, for all n and c is a finite constant.

Another approach is the Fourier transform of a set of filter coefficients b_j given by

$$Y(\omega) = \sum_j b_j e^{ij\omega}. \tag{3.1}$$

Here Y is a function periodic in 2π and it defines the wavelet $y(n)$ in Fourier domain.

Hence, in contrast to time series, wavelets are self-contained with a definite origin and arrival time. In other words, a wavelet has a finite time independent *index*. A wavelet in electromagnetic imaging is the measurement data set subsequent to a finite emitted pulse. Wavelet theory is extensively used in seismic applications [78] [14].

The retrieval, by means of coherent techniques, of information from wavelets has been demonstrated [46]. Although we will use the same principle of system identification, we will further explore additional concepts of uncertainty for optimal information retrieval (Section 5.3). However, in this chapter the significance of wavelet theory is that the measured signal can present a maximum capacity of information and in order to demonstrate the potential of information retrieval from an uncertainty perspective, we may realize that the problem is related to Heisenberg's uncertainty principle [76]. In general, the problem can be divided into two categories:

1. Linear wavelet identification

2. Uncertainty and information retrieval

The following part of the chapter will concentrate on the system identification aspects whereas the second category will be approached in Chapter 5. Wavelet identification is governed by the principle of separability in the frequency domain such that

$$Y(\omega) = H(\omega) \prod_i S_i(\omega) \tag{3.2}$$

and hence system identification is to find the realizability of the inverse filters that will solve for $H(\omega)$ or

$$H(\omega) = Y(\omega) \prod_i \frac{1}{S_i(\omega)}. \tag{3.3}$$

3.3 System Identification Methodology

System identification concepts developed in this research rely exclusively on the measured data. As a result, assumptions regarding the radar waveform and the nature of noise involvement are reduced to a minimum. In fact, we will consider these variables to be associated with the identification procedure of the characteristics of the system being investigated.

System identification can also be hardware-dependent. For example, since the scattering and measurements are performed with a pseudo-single antenna, data becomes highly correlated and excessive symmetries are found. When the source and receiver are coincident, the path of the energy from the transmitter is retraced exactly by the energy coming back to the receiver. Known as *zero offset* imaging, it simplifies the inverse problem and spatial distortion recovery by knowledge of the travel time which is exactly doubled. Another important aspect is that only normal incident wave fields are reflected and measured [1] and hence the reflection index assumptions are not required. The concept of zero offset was first introduced in seismic applications as described by Cassinis [13]. In fact, due to the similarities in the principle of wave propagations, the same concept of the zero offset is embodied in modern ground penetrating radar designs. Reducing uncertainties and assumptions allows an increase in imaging resolution and reduces the computational costs allowing the freed time to be used in resolving spatial distortion relating the wave velocities [64].

[1]Refracted signals are of higher order and can be eliminated. Reflected refractions are identified by their weak signal energy compared to their first reflection through correlation.

3.4 Input-Output System Description and Modeling

The input-output description of a system consists of deriving mathematical expressions which specifically define the relation between the input and output signals. An electromagnetic signal $x(t)$ is emitted at t_0 and is of finite length; a signal $y(t)$ is recorded for an interval T units of time starting at t_0.

Let $h_1(t)$ define the characteristics of the medium in the direction of the transmitted wave, and $h_2(t)$, the characteristics of the medium in the opposite direction for the interval T or

$$h_2(t) = h_1(T - t) \tag{3.4}$$

where $t_0 + T$ becomes the cutoff time. In other words, if a signal excitation $x(t)$ is emitted and a measurement $y(t)$ is recorded in the interval T, the signal uses at maximum $T/2$ units of time to penetrate the medium and $T/2$ units of time to return. In addition, the signal employs exactly the same path in both directions.

Also, the signal that penetrated the medium was affected by the material characteristic $h_1(t)$ until it was reflected, and affected again by $h_2(t)$ and finally measured. In general such behavior is represented by the convolution of the signal with the medium characteristics and

$$y(t) = h_1(t) * h_2(t) * x(t) \tag{3.5}$$

and expanded, given a *zero offset* system , to

$$y(t) = h_1(t) * h_1(T - t) * x(t) \tag{3.6}$$

which displays an autocorrelation sequence, and

$$h(t) = r_{h_1 h_1}(t + T). \tag{3.7}$$

In reality, the signal $y(t)$ is sampled and modified arbitrarily to yield the wavelet $y(n)$, which is the only measurement recorded. Hence, it is relevant to assume a corresponding sequence $x(n)$ and $h(n)$ where k describes the index each time a measurement is recorded and

$$y_k(n) = s(n)[h_k(n) * x(n)] \tag{3.8}$$

where $y_k(n)$ is multiplied point-by-point by the arbitrary sequence $s(n)$. Here, $s(n)$ has no physical meaning, but rather introduces a gain to compensate for the attenuation of electromagnetic waves as a function of distance.

3.4.1 Estimating Noise Corruption

Suppose that an electromagnetic excitation input sequence $x(n)$ is repeatedly used and $y_k(n)$ is the set of measured sequences. In reality, the current problem is that measurements are corrupted with noise as shown in Figure 3.1, and the objective is to estimate the noise spectrum. In general the measured signal $y_k(n)$ may be expressed as

$$y_k(n) = h_k(n) * x(n) + e_k(n) \tag{3.9}$$

where k is the k^{th} scan, $h_k(n)$ depends on the medium characteristic and $e(n)$ is the corruption noise. In the z-domain transform

$$Y_k(z) = H_k(z)X(z) + E_k(z). \tag{3.10}$$

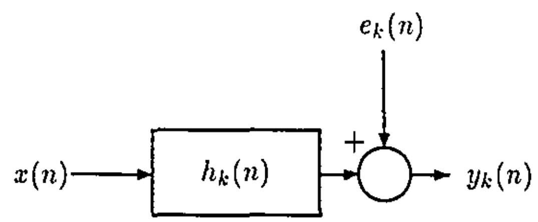


Figure 3.1: Model representation with noise corruption. In the proposed model, we only assume measurement noise.

Assume the fact that the scans are consecutive and narrow enough to have the variation between $h_{k-1}(n)$ and $h_k(n)$ negligible. As a result the fractional noise change becomes sharp and locally identified

$$Y_k - Y_{k-1} = (H_k - H_{k-1})X + E_k - E_{k-1} \tag{3.11}$$

which leads to

$$\lim_{\Delta_{k,k-1}\to\epsilon}(Y_k - Y_{k-1}) = E_{k,k-1}(z) \tag{3.12}$$

which is the zero mean noise spectrum and ϵ is an arbitrary value. The inverse z-transform $e_{k,k-1}(n)$ can be considered the best local estimated noise and thereby, global noise filtering can be achieved. What we have introduced here is a dynamic adaption to a time-varying noise spectrum.

The validity of this approach is to investigate the nature of $e_{k,k-1}(n)$. The auto-correlation sequence of $e_{k,k-1}$ should resemble

$$r_{e_{k,k-1}e_{k,k-1}}(l) = \begin{cases} E_{e_{k,k-1}} & l = 0 \\ 0 & l \neq 0 \end{cases} \tag{3.13}$$

if $e_{k,k-1}(n)$ is white noise sequence, and $E_{e_{k,k-1}}$ is a constant. What is of interest is to examine the effect of the distance of the scans on the noise estimation, which can be accomplished by evaluating the sequences $e_{k,k-1}(n)$, $e_{k,k-2}(n)$, $e_{k,k-3}(n), \ldots$ and comparing their corresponding auto-correlations.

3.5 Deconvolution in Remote Sensing

A large class of systems can be represented by a convolution of the input and the system impulse response. Even the system impulse response can be represented as a convolution of impulse response functions of many subsystems. The problem of deconvolution has been solved by many different approaches, as for some time this was a major issue in seismic applications. Among others we can identify the following filtering techniques: Weiner filtering [92] and Homomorphic deconvolution [90].

3.5.1 Example: Least Squares Prediction in Radar

Assuming the signal $y(n)$ was recorded at the antenna for an excitation sequence $x(n)$, the output wavelet may be written as

$$y(n) = \sum_{i=1}^{L} k_i x(n - D_i) \tag{3.14}$$

where k_i is the reflection coefficient at the ith depth and D_i represents the corresponding propagation delay. The number L is the number of points recorded. The received signal $y(n)$ can be viewed as the convolution of the excitation $x(n)$ with the sequence

$$h(n) = \sum_{i=1}^{L} k_i \delta(n - D_i) \tag{3.15}$$

where $h(n)$ is a function of the characteristics of the medium. What is of interest is to recover the sequence $h(n)$ from the measured sequence $y(n)$. Consequently the problem becomes an inverse filtering to remove the excitation effects of $x(n)$. In reality, the current problem of inverse filtering is more complex. The *ghost patterns*, which are a series of reverberations, are known to strongly mask the measurements which are represented by the sequence $c(n)$. In general, when all variables are considered, the received signal can be represented as

$$y(n) = x(n) * c(n) * h(n). \tag{3.16}$$

As a result, the sequences $x(n)$ and $c(n)$ are to be eliminated. Let

$$p(n) = x(n) * c(n) \tag{3.17}$$

as a single sequence to be eliminated. The sequential effect of a signal path may be expressed as

$$p(n) = x(n) - cx(n - D_i) + c^2 x(n - 2D_i) - c^3 x(n - 3D_i) + ... \tag{3.18}$$

and in the z-transform domain

$$P(z) = (1 - cz^{-D_i} + c^2 z^{-2D_i} - c^3 z^{-3D_i} + ...)X(z) \tag{3.19}$$

which converges to

$$P(z) = \frac{1}{1 - cz^{-D_i}} X(z). \tag{3.20}$$

Therefore, the system function for the ghost pattern can be expressed as

$$C(z) = \frac{1}{1 - cz^{-D_i}}. \tag{3.21}$$

Since the coefficient $c < 1$, the system is stable and its inverse is FIR and minimum phase. The deconvolution is designed to remove the effect of $x(n)$ and $c(n)$ simultaneously. To derive the inverse system for deconvolution, it is necessary to adopt a well based statistical approach. The assumption is that the sequence $h(n)$ is uncorrelated [92] [78] and consequently has the following autocorrelation

$$r_{hh}(l) = \begin{cases} E_h & l = 0 \\ 0 & l \neq 0 \end{cases} \tag{3.22}$$

where E_h is some arbitrary constant. On the other hand, the sequence $p(n) = x(n) * c(n)$ is highly correlated. As before in the previous section on noise estimation, we will assume the fact of consecutive narrow scans for small variations in the output observation and, hence, we can predict future observations. If $y(n)$ is the observed output sequence of the real system, let $\hat{y}(n)$ be the output of the predicted model, and

$$\hat{y}(n) = \sum_{k=0}^{M} b_k x(n-k) \tag{3.23}$$

where b_k is the prediction coefficient. Sine the previous equation describes an FIR filter, of length M and coefficients b_k, the filter coefficients are selected to minimize the sum of the squared-error sequence, that is

$$\varepsilon = \sum_{n=0}^{\infty} \left[y(n) - \sum_{k=0}^{M} b_k x(n-k) \right]^2 . \tag{3.24}$$

The minimization of ε with respect to the coefficients b_k leads to the equations of the form

$$\sum_{k=1}^{M} b_k r_{yy}(k-l) = r_{yy}(l) \tag{3.25}$$

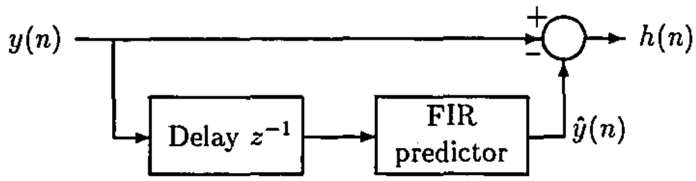


Figure 3.2: Linear prediction and deconvolution. The FIR predicts a sequence $\hat{y}(n)$ from past samples and subtracts it from the observed values to yield the desired sequence $h(n)$.

where $r_{yy}(l)$ is the auto-correlation of the sequence $y(n)$ defined as

$$r_{yy}(l) = \sum_{n=0}^{\infty} y(n)y(n-l). \tag{3.26}$$

We can express again the linear equation in the matrix form as

$$\begin{bmatrix} r_{yy}(0) & r_{yy}(1) & \dots & r_{yy}(M-1) \\ r_{yy}(1) & r_{yy}(0) & \dots & r_{yy}(M-2) \\ . & & & \\ r_{yy}(M-1) & r_{yy}(1) & \dots & r_{yy}(0) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ . \\ b_M \end{bmatrix} = \begin{bmatrix} r_{yy}(1) \\ r_{yy}(2) \\ . \\ r_{yy}(M) \end{bmatrix} \tag{3.27}$$

or equivalently, as

$$\mathbf{R_{yy}b} = \mathbf{r_{yy}}. \tag{3.28}$$

These equations, often called the *Yule-Walker equations* [62], admit an efficient solution due to the Levinson and Durbin algorithm [57]. Since the correlation sequence $\{r_{hh}\}$ is an impulse, it follows that $\{r_{yy}\} = E_h\{r_{pp}\}$. In view of the foregoing correlation, the FIR filter predicts the sequence $p(n)$ from past samples of $y(n)$. The predicted $\hat{y}(n)$ is basically an estimate of $p(n)$, and is subtracted from the observed values to finally yield the desired sequence $h(n)$ as shown in Figure 3.2.

Chapter 4

The Inverse Problem

> The moment one begins to investigate the truth of the simplest facts which one has accepted as true, it is as though one had stepped off a firm, narrow path into a bog of quicksand – every step one takes one sinks deeper into the bog of uncertainty.
>
> *L. Woolf.*

In formulating the solution techniques for the inverse problem as well as for the inverse-scattering problem, an important consideration has been to maintain their capability of being generalized to more complex configurations as much as possible. A consequence of this approach is that the conclusions obtained from the abstract formulation must be verified by computing numerical results for adequate inversions.

The inverse problem is a complicated process. Techniques vary in their approach to solving the inverse problem. Trial and error techniques are too cumbersome. Analytical inversion of the direct problem does not leave much space for a realistic application and is possible for few cases, and even then meets with many difficulties. Problems such as uniqueness and stability are to be dealt with. There is a branch of mathematics in which problems of uniqueness and stability have been studied extensively: Linear Algebra.

4.1 Introduction

Our dull description of an inverse problem was based on terms of common knowledge in scattering and slightly on terms of electromagnetic scattering, but, with a little stretch of the imagination, the analogy can be transposed to a generic infinite scattering phenomenon with a finite measurable space.

The interpretation of scattered radiation is rarely straightforward. Often the information contained in the raw measurement data requires a considerable degree of processing before a sensible conclusion can be made about the object. Indeed, it is only after the implementation of some inversion procedures, which necessarily contain within them a description of the scattering mechanism, that the measurement can be transformed into an understandable cognitive form. For example, the images of an object produced by two electromagnetic detection devices operating at different frequencies reveal distinct differences when compared, despite their both having emerged from antennas and being formed from scattered radiation. More restrictively, some electromagnetic signals can be manipulated more easily than others, whereas other wave fields can provide valuable information despite the encountered difficulties in their manipulation.

In the development of electromagnetic theory, emphasis was placed on discovering the implications and exploiting a few specific applications of the theory. As a consequence, all effort was invested on the forward problem, that is, on finding solutions to Maxwell's equations rather than on the corresponding inverse solutions.

The individual constraints and challenges imposed by electromagnetic theory have led to different inverse solutions. In recent years, the emphasis has shifted from solving the exact inverse problem to adapting techniques which can be applied with confidence in certain areas which have greater flexibility with relation to the measurements.

4.2 The Inverse Problem Formulation

Any scattering problem can be represented in terms of a mapping between certain sets of functions. Before presenting the inverse problem itself, it is appropriate to consider the forward problem within this context. The direct problem consists of finding how a set of functions p forming the elements of the parameter set P, which describes the scatterer, are mapped to the set of functions d, being the elements of the data set D and hence describing the scattering phenomenon in its conceptual form. The mapping A is an operator which acts on the elements $p \in P$ to produce the data set $d \in D$, or

$$A : P \to D. \tag{4.1}$$

Provided that the operator and parameter set are known, the data set can be defined as:

$$D = \{d : A(p) \to d\}. \tag{4.2}$$

This formal definition of the forward problem can be restated in words such as the set D is the collection of elements d such that the operator A maps p to d. Solution of the inverse problem can be defined as finding the inverse mapping and the inverse operator A^{-1} which constructs the parameter set from the data set. In this sense, the inverse operator performs a reconstruction of the parameter set and, again, provided the inverse mapping and data sets are known, the parameter set can be defined as:

$$P = \{d : A^{-1}(d) \to p\}. \tag{4.3}$$

These operations are shown schematically in Figure 4.1.

If these general concepts are related to a real scattering experiment, then the complicated relationship between the mapping and the respective sets soon becomes apparent. First of all, consider the parameter set. This set of functions is unlikely to represent a complete description of the object, being deficient in specific parameters which may contribute to measurements, such as geometrical constraints. In addition,

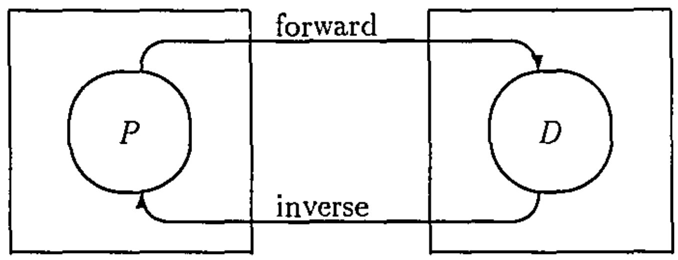


Figure 4.1: Demonstrating the forward and inverse problem as a mapping between parameter and data sets P and D respectively. The mapping itself is a representation of the scattering phenomena where forward and inverse are function operators.

measurements are made from a finite set of wavelets or finite discretization of spatial and temporal domains, and this inevitably means that we cannot deny the data incompleteness. Specific to radio detection and ranging, knowledge of the source signal is also deficient and results in an additional complexity in the accuracy of the forward problem itself prior to any inversion.

The mapping itself is a representation of the scattering phenomena, usually deduced from a knowledge of the propagation characteristics. For the majority of electromagnetic applications, including radio detection and ranging, the mapping is derived from a wave operator. It is important to realize that this wave operator is linear in the forward mapping problem and nonlinear in the inverse mapping problem of the data set to the corresponding parameter [45].

One critical aspect of inverse solutions is the effect that inaccuracies in the measured data have upon the accuracy of the reconstruction. These inaccuracies can be introduced by the measuring equipment. The problem of errors can be considered of high importance and is treated on its own as there is no intention to study the impact of noise on the inverse solution. Some noise analysis was presented in Section 3.4.1.

4.2.1 Labelling the Inverse Problem

Many issues have obvious bearing on the validity of an inversion in the existence of the reconstruction, that is, the inverse mapping maps elements from the data set to those elements which are proper members of the parameter set. One wishes that the inversion should also be unique, in the sense that every element of the data should correspond to a single element in the parameter set. However, it will become evident that the mappings describing inverse scattering in electromagnetic applications are nonlinear, indicating the existence of more than one inverse solution to the scattering problem. In this thesis we classify the inverse problem into four categories.

Definition 4.1 (Inverse Uniqueness) *A unique inversion is one in which every element of the data set should correspond to a single element in the parameter set.*

$$P = \{p : A(p) \rightarrow d; A^{-1}(d) \rightarrow p\} \tag{4.4}$$

Definition 4.2 (Inverse Stability) *A stable inversion is one in which an infinitesimally small change in the data set gives rise to correspondingly small changes in the parameter set. If the changes in the parameter set are large, then the inversion is deemed to be unstable.*

$$P = \{d : A^{-1}[d + \varepsilon] \rightarrow [p + \epsilon]\} \qquad \varepsilon, \epsilon < \infty \tag{4.5}$$

and ε, ϵ are arbitrary values.

Definition 4.3 (Inversion Divergence) *A divergent inversion is one in which, for every element in the data set, there is at least more than one corresponding element in the parameter set. Hence the divergence factor of the inversion is the ratio of the corresponding element in the parameter set to the whole set.*

$$P = \{p : A^{-1}(p) \rightarrow \alpha P\} \qquad 0 \leq \alpha \leq 1 \tag{4.6}$$

where alpha is the divergence factor.

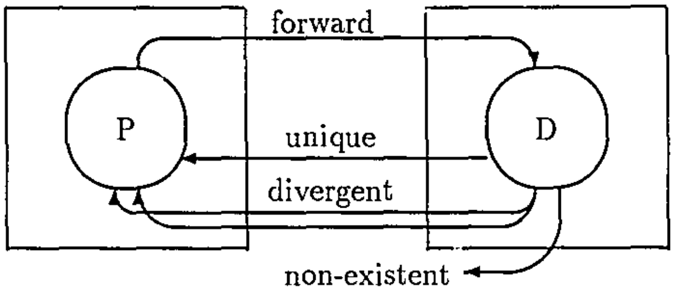


Figure 4.2: Depiction of a unique, divergent and ill-posed inverse problem. A perturbation of the set can lead to an unstable solution which is a point outside the parameter P space but with some feasible dimensions. The ill-posed inverse problem occurs when the reconstructed parameters lie outside the domain of feasible dimensions.

Definition 4.4 (Ill-posed Inversion) *An ill-posed inversion is one in which there is at least one element in the data set that does not have at least one corresponding element in the parameter set.*

$$P = \{\exists d : A^{-1}(d) \rightarrow \{\}\} \tag{4.7}$$

In general, ill-posed problems are the inverse problems that do not admit solutions [27], that is, numerically, the inverse operator A^{-1} does not exist or is the null operator [2]. However, in the realistic aspect of electromagnetic scattering, the unique inverse problem and the ill-posed inverse problem will not be stressed in this thesis.

An ill-posed problem is schematically described in Figure 4.2. The inversion of the data does not yield anything in the parameter set P. However, in Figure 4.2, the same element deriving from the data set D breaks off into two possible elements in the parameter sets. The consideration of solving a divergent inverse problem is at best ambiguous. One reason is that the inverse problems in electromagnetic scattering are divergent problems by definition of the scattering wave operator. In other words, in electromagnetic imaging the formulation of an inverse operator is feasible at least in an approximate perspective.

4.3 The Backus-Gilbert Solution

Considering that initially there are a finite amount of sensor measurements, the hope that there exists only one satisfying model is vain. In fact, Backus and Gilbert have proved [2] that the set of models satisfying the finite amount of data is either empty or infinite. As Backus and Gilbert point out, the uncertainty in the final model results mainly from the finite number of measurements and provides a basic insight into the approach of how to choose the model which is the "smallest" in a least-square sense, that is, to minimize the Euclidean norm of the sensor measurements [3] [4].

Similar world modeling methodologies introduce some concepts of linear regularization of elements as they seek to maximize the stability of the solution. The Backus-Gilbert method looks at the relationship between the solution and measurements and proceeds to minimize what is called the *resolution function kernel.* Once the model has been calculated, it remains to determine its uncertainty which is computed as a function of the smallest model and the magnitude of the image. As a result, it becomes clear that there is a trade-off between the uncertainty and the resolution range of the image. This process is termed *regularization* in inverse problem nomenclature. Nowadays, Backus-Gilbert is often recommended as the generic method of choice for designing and predicting the performance of experiments that require data inversion.

4.3.1 Resolution Limits and Resolving Power

This thesis plainly defines resolution as the ability to distinguish between two closely spaced parameters. A *spatial resolution* defines more specifically the minimum distance between the spatial coordinates in the parameter space. Forshaw and others [26] discuss alternate definitions of spatial resolution. Resolving power and spatial resolution are closely related concepts. The term *resolving power* applies to the imaging and transformation procedure, whereas spatial resolution applies to the Cartesian coordinates produced by the imaging or transformation procedure.

In the inverse problem, the resolving power is of interest since the inversion is validated by the spatial resolution. In fact this concept is broadly applied in the forward modelling and residual error computation in the least square sense where the resolution is set to a numerical threshold. Though some computational perspectives become redundant, since the correct validation is the visual perception of the image rather than the algorithm's convergence, this discussion is expanded in Chapter 8 from an image perception concept. In this thesis our concern is on the qualitative understanding of the resolving power that leads to the spatial resolution knowledge maps.

The advantage of these approaches is that, in principle at least, they are applicable to all remote sensing and scattering phenomena, with any amount of available information. More information simply leads to better numerical stability, faster convergence, larger tolerance and probably uniqueness, although the last property may be difficult to achieve in a complicated situation, even with sufficient data.

In any real life situation, noise and other uncertainties associated with a measuring process are always inevitable. However, there are few authors who have concerned themselves with the resolution limit imposed by these uncertainties [4]. One explanation may be that approximations of one sort or another must be made in all the methods used. Hence the resolution limits are actually governed by the degree of approximation rather than the measurement uncertainties.

According to the Backus-Gilbert method, the set of models satisfying the finite amount of data is either empty or infinite. In our context of remote sensing in general, the Backus-Gilbert solutions sets are of the ill-posed inverse type. However, the inverse problem in electromagnetic scattering admits a finite amount of solutions which therefore are of a divergent inverse type. The resolution limit in our context is to determine uncertainty factors of the estimated parameters.

4.4 Inverse Problem: A Qualitative Approach

The starting point of our inversion investigations is to clearly define the inversion problem and what is exactly required of the inversion process. As the proposed inversion process described in this thesis is applicable in a generic fashion, our investigations revolve around the macroscopic description of Maxwell equations in their application in radio detection and ranging. There is no doubt that the proposed concepts in this chapter can be validated in areas other than remote sensing which involve somewhat incomplete data inversion.

There are no methods at our disposal to correct the inverse of ill-posed problems, however, we can restrict the divergent inverse problem to one from a stability sense (Figure 4.3). The study of the convergent inverse problem in its practical context has led to the development of techniques which incorporate constraints on the inversion or reconstruction operation which have some physical relevance to the problem being considered. These constraints consist of information which is independent of the data set and therefore should be known *a priori.* Constraint is a useful description since it conveys the correct impression that the information is used to confine the inversion process to solutions which are considerable or sensible in electromagnetic applications and radio detection and ranging.

In Section 4.2, we have briefly mentioned the effect that limited and erroneous data can have on the qualitative inversion. A proper examination of the scattered field data is therefore of major importance to the successful implementation of an inverse solution. As a matter of fact, the theoretical development of inverse scattering solutions might profitably be influenced by the quality of the measured data. It is also true that compromises must be made in the implementation of the inverse scattering

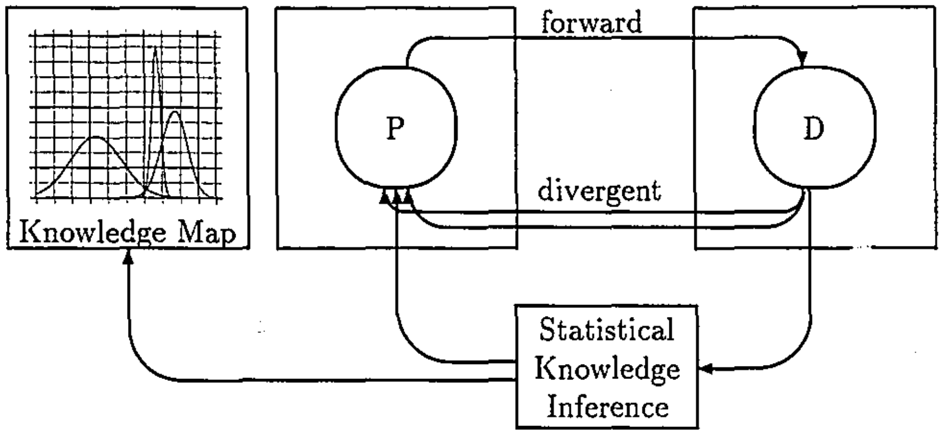


Figure 4.3: Schematic illustration of knowledge inference to stabilize divergent inverse problems. Knowledge inference accepts many types of *a priori* information about either the data or the parameters and may include constraints of conformity to certain statistical distributions and sensing resolutions. Manipulating the *a priori* information results in the *regularization synthesis*, although the derived solution is not a unique solution but rather some *a posteriori* knowledge sets that provide *knowledge maps* of the parameters.

theories. The art of this thesis subject is to ensure that, in making such compromises, some knowledge is retained in a descriptive manner. Hence, qualitative inversion can be performed, and this process will identify the particularities of the application.

In this thesis we have adopted a different perspective to approaching the inverse problem. The implication of knowledge inference and higher order statistical interpretation of the inverse problem provides a qualitative approach to associating measurement truth parameters with understanding. As the proposed technique for the inverse problem in this thesis is given the generic term *synthesis and regularization* method, we will show in future chapters the inversion scheme based on knowledge inference methodologies that exploit concepts of Information Classification (IC) and Information Visualization (IV). Similar work has been approached by Zucker and others [99] [47].

4.4.1 The Inverse Problem in Radio Detection and Ranging

Although measured data consist of real numbers, it is often convenient to analyse the properties of signals by constructing the analytical signal which is the extension of the real signal in the complex plane. In this form it is straightforward to deduce the signal amplitude and phase which are natural descriptions of electromagnetic wave fields. In most circumstances, however, only the amplitude of the scattered field is recorded. In this case, the inverse computation of the phase field from the direct measurements obeys the uniqueness and stability requirements and is directly computed as wavelets [36] [16] [77]. Section 7.3.2 describes the computational approximation of the phase signal as the application to ground penetrating radar. The deconvolution of the measurement data into two separate data sets of phases and magnitudes is used associatively in the inversion process. The deconvolution example in Section 3.5 can be transformed to a generalized deconvolution prediction process and therefore solved for the phase. Other techniques can be employed in estimating the phase from the measurements by a minimal knowledge of the electromagnetic field. What prompted the topic of phase retrieval from the amplitude data is the verification of an assertion by researchers [97] disputing the minimal phase of recorded signals as proposed by [92] and [78]. The phase may contain coherent information and may not have been properly investigated in ground penetrating radar applications [35].

In order to be able to handle divergent inverse problems adequately in electromagnetic scattering, one must use the recorded measurement to its full extent. In other words, maximum knowledge has to be extracted from the measurements and used thereafter in estimating the model parameters and distributions 4.4. However, one must bear in mind that the solution is only as good as the accuracy of the measured data and optimal knowledge separation and inference. There is no dispute about the mathematical model of the inverse operator in radio detection and ranging, although one must note that the essential aspects of divergent inverse problems

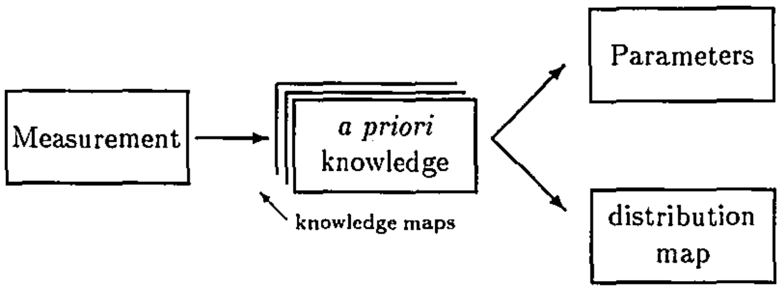


Figure 4.4: *A priori* knowledge has to be extracted from the measurements and used thereafter in estimating the *a posteriori* model paramcters and distributions.

reside in handling the inverse operator solutions.

4.4.2 Synthesis and Regularization

Statistical inversion methods have recently been shown to be able to produce essential new information in most remote sensing applications [53]. The inversion problem we are approaching is the inversion theory for multi-valued variables, which have some dependent probabilistic distributions. In fact, this would make it possible to handle white noise in a convenient way. But, mainly, the estimated inversion could be described by infered stochastic processes which could clarify the interplay between resolution and accuracy.

In our model, we construct the knowledge space by standard deductive methods of the hyperbolic inverse operator but we may have no source of information about the weights of the propagation (e.g. aperture angle). The question is addressed whether it is possible to compute accurately the measure of uncertainty of the resulting inverted *knowledge map*. For example, it can be proven that the longer the hyperbolic sensor line (the inverse path of point estimates), the larger the evidential space and the more certain a conclusion. However, technical descriptions will be kept for the application to radio detection and ranging chapter (Chapter 7).

As the conceptual approach of the regularization synthesis is analysed, the mathematical formulation of the inversion procedures will be discussed in the next two chapters. The philosophy of the regularization synthesis resides in statistical inversion theory with the initial constraint that measurement data do have an *a priori* knowledge. Typical measurement data resulting from a forward problem as defined in equation 4.2 may be represented by series of indexed data, or

$$D = [u^i(\mathbf{h}, t)] \qquad i = 1, 2, \ldots \qquad t_0 < t < t_n \tag{4.8}$$

where u^i is the measurement space, and $\mathbf{h}$ denotes a generic measurement wavelet time indexed by t. Considering the problem is linearly separable, $\mathbf{h}$ is broken down to one or more of its components such that

$$[u^i(\mathbf{h}, t)] = \sum_k ([u^i(\mathbf{h_k}, t)]). \tag{4.9}$$

Therefore the probability space where u^i is measured can be expressed by the *a priori* density functions

$$S_k^f = \frac{[u^i(\mathbf{h_k}, t)]}{\sum_k [u^i(\mathbf{h_k}, t)]}. \tag{4.10}$$

On the other hand, the inverse problem as in equation 4.3 can be expressed as

$$P = [v^i(\hat{\mathbf{h}}, t)] \qquad i = 1, 2, \ldots \qquad t_0 < t < t_n \tag{4.11}$$

where u^i is the computed space and $\hat{\mathbf{h}}$ denotes a corresponding estimated wavelet time indexed by t. As the problem was assumed to be separable, this leads to

$$[v^i(\hat{\mathbf{h}}_\mathbf{k}, t)] = f_k^{-1}([u^i(\mathbf{h_k}, t)]) \tag{4.12}$$

where $\hat{\mathbf{h}}_\mathbf{k}$ is the estimated wavelet and f_k^{-1} is some inverse operator function. Hence the collection joint probability density functions of the estimated wavelets are expressed as

$$S_k^{f^{-1}} = \frac{[v^i(\hat{\mathbf{h}}_\mathbf{k}, t)]}{\sum_k [v^i(\hat{\mathbf{h}}_\mathbf{k}, t)]}, \tag{4.13}$$

which is a good way to give marginal densities of inverse problems with respect to each inverse operator involved in the process.

The approach we are taking towards developing the regularization synthesis, which may be used to represent all of the various symbolic and numeric aspects of *a priori* knowledge delineated by beliefs and uncertainty, is to consider a logic of argumentation. We extend the logic so that not just one argument, but all arguments, supporting or opposing an inversion hypothesis are considered in a given decision-making context, that is, the logic used to structure the inverse problem. We hold this to be the key component of the practical inversion synthesis and regularization. As arguments are identified, the support they confer on a hypothesis or its negation is aggregated to provide a measure of the degree of belief in the hypotheses of interest. The aggregation operation will depend on the calculus used to represent the uncertainty or vagueness associated with the arguments. The choice of calculus will in turn depend on the representation requirements and the information which is available from the given *a priori* knowledge maps. By inversion theory, the synthesized inverse solution is objectively an *a posteriori* density function regardless of the innovation of the regularization factor (rf) (Section 5.4) that largely influences the knowledge map density functions. As the *a priori* densities are approximately constant and the knowledge maps densities are resolved, the problem can be formulated as the minimization of the norm of the covariance matrix $[cov_{f,f^{-1}}]$ or

$$\boxed{Min: \quad \|E\left[\left([u^i(\mathrm{h},t)] - \mathbf{S}^f\right)\left(f^{-1}([u^i(\mathbf{h},t)]) - \mathbf{S}^{f^{-1}}\right)^{\mathsf{T}}\right]\| \qquad (4.14)}$$

where E is the expected value.

The analysis of evidence supporting hypotheses is a promising framework for drawing conclusions efficiently without losing resolution limits. Two different assumptions lead to stable expressions of the *a posteriori* knowledge maps derived from the hypotheses based on the belief functions. The first assumption is one of independence of each measurement or evidence from all other measurements (Chapter 5). The second assumption is one of independence between each measurement and the hypotheses derived from the remaining evidence. The required independency in equation 4.14 is not an assumption but rather a requirement of the inverse problem. In reality, the independency between the *a priori* and *a posteriori* knowledge sets only guarantees a solution to the inverse problem [53] [27] for all the inverse operators for maximum independence. The analogous operation yielding maximum *a posteriori* estimates is found in many stochastic relaxation processes with Bayesian restoration in pattern analysis [1] [31]. A common interest of this thesis and the stochastic relaxation is stability. Image restoration by a maximum *a posteriori* estimate by annealing [31] is in fact creating independence between the marginal distributions.

The ability of deriving *a posteriori* knowledge map distributions provides many possibilities for approaching various statistical testing problems [88]. In our derivation, when there is more than one knowledge map to choose from, each of which has its own *a priori* spaces for its corresponding variable, an inversion approach might be formulated by considering the union of these *a priori* spaces. The test would be based on calculating the density functions of the inverted space and minimizing the partial density function dependencies as in equation 4.14. In other words, the objective is to utilize the density functions as in equation 4.13 in the synthesis process which can be carried out in optimal fashion through Kalman Filtering which will be expanded in Chapter 6.

4.4.3 Discussion

It is true, nonetheless, that a highly divergent inverse problem needs more regularization, that is, a higher order complex knowledge map, than a problem which is only weakly divergent. We can remark also that the complex derived knowledge maps for weakly divergent inverse problems are trivial in the sense of the excess of information they represent. The major disadvantage associated with imposing trivial knowledge constraints is the cumbersome mathematical and programming methods required to determine the numerical solution when the solution already exists.

A basic difficulty of this kind of method is that one has to specify adequate *a priori* distribution knowledge. As the probability law that governs measured data is unknown, numerical modeling is needed to define the separate variables involved in the forward problem, and hence form the *a priori* knowledge map for every identified variable.

The numerical solutions constructed using high order synthesis regularization as in Chapters 5 and 6 converge non-iteratively in a consistent manner, provided of course that the inverse problem is at worst a stable and divergent case. As far as we can see, application of the proposed inverse problem solution will require, sooner or later, the experimenter to speculate constructive information about the proposed solution from its numerical representation, as will be described in Chapter 8.

Chapter 5

Synthesis: A world of Beliefs

> In maximum entropy restoration, the entropy of a physical system in some macroscopic state of an image is the logarithm of the microscopically distinct configuration knowledge nodes leading to the same macroscopic image. In some situations there is reason to believe that the entropy of a *stable* system is of lower entropy whereas any *ergodic* behavior only increases the entropy.
>
> *adapted from Boltzmann.*

5.1 Introduction

One of the important aspects of using evidential reasoning concepts to solve inverse problems lies in the representation of knowledge uncertainty. The variety of theories and models for the representation of knowledge with uncertainty factors has been addressed often in research [11] [19] [73] [81] [98]. Several optimality criteria, such as maximum likelihood and minimum entropy, have been used in handling the uncertainty of knowledge. Shannon [82] introduced in 1948 the notion of entropy as a measure of uncertainty in information theory, and, since then, this concept has been intensively applied in probability contexts. The entropy computation concept is also used in the current framework to evaluate the uncertainty and in fact forms the emphasis of the regularization synthesis.

Although the concept of entropy as a representation of uncertainty factors can be used in a large range of applications, we employ in the derivation and validation of the proposed regularization synthesis method a specific model in knowledge representation. The current work draws on research in knowledge inference. In particular, it is validated by the methodology of empirically constructed knowledge networks (i.e., inference networks) [58] [59]. Such networks serve as a basis for making inferences about knowledge assertions where knowledge maps can be extracted. The present study basically employs similar maps by augmenting the implications with certainty measures and optimization methods.

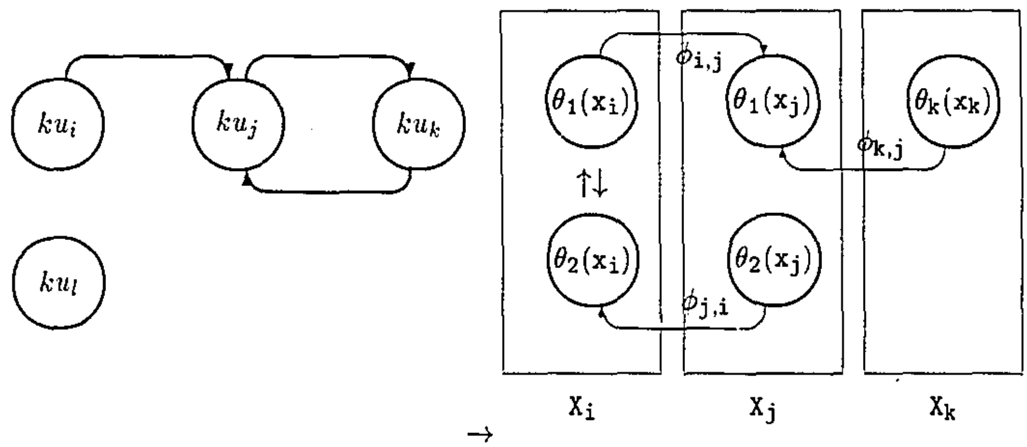


Figure 5.1: Typical networks. Right: Node X_i, X_j and node X_k are the frame of discernment in Dempster-Shafer. Left: Any two knowledge units can be linked in a single direction, in both directions or not linked at all.

The described electromagnetic image inversion problem fits into the general objective to develop a method of building a *knowledge map* from sensor based data. To approach the inverse problem as an issue, it is necessary to analyse the simultaneous effect of both types of characterization of the *belief functions* and the *evidence propagation* scheme as they all affect the accuracy and variability of the reasoning.

Therefore, the objective becomes that of addressing the optimization of the belief functions in the construction of the knowledge map. In other words, the interaction between the belief construction method and the evidence propagation scheme that would affect the accuracy and variability should be optimized. We propose an optimization method based on entropy computation, and hence minimizing uncertainties will result in a series of classified knowledge maps within a context of probabilities. For example, it could happen that some initial beliefs (assumptions) based on some evidence propagations are more valid with some constructed belief functions than others, thereby confirming that the sequential image classification (reasoning) must be considered a function of both the evidence propagation scheme and the belief function's construction.

5.2 Belief Function Propagation of Evidence

Bayesian inference is based on the mapping of an implication relation into conditional probabilities. Let $\{z_1, z_2, ... z_N\}$ be N independent radar measurements of x_k such that

$$x \rightarrow \{z_1, z_2, ... z_N\}. \tag{5.1}$$

Given the conditional probability function $f_1(x|z_i)$, updating the knowledge of x given z_i measurement would be based upon $P(x|z_i)$. The difficulty with the scheme stems from the fact that with further estimation of y and with a relation $y \rightarrow z_i$, then there is a need to update the value estimated of x based upon $P(x, y|z_i)$, and so on. As more observations occur, the conditional probabilities become practically impossible to estimate, whether subjectively or from sample data and consist of the inverse problem. To address this difficulty in Bayesian belief network, we consider the fact of dependence between implication relations. In other words, x and y would be dependent and we would not need to obtain the joint conditional probability $P(x, y|z_i)$ but would need only the simple conditional probabilities to compute the new probabilities of $P(y|x)$ and $P(x|y)$ or

$$\{x, y \rightarrow z_i\} \Longrightarrow \{x \rightleftharpoons y\}. \tag{5.2}$$

These dependencies, when they exist, form the context of *belief networks* to the extent of this thesis and the regularization synthesis in the sensor fusion.

Belief networks (Figure 5.1) are directed cyclic or acyclic graphs in which each node represents a knowledge unit (ku) and each link represents an implication relation [84] [43] [15]. Suppose that we are given a certain knowledge map about which a new knowledge state is to be analyzed resulting from an additional sensor measurement. Thus, a complete knowledge estimation of all the knowledge units would indicate the new knowledge map state. It is this estimate and assessment that form the basis of the proposed regularization in its logical form.

Initially, our present approach to *knowledge regularization* relies on an inference network [10]. The inference network, in conventional knowledge-based applications, usually refers to a representation of knowledge structure (Figure 5.1) as defined by equation 5.2, which enables an inference engine to make explicit conclusions given some specific measurement data. For the purpose of knowledge mapping, we use an instantiation of such a representation, that we identify as a knowledge structure, as a basis of performing inferences and thereby handling the regularization of uncertainty factors to each sensor individually.

Bayesian inference is based on the mapping of an implication relation into conditional probability relations which form the skeleton of the knowledge structure. The Dempster-Shafer theory, as introduced by Shafer [81], offers a powerful methodology for revising beliefs about uncertainty in the presence of new information (i.e., accumulated evidence). We suspect optimization in the belief updating process can best characterize the regularization reasoning. Other related formalisms exist [98], such as Pearl's Bayesian networks [72] [73], though we presume that Dempster-Shafer will fit in our application of subsurface imaging, and be the first step in solving an inverse problem through evidential reasoning.

5.2.1 Bayesian and Dempster-Schafer Belief Functions

The Dempster-Shafer theory of evidence accepts partial specifications in the form of logical sentences and allows a probability assignment to a subset of these sentences. Mathematically, it offers a rigorous way of combining beliefs from distinct sources (e.g., confirming and dis-confirming supports) to obtain a set of aggregate beliefs. Of most significance, it distinguishes the state of ignorance about a proposition from the relative weight afforded to the proposition versus its negation. Therefore, as a system for representing and manipulating degrees of uncertainty, we believe that uncertainty optimization in Dempster-Shafer theory would be well-suited to modeling the process of assessing knowledge based on the accumulation of evidence.

Unlike the Dempster-Shafer scheme, the Bayesian belief network treats rules as conditional probabilities. The axioms of probability require that

$$P(K) + P(\neg K) = 1 \tag{5.3}$$

and, hence, may sometimes raise concerns about representing belief measures. For instance, an observation leading to the belief of K does necessary commit the complementary dis-belief about $\neg K$. In general, the amount of truth observation is not bounded and the axioms are to be questioned about handling uncertainty in order to reach conclusive judgements.

There exist various interpretations of the imprecision [1] measures associated with an implication rule [49]. Each interpretation dictates the way in which inferences are to be performed. In our knowledge assertion, we have chosen the Dempster-Shafer model of evidence and Bayesian model of inference, where the deductions take place within logical constraints, and the belief information is treated as an empirically formed meta-constraint as a function of the inverse operator that modifies these logical constraints. In addition, examples of integrating the evidence theory into real

[1]Uncertainty.

world systems can be found in the literature such as inference [37] and multi-sensor integration [30].

5.2.2 Dempster-Schafer Representation

According to the Dempster-Shafer theory, the set of possible outcomes of a unit in the knowledge map is called the *frame of discernment* , denoted by Θ. Let us define the following sets:

p:	number of positive instances
n:	number of negative instances
P:	total number of positive instances
N:	total number of negative instances
m:	basic probability assignment

Table 5.1: Possible measurement distributions.

Among other possible notations [20] [19], the Dempster-Shafer belief functions may be written as

$$m(h) = \begin{cases} \frac{p-n}{P} & if p \geq n \\ 0 & otherwise \end{cases} \tag{5.4}$$

$$m(\neg h) = \begin{cases} \frac{n-p}{N} & if n \geq p \\ 0 & otherwise \end{cases} \tag{5.5}$$

$$m(\Theta) = 1 - m(h) - m(\neg h) \tag{5.6}$$

where $\{h, \neg h\}$ denotes the hypothesis induced from the observations data. In this context, when the confirmation $m(h)$ and dis-confirmation $m(\neg h)$ tend to zero, the frame of discernment mass $m(\Theta)$ tends to unity. Although the proposed model increases the complexity, there have been suggested more complex representations of the Dempster-Shafer theory [96]. In our context, we maintain the entropy computation independent of the proposed model which is generalized as a function of the frame of discernment.

Definitions and Assumptions

As mentioned before, the set of possible outcomes of a node is the frame of discernment Θ. If the antecedents of a rule resulting in an inverse operator confirm a conclusion with degree $m(h)$, where $m(h)$ is above a certain threshold value, the rule's effect on belief in the subsets of Θ can be represented by so-called probability masses. (Note that in Bayesian formalism, probability masses can be assigned only to singleton subsets of Θ). When a source of evidence assigns the probability masses to the conclusion represented by subsets of Θ, the resulting function is called a *basic probability assignment.*

In the Dempster-Shafer model, the probability mass assigned to Θ represents ignorance. If a basic probability assignment assigns $m(h)$ to a singleton corresponding to the conclusion of a rule, for example K, then it assigns $1 - m(h)$ to Θ. If it is a negative implication and the evidence dis-confirms the conclusion with degree $m(\neg h)$, then the basic probability assignment assigns $m(\neg h)$ to the subset corresponding to the negation of the conclusion, $\neg K$, and assigns $1 - m(\neg h)$ to Θ. Unlike the Bayesian approach, in the Dempster-Shafer model, a subset cannot be proved by any rule set unless it appears in a consequent of at least one rule.

Formally, a basic probability assignment is a function:

$$m : 2^{\Theta} \rightarrow [0, 1] \tag{5.7}$$

where

$$\sum_{X \subseteq \Theta} m(X) = 1.0; \qquad m(\emptyset) = 0. \tag{5.8}$$

Accordingly, a belief function, $Bel(X)$, over Θ is defined as the total belief committed in all subsets of X, i.e.,

$$Bel(X) = \sum_{Y \subseteq X} m(Y). \tag{5.9}$$

Combination of Belief Functions

The Dempster-Shafer theory provides a means for combining beliefs from distinct sources, known as *Dempster's rule of combination.* This rule states that two basic probability assignments, corresponding to two independent sources of evidence, may be combined to yield a new basic probability assignment, and that is,

$$m(X) = k \sum_{X_i \cap X_j = X} m_1(X_i) m_2(X_j) \tag{5.10}$$

where k is a normalization factor that ensures equation 5.10 be satisfied,

$$k = \frac{1}{1 - \sum_{X_i \cap X_j = \emptyset} m_1(X_i) m_2(X_j)}. \tag{5.11}$$

As will be shown in Chapter 7, this rule of combination plays an important role in deriving the knowledge maps from the accumulation of evidence and inferencing results.

5.2.3 Belief Function Propagation of Evidence

The general problem of drawing inferences from objectively assessed evidence is one in which there is a renewed interest because of the current work in the field of artificial intelligence. It is natural to attempt to apply Bayesian methods in the analysis of such a problem [22]. These methods have been grounded in the concept of subjective probability [49], for which there exists a solid theoretical foundation. In general, our concern is with an inference network in which there are chains of evidence and hypotheses, several hypotheses supported by the same evidence, and a single piece of evidence supported by several pieces ofevidence. Pearl [73], for example, has developed an updating scheme for inference networks.

We assume that there are only two possible outcomes for each sensor measurement, z_i; namely, "the sensor *does* or *does not* confirm to the parameter hypothesis in question". In the Dempster-Shafer model, this implies that our frame of discernment will be of the form in equation 5.4. The basic probability assignment, corresponding to the frame of discernment, to the propagation of knowledge in the knowledge map can be formulated in the following fashion:

The Algorithm: `(BIND &facts)`

For each single node in the knowledge structure, there can exist from the inverse operator, without loss of generality, m nodes confirming the node u_k and n nodes disconfirming the node u_k. In this case, combining all basic probability assignments for each of the possible outcomes can be thought of as grouping all the rules into two rules, one confirming u_k with a basic probability assignment equal to $m^k(h)$ and the other disconfirming $m^k(\neg h)$.

By the definition of the basic probability assignment, we know that $m^k(h)$ and $m^k(\neg h)$ can be derived by repeatedly applying

$$m^k(h) = 1 - \prod_{1 \le i \le m} (1 - m^i(\neg h)) \tag{5.12}$$

$$m^k(\neg h) = 1 - \prod_{1 \le j \le n} (1 - m^j(h)) \tag{5.13}$$

and also we can compute

$$m^k(\Theta) = 1 - m^k(h) - m^k(\neg h). \tag{5.14}$$

Hence, the belief propagation algorithm can be reformulated as: Each of the knowledge units propagates the belief to its neighboring nodes (as specified by the inverse operator), following equations 5.12, 5.13 and 5.14. In general, if a node is confirmed, it performs backward chaining, otherwise it performs forward chaining and results in the branching of the propagation into a continuous direction change.

In our case, the belief functions, defined as the set of the beliefs committed to every basic probability assignment, can be easily represented as follows,

$$Bel(u) \doteq \{m_t^k(h), m_t^k(\neg h), m_t^k(\Theta)\} \tag{5.15}$$

and here t represent the time index of the knowledge state. From the preceding discussions, we can readily work out a procedure for automatically deriving the stability factor of the knowledge unit u which is governed by the relative ($e.r.$) standard error estimate

$$\boxed{e.r. = |m_t^k(\Theta) - m_{t-1}^k(\Theta)|} \tag{5.16}$$

and the static ($e.s.$) standard error estimate

$$\boxed{e.s. = |m_t^k(\Theta) - m_0^k(\Theta)|.} \tag{5.17}$$

which are computed as a function of Θ.

In what follows, we provide basic search and computing steps for regulating the belief functions.

5.3 Entropy Computation and Optimal Search

In the notion of information theory [82], quantitative concepts were derived with the intention of optimizing the information process. Whether for simple systems, or for systems that have a tendency to grow in complexity and size such as belief network architecture, a standard method of measuring the system is essential.

In the context of this thesis, the knowledge graph consists of the information system model, which is to be somehow measured and which is to acquire a significant and informative measurable index. Since the knowledge structure faces some alterations when faced with modifications of one or more knowledge units, the knowledge structure undergoes a state transition. Here, it is logical to assume that the new state is more certain than its antecedent as the nature of the modifications added to the knowledge units is of an informative nature (cumulative evidence).

It is obvious that if a knowledge structure can be controlled to any state, there is no need to measure the knowledge structure since the final state (i.e. maximum information) of the structure is readily reached. On the other hand, in the case where the knowledge states are not predictable, it is essential to measure the knowledge structure at every update or state transition and acquire an index of uncertainty and, hence, the largest possible transition between a state and its antecedent minimizes the uncertainty and leads to the optimal knowledge state. This procedure is known as the minimum entropy search.

5.3.1 Uncertainty in Bayesian Networks Revisited

Realizing that the knowledge structure can be at any state, which characterizes the notion of a degree of uncertainty, it is evident that the information gathered regarding the knowledge structure can dramatically change the uncertainty index. It becomes a problem of a sequential selection of knowledge units to be governed by the degree of uncertainty. Generally speaking, it is clear that the more information gained, the more the degree of uncertainty regarding the knowledge structure decreases. Hence, in Bayesian space when assigning to each knowledge unit in the knowledge structure a probability space, it becomes obvious that the degree of uncertainty is a function of the probabilities associated with the knowledge units or

$$H = f(P(u_1), P(u_2), ..., P(u_N)) \tag{5.18}$$

where $P(u_n)$ describes the probability of the n^{th} knowledge unit. It is essential to mention that the degree of uncertainty is a function of the number of units in general, and, in particular for the knowledge structure, the amount of units is fixed at N which is also the amount of nodes in a specific knowledge map.

In practice, to measure the uncertainty of a knowledge unit, we will use the concept of entropy which was initially developed in information theory [82]. The entropy of a knowledge unit or node can be valued as

$$H_u = -\sum P_i(u) log(P_i(u)) \tag{5.19}$$

where log is a logarithmic function and $P_i(u)$ describes the probability space of the event u. It is important to mention that the probability space is confined to unity or $\sum P_i = 1$. The index i describes the individual exclusive sets of the assigned space.

In building the knowledge map, as refered to in earlier sections of this document, adapted to Bayesian space, knowledge representation is based on the existence of the $\{confirm, \neg confirm\}$ pair, where each contributes equally to the assessment process. This reflects the aspect of knowledge introduced by the knowledge units with probability values approaching zero which is considered as valuable as knowledge units approaching unity, and the sample space of knowledge is divided into two exclusive spaces ($i = 2$). Hence the entropy can then be evaluated exactly as

$$H_u = -[P(u)log(u) + (1 - P(u))log(1 - P(u))] \tag{5.20}$$

where u is any knowledge unit in the knowledge map.

Since the basic concept of knowledge assessment is to obtain an impression of a certain measure of the degree of uncertainty of a knowledge structure of more than one node (i.e., N nodes), the knowledge structure can be at any state dominated by the probability sets in all knowledge units. Hence, the uncertainty measure may be viewed as a function of the probability of all units in the knowledge map and,

$$H = f(P_{i_1}(u_1).P_{i_2}(u_2)....P_{i_N}(u_N)) \tag{5.21}$$

which can be expanded and arranged to a successive sum of the composition of all the possibilities of all nodes and hence

$$H = -\sum_{i_N} ... \sum_{i_1} P_{i_1}(u_1)...P_{i_N}(u_N)log(P_{i_1}(u_1)...P_{i_N}(u_N)) \tag{5.22}$$

where the index i describes the individual exclusive sets of the assigned probability space and N the number of knowledge units. In particular for our case, the number of states of the knowledge unit is fixed to two. As a result, the total number of combinations is bounded to 2^N. Since the logarithmic function of a product can be expressed as summation, the previous equation may be expanded and rearranged as

$$H = -\sum_{n}^{N}\sum_{i}^{2} P_i(u_n)log(P_i(u_n)). \tag{5.23}$$

which is the sum of the entropy of each individual node [2].

5.3.2 Uncertainty in Dempster-Schafer Networks

Unlike the entropy computation for the Bayesian approach which has many practical implementations in expert systems, entropy computation for the Dempster-Shafer approach has few implementations which have been carried out on the same criteria as its Bayesian counterpart. The controversy relies on the exactness of the entropy computation in Dempster-Shafer and the linear projection of the entropy in the Bayesian case.

In Figure 5.2.a, the entropy computed in Bayesian space can be viewed as a direct cost function of the associated probability. Figure 5.2.b demonstrates the erroneous attempt to construct the entropy for Dempster-Shafer as a linear projection from its Bayesian counterpart. In Figure 5.2.b, values $\{1,1\}$, $\{\frac{1}{2},\frac{1}{2}\}$ and $\{0,0\}$ have the same entropy although each $\{confirm, \neg confirm\}$ set represents different amounts of knowledge. Note that the $\{confirm, \neg confirm\}$ pair divides completely and exclusively the probability space. Although a formal investigation of the linear projection of entropy can lead to an existing relation between the $\{confirm, \neg confirm\}$ pair which defies the definition of Dempster-Schafer induction rules, we will concentrate our effort on directly deriving the entropy cost function for Dempster-Schafer.

[2] Hint for proof: expand the log to a series of sums and regroup by units and finally collapse the probabilities to unity.

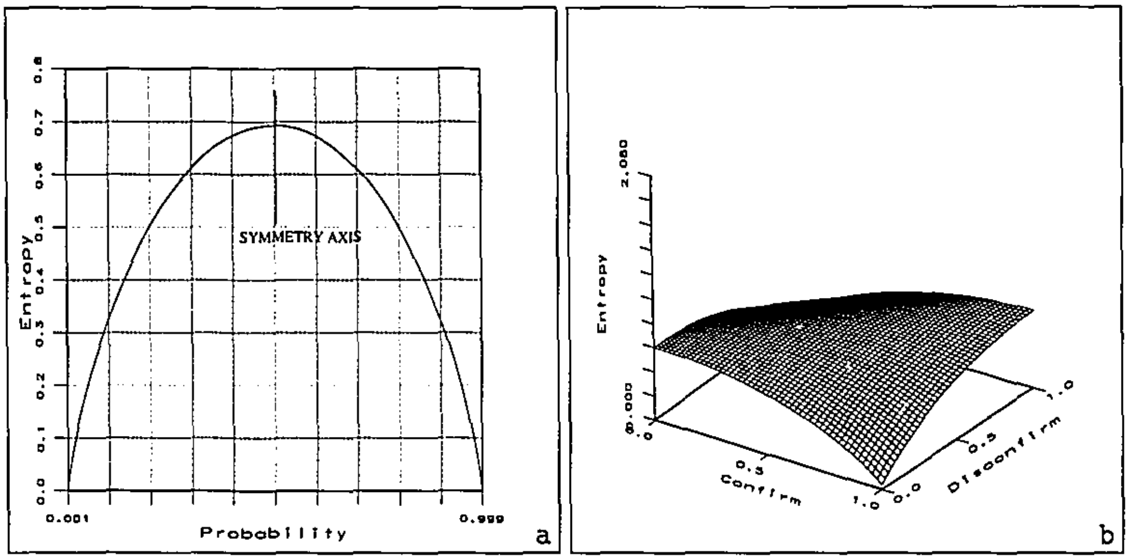


Figure 5.2: Entropy computed for Bayesian belief networks. Notice the linearly projected cost function for Dempster-Shafer at values: {1,1}, $\{\frac{1}{2}, \frac{1}{2}\}$ and {0,0}.

Entropy Computation in Dempster-Schafer

As mentioned before in Section 5.2.2 on Dempster-Schafer theory, there is a fixed set of mutually exclusive and exhaustive elements of the *environment* which is formally symbolized by

$$\Theta = \{\theta_1, \theta_2, ..., \theta_M\}. \tag{5.24}$$

When an environment's elements are interpreted as possible answers, the environment is called a frame of discernment, and only one answer is correct in the frame of discernment [32]. The term discern means that it is possible to differentiate correct knowledge states from all the other possible knowledge states to a specific node. If the knowledge state is not in the frame, then the frame is expanded to accommodate the additional knowledge elements Θ_{M+1}, Θ_{M+2} and so forth. One correct state requires the set be exhaustive and that the subset be disjoint [19] [81]. The power set of the frame of discernment has as its elements all knowledge states of the environment and

$$\mathcal{P}(\Theta) = \{\emptyset, \{\theta_1\}, \{\theta_2\}, \{\theta_1, \theta_2\}, ..., \{\theta_1, ..., \theta_M\}\}. \tag{5.25}$$

The entropy can then be defined as a function of the power set of the frame of discernment

$$H_u(\Theta) = f(\mathcal{P}(\Theta)). \tag{5.26}$$

Defining the joint distributions,

$$\mathcal{Q}(\Theta) = \{\emptyset, \{\theta_1\}, \{\theta_2\}, \{\theta_1\theta_2\}, ...\}; \tag{5.27}$$

hence entropy becomes the sum of the individual entropy of each element in the frame of discernment. The partial entropy of the k^{th} element in the frame can be written using Equation 5.19 as

$$\boxed{H_u(\Theta) = \sum_{k=1}^{2^M} H_u(\mathcal{Q}_k(\theta)).} \tag{5.28}$$

Using the same derivation devised for the Bayesian approach, the entropy for the knowledge structure can be written as

$$\boxed{H = \sum_{n}^{N} \sum_{k=1}^{2^M} H_{u_n}(\mathcal{Q}_k(\theta))} \tag{5.29}$$

where N is the total number of knowledge units or nodes.

In Figure 5.3.a, the overlap plots of the linearly projected Bayesian entropy cost function and the exact entropy cost function for Dempster-Shafer are shown. Figure 5.3.b is the proposed entropy cost function for Dempster-Shafer. Note that entropy increases for values {0,0} and {1,1} and reaches a maximum entropy at $\{\frac{1}{2}, \frac{1}{2}\}$.

5.4 Regularization and Minimum Entropy

Since one of the intentions is knowledge assessment, minimization of the uncertainty becomes essential to satisfying the goal. The concept of minimum entropy computation as the measure of degree of uncertainty provides optimal perspective to

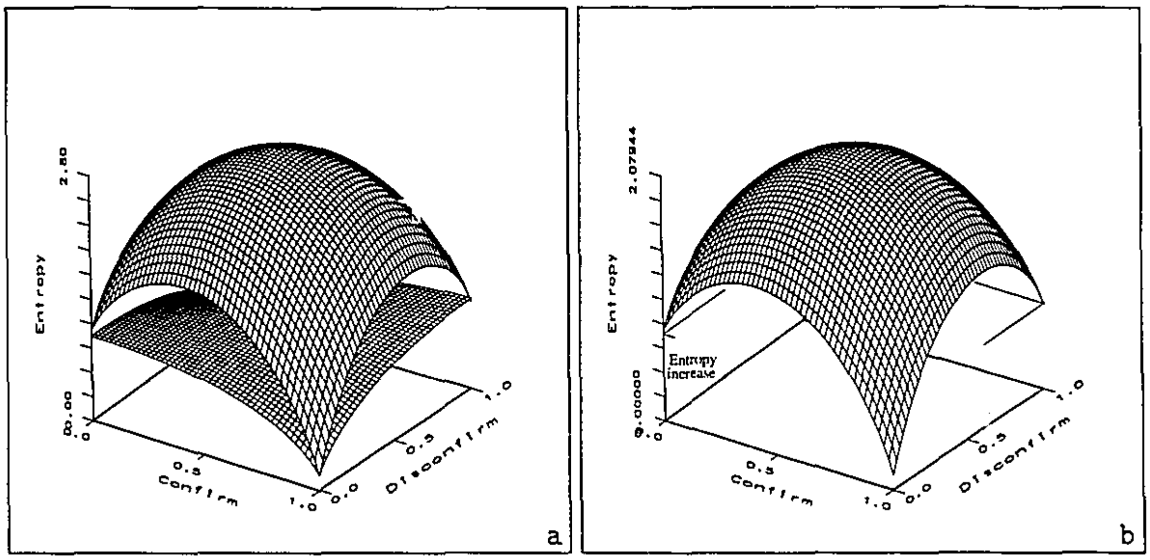


Figure 5.3: Entropy computed for Dempster-Shafer environment of two elements. Notice the cost function at values: {1,1}, {0,0}, {1,0} and {0,1}.

measuring the knowledge structure although that is only valid when the interest of the knowledge assessment is fair and is particular to the $\{confirm, \neg confirm\}$ set.

To understand the concept of the regularization of entropy for the knowledge maps, it is imperative to realize that the goal of entropy is to locate the maximum uncertainty in the network and require the measurements of the knowledge unit that minimizes the uncertainty. However, when a large amount of the knowledge units are distributed around a certain mean different from $\frac{1}{2}$, the entropy of the total system increases around the mean and results in a momentum bias, whereas uncertain knowledge units located symmetrically away from the mean are not well represented.

A practical way to view the impact of the bias is to consider a specific knowledge map with a probability assignment to each knowledge unit. Since the entropy of a knowledge structure is the sum of the entropy of the individual knowledge units, it becomes clear that the mean of the probabilities is reflected in the knowledge structure entropy computation and hence forms a momentum towards an extreme in the devised set $\{knowledge, \neg knowledge\}$. A control of the mean would result in a moderate control in the minimum search and hence regularize the entropy search by the cumulative momentum.

5.4.1 Sensitivity and Regularization

In practice, the entropy of complex systems of multiple units can exhibit eccentric behavior due to the effect of any possible cumulative bias in the structure. Such behavior can occur in dynamic systems when the minimization of entropy is used. This can be viewed as the belief values being driven numerically towards zero or one and hence any initial bias will force the minimization to follow a path constrained by the momentum.

The computation of the momentum of the entropy in a knowledge map is some function of the mean μ which is the average sum of the assigned probabilities $P(u_n)$ of the knowledge units. The mean can be written as

$$\mu = \frac{1}{N}\sum_{n}^{N} P(u_n) \tag{5.30}$$

and hence we can warp the knowledge map probabilities to the space normalized by the mean which can then be expressed by the posterior probability

$$P'(u_n) = P(P(u_n)|1-\mu) \tag{5.31}$$

and the Bayesian posterior probability is expressed as

$$P' = \frac{(1-\mu)\ P}{(1-\mu)P + \mu(1-P)}. \tag{5.32}$$

The transformation is one to one, mapping every unit to the new probability space. Such a warp will ensure that the entropy cost function is balanced and the bias momentum is reduced. The direct impact of warping the data around the mean will result in as much interest in resolving uncertain data as in resolving more certain data. The reduction in the bias momentum is not sufficient by itself to guarantee a good estimate. In fact, there is no advantage to performing statistical warping besides that of escaping the unpredictable bias. However, achieving a control on the bias is highly significant to the inverse problem. Let λ be the bias momentum we would like to achieve and hence the desired mean μ', which is also the second order warp, is computed as

$$P''_n = P(P'_n|\lambda); \qquad \lambda \in [0,1]. \tag{5.33}$$

λ is the proposed *regularization factor* (rf) in this thesis to accomplish the second order warp around the mean. As a result the entropy, for example in Bayesian space, can then be expressed as

$$H = -\sum_{n}^{N}\sum_{i}^{2} P''_i(u_n) \tag{5.34}$$

where such mapping introduces an entropy decrease for the units neighboring the bias μ and an entropy increase for the units neighboring λ.

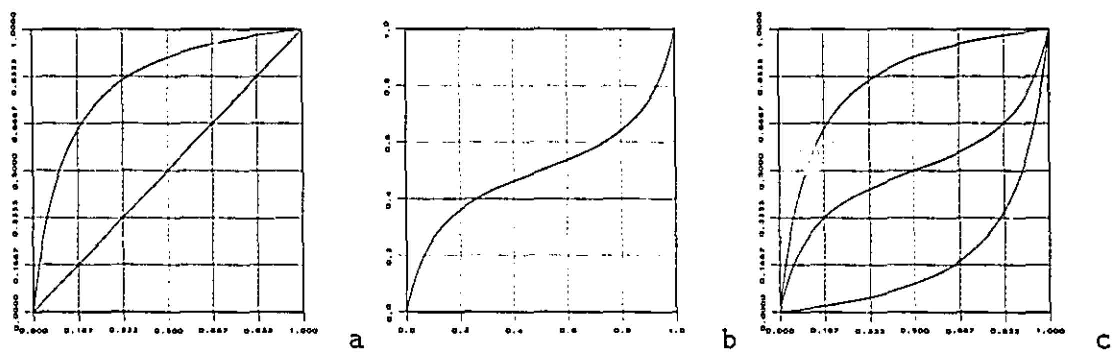


Figure 5.4: Left: A warp of $\lambda = 0.2$ compared with warp of $\lambda = 0.5$. Middle and Right: A combined warp of $\mu = 0.8$ and $\lambda = 0.8$.

Similarly, the approach can be derived for the Dempster-Shafer belief functions. The mean of the mass beliefs is computed separately as

$$\mu_h = \frac{1}{N}\sum_{n}^{N} m_n(h); \qquad \mu_{\neg h} = \frac{1}{N}\sum_{n}^{N} m_n(\neg h). \tag{5.35}$$

Mapping the mass m to a new space m' normalized by its mean is expressed again by the posterior probability function or

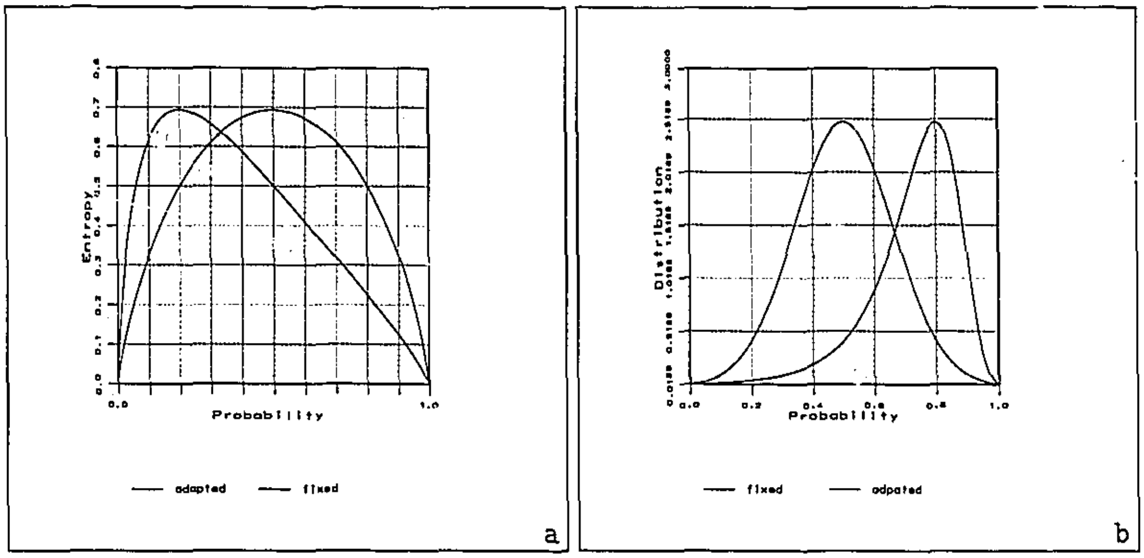


Figure 5.5: Entropy functions as a notion of uncertainty measure: left to right, a comparison of the resulting entropy functions for a warp for $\mu = 0.8$ and $\mu = 0.5$, given two distributions with mean 0.5 and 0.8.

$$m''(h) = P(m(h)|1 - \mu_h, \lambda) \tag{5.36}$$

$$m''(\neg h) = P(m(\neg h)|1 - \mu_{\neg h}, \lambda) \tag{5.37}$$

$$m''(\Theta) = 1 - m''(h) - m''(\neg h). \tag{5.38}$$

Typical warp is shown in Figure 5.4. In Figure 5.5, two probabilistic distributions are shown. The first fixed plot is normalized with mean $\frac{1}{2}$ whereas the second is skewed to the right and has a mean of 0.8. The entropy computation for both plots is shown in Figure 5.5 and the adapted plot demonstrates the adjustments introduced.

Finally, before concluding this chapter, the notion of regularization in adapting the sensitivity can be viewed as the concept of having maximum conceivable response of the system to the smallest knowledge change. That is, in the context of the biased knowledge structures, the entropy around the bias decreases when mapped. The entropy become more sensitive to values neighboring the bias centered at λ.

Chapter 6

Synthesis: Multi–Sensor Fusion

6.1 Introduction

All deterministic phenomena are inevitably followed by stochastic variations. However, in practical problems the theory is not perfect and there exist random variations. The methodologies of *perfect science* often fail to justify certain behaviors which could be attributed to lack of knowledge of certain parameters, or simply randomness.

From a completely theoretical point of view, the parameters, which have been labelled as random, do not differ in principle from any other parameters in question. Theoretically, the resolution of the problem can grow difficult when additional parameters are taken into consideration [71]. However, the realization of such an approach is practically unfeasible, or will result in complicated solutions which are also not practical [74].

It is evident that there must exist a difference in principle among the methodologies that permit us to take into consideration the essential factors governing world modeling, and also secondary factors that manifest through errors or more simple random variations. As the essential factors of the inverse problem were developed in Chapters 4 and 5, this chapter deals with the secondary factors of the inverse problem, namely, interpreting the uncertainty.

6.2 Uncertainty and Probability Distributions

To compute the probabilities of a random variable, it is not always indispensable to determine the frequency directly from empirical data. Empirical analysis in remote sensing is quasi-impossible while the system is under investigation, which leads to a contradiction in the *a priori a posteriori* definition. The theory of probability disposes of numerous methods that permit the indirect identification of probabilities in function of other events or measurements somehow related to the investigated problem. It is these methods on which we try to base this thesis. Chapter 5 introduces an initial approach in multi-sensor fusion to infer the knowledge through direct and indirect approaches. Nevertheless, it is fundamental to define the random behavior in truth as stable in the sense of a density function, which is validated in calibration procedures and laboratory experiments. In most statistical distributions, chance intervention is more or less high, and that is due to the fact that the number of the selected samples is limited. To resolve this issue while processing the sample space, it is necessary to choose a theoretical curve to approximate a statistical distribution and to express the essential elements of the samples without expressing the circumstances.

The problem consists of finding a theoretical curve which in a way provides a satisfactory description of the sample space. The search for the best approximation of the distribution of a sample space, which is similar to the problem of the best analytical representation of an empirical function, is a problem which is vaguely defined. However, the solution depends explicitly on what "best" means. The range of functions which provide the approximation is chosen, depending on the nature of the physical problem to be solved rather than on mathematical perspectives, and hence depends on the characteristics of the empirical curve [80].

To represent the statistical observation in a compact and ordered form, the frequencies are grouped and sorted and finally approximated by an analytical density function. In the most common sensor technology, the normal distribution $n(x;\mu,\sigma)$, also known as the Gaussian distribution, can accommodate the situation and is formally written as

$$n(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-1/2[(x-\mu)/\sigma]^2}, \tag{6.1}$$

which has, in fact, proven to be efficient in most histogram approximations. In general, the normal curve is directly dependent on the mean μ and the variance σ of the distribution under investigation. The variance of a random variable is characterized by the dispersion of its value in the neighborhood of its mathematical mean. Although we operate in a single dimension, we should note that in the application of the inverse problem, the inverse estimates are governed not by a single random variable, but rather by three random variables in the x, y, z Euclidean space.

6.2.1 Finite Semi-Finite Spaces and Differential Entropy

As we envisaged in Chapter 5, entropy computation as an index for uncertainty in the belief functions, continuous distributions also present entropy as an information aspect. Single dimension continuous probability functions are characterized by the random variable x and probability density function $f(x)$ and a zero mean in which case the normal distribution can be represented by $f(x) = n(x;\sigma)$. Above all, continuous functions are ideal in their form and in fact they approximate empirically constructed distributions. For the mathematical derivation simplicity, we will consider a precision increment Δx for the normal distribution $n(x;\sigma)$ such that the variation of x in Δx has an insignificant outcome. The equivalence of this description is to evaluate the continuous functions $n(x;\sigma)$ by a discrete histogram such that $\Delta x f(x)$ is the probability of the event occuring in the segment Δx.

If we consider the segment Δx small enough to justify the insignificant variations of x in Δx, the approximation of the entropy for all the Δx is determined by the standard entropy expression or

$$H_{\Delta x}(x) = -\sum_i f(x_i)\Delta x \; log[f(x_i)\Delta x] \tag{6.2}$$

which is also equal to

$$H_{\Delta x}(x) = -\sum_i f(x_i)log[f(x_i)]\Delta x - \sum_i log[\Delta x]f(x_i)\Delta x. \tag{6.3}$$

Letting Δx be very small, the following approximation becomes valid and

$$H_{\Delta x}(x) = -\int_{-\infty}^{\infty} log[f(x)]f(x)dx - \int_{-\infty}^{\infty} log[\Delta x]f(x)dx \tag{6.4}$$

and hence equation 6.4 becomes

$$H_{\Delta x}(x) = -\int_{-\infty}^{\infty} log[f(x)\Delta x]f(x)dx. \tag{6.5}$$

The first term in expression 6.4 does not depend on Δx, which is the precision in evaluating the uncertainty. It is rather the second term $(-log[\Delta x])$ which depends on the precision Δx which tends towards infinity for $\Delta x \rightarrow 0$, which is perfectly logical, as the more precision there is, the more uncertain the random variable is. However, giving Δx the required length as direct impact of the sensitivity of the sensor measurements, the entropy can be approximated by equation 6.5 and may be rewritten in an expectation form

$$H_{\Delta x}(x) = cov(-log[f(x)\Delta x]). \tag{6.6}$$

For the zero mean normal function $n(x;\sigma)$, equation 6.6, the entropy is then derived as

$$H_{\Delta x}(x) = E\left\{-log\left[\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}\Delta x\right]\right\} \tag{6.7}$$

which reduces to

$$H_{\Delta x}(x) = log\left[\frac{\sigma\sqrt{2\pi e}}{\Delta x}\right]. \tag{6.8}$$

The complete entropy was derived in Section 5.3 and we can rewrite equation 5.29 as the sum of three components or

$$H = I_h + I_{\neg h} + U_\Theta \tag{6.9}$$

where I_i is the information index and U the uncertainty index as a function of Θ and, for the Dempster-Schafer model as in 5.4, the U_Θ component is expressed as

$$U_\Theta = -(1 - m(\Theta))log(1 - m(\Theta)), \tag{6.10}$$

hence equating U_Θ to $H_{\Delta x}$ and finally we get

$$\sigma^2 = \left(\frac{e^{1-m(\Theta)}}{1 - m(\Theta)}\right)^2 \left(\frac{\Delta^2 x}{2\pi e}\right), \tag{6.11}$$

which can be generalized for any uncertainty function U_p and

$$\boxed{\sigma^2 = \left(\frac{e^{U_p}}{U_p}\right)^2 \left(\frac{\Delta^2 x}{2\pi e}\right); \qquad U_p \in [0,1], \tag{6.12}}$$

where Δx is the physical resolution limit.

Having derived the *finite* to *semi-finite* mapping from an uncertainty concept to a normal distribution variance, one should keep in mind that equation 6.12 is only valid if the density function is a zero mean normal (Gaussian) distribution.

6.3 Sensor Fusion and Kalman Filtering

Interest now arises from the need to combine the knowledge maps into an optimal estimate. The best estimate of parameters is that which maximizes the expected consequences. In the statistical literature, the effects of variations in the distribution is called *model robustness* [80]. In this work, we first consider the possible advantage to representing knowledge by approximate distribution and associated uncertainty. Secondly, we realize that the type of error modelled is properly described as a function of a stochastic model. This variation may arise from two possible analyses: variations among the knowledge maps themselves, or discrepancies between the estimated knowledge and the real model, and that is of course under simulation conditions. This distinction is important – the former is an error which must be minimized by the final stage of the sensor fusion while the latter, if available, may be an important source of information about the suitability of the model. However, nothing much can be done about this.

Since error free point estimates among knowledge maps are generally not possible, an important element in the determination of a strategy for merging the knowledge maps is the effect of the estimation error. Merging data has been the focus of much research, especially in robotics applications where the *sensor fusion* became common in the literature [23] [39] [25]. The essential problem in fusion is the conservation of the patterns that exist in either knowledge map. On the other hand, merging must preserve the knowledge without introducing spurious elements. A simple method can be that of simply averaging the probabilities of the point estimates; however, we will keep a linear aspect in the fusion although some optimality will be required.

6.3.1 Theory of Kalman Filtering

The application of linear filters in sensor fusion provides simple processing operations and that is something we should investigate before any attempt at nonlinear optimization. Fortunately, the known Kalman filter clearly outlines the proposed requirements with minimum assumptions. The theory of Kalman filtering encompasses a wide range of classical mathematical topics, especially when dealing with random process theory and estimation theory. Kalman filtering can be considered as a mature engineering discipline at this point in time. Since the original concept was first published by R. Kalman back in 1960, literally thousands of technical papers have been written about Kalman filtering. Unfortunately, Kalman filtering will only be used in its simple form. The mathematical theory behind the concept of Kalman filtering involves probability and statistics together with linear systems theory in a state vector formulation. In this thesis, the theory will be outlined, however, in its decentralized form. Many references are available for more complete analysis of the theory [5] [52].

When presented in its simple form [6], the Kalman filtering process consists of combining two estimates of a random variable to form a weighted average. The optimal weighting factor is chosen so as to produce a weighted average having the *minimum variance* . In order to apply Kalman filtering technique to any particular physical problem involving a dynamic process. the equations of motion of the process must be expressed in a state variable formulation, with any random noise included as well – the underlying statistical theory will be that of the Markov process.

In principle, the Kalman filter is a non-discriminating classifier. which, in a sense, attempts to combine a statistical component of estimates rather than selecting a best estimate. The main stage in the operation in the image synthesis or reconstruction, that is, the Kalman filter will search among the knowledge maps for the most consistent sensors and generate an output estimate based on the optimized weighted estimate. Each estimated node deals with the corresponding point estimate in the knowledge map; given that the number of knowledge maps is larger than one, the failure of any node will not result in failure in the final estimate of the node. In fact, the Kalman filter advantage is that it is able to degrade gracefully in the face of merging failure.

In order to make the design of a Kalman filter at all feasible, there is the major requirement that all noise sources be Gaussian. As demonstrated in Section 6.2.1, point estimates are normal, by definition, with a stable variance (σ^2) evaluated as a function of uncertainty (equation 6.12). The statistical mean (μ) and variance (σ^2) completely define the normal distribution. Thus the former factor, or the mean, has not been evaluated. However, the mean can be computed as

$$\mu = m(h) + m(\neg h), \tag{6.13}$$

which is a straightforward evaluation given the model as described in equation 5.4.

6.3.2 The Kalman Filter and the Kalman Gain

Formally, let x be a point located in the physical system under investigation and let $\{x_1, x_2, ...x_N\}$ be N independent point estimates of x. Given the conditional probability functions $f_n(x|x_n)$ for all $n = 1, 2, ..N$ expressed by their means $\mu_n = E\{x_n\}$ and error variances σ_n^2, the N independent estimates set can then be combined to generate an optimal estimate of the physical model x. The general weighted average of the estimate x_n is denoted by $\bar{x}$, as the expression of the optimal estimate and may be written as

$$\bar{x} = \sum_i^N K_i x_i; \qquad \sum_i^N K_i = 1 \tag{6.14}$$

and the expected value is

$$E\{\bar{x}\} = E\left\{\sum_i^N K_i x_i\right\} = \sum_i^N K_i E\{x_i\}. \tag{6.15}$$

The variance of $\bar{x}$ becomes

$$\sigma_{\bar{x}}^2 = E\left\{(\bar{x} - E\{\bar{x}\})^2\right\}, \tag{6.16}$$

which reduces to

$$\sigma_{\bar{x}}^2 = \sum_i^N K_i^2 \sigma_{x_i}^2, \tag{6.17}$$

as the x_i estimates are initially independent and cross covariance is zero. The goal turns into a minimization of the variance with respect to K_n coefficients. The partial derivation with respect to K_n may then written as

$$\frac{\partial \sigma_{\bar{x}}^2}{\partial K_n} = \frac{\partial}{\partial K_n}\{\sum_i^N K_i^2 \sigma_{x_i}^2\}, \tag{6.18}$$

yielding the Kalman gain set or weight estimates $\hat{K}_n$ for all $n = 1, 2, ..N$ and

$$K_n = \frac{1}{N-1}\left(\sum_i^N \sigma_{x_i}^2 - \sigma_n^2\right)\left(\sum_i^N \sigma_{x_i}^2\right)^{-1}. \tag{6.19}$$

Part II

Ground Penetrating Radar

Chapter 7

Application to Radio Detection And Ranging

The concept of using radar to penetrate the ground has been in design for three decades. Historically, the initial ground penetrating radar work originated from the fact that altimeters on aircrafts used in the Arctic would penetrate through ice sheets. This discovery led to the exploitation of radar in other materials and in time led to great success with the advances of computer technology and visualization methods.

7.1 Ground Penetrating Radar

Actually, ground-penetrating radar can be used with success in almost all environments. In general, the higher the ground resistivity, the better the chances of utilizing ground-penetrating radar. As a result, hardware adaptations to radar systems were added to compensate for the large variety of materials. Higher resistivity has excellent dielectrics through which radio waves easily propagate. Early work has led to the assumption that frozen material should be transparent to radar signals. In fact, field measurements at later stages have verified this suggested hypothesis.

Similar to conventional radar systems, which are used for ranging the distance to specific targets, an antenna is used in ground penetrating radar to generate an electromagnetic radio frequency pulse. The differences between ground penetrating radar and the conventional radar system are governed by the major aspects of prone to wave velocity changes, excessive attenuations. On the other hand, the major resemblance is related to its historical discovery, the reflectivity characteristic.

The basic concept of ground penetrating radar is very simple. A signal is radiated from the antenna and part of the energy propagates with the corresponding electromagnetic wave velocity in the medium. When the signal reaches a critical point of a medium change, part of the energy is reflected. This return radiation forms the basis of ground penetrating radar.

A ground penetrating radar system has two initial requirements. The first is to obtain hardware which can generate and receive the appropriate electromagnetic signals. The second is to make this hardware portable. Once these two requirements are resolved, profiling can then be achieved. Since the wave lengths and scales of the measurements are quite often small, of the order of MHz and centimeters respectively, it becomes essential to have very close spatial sampling. As a result, it turns out to be most practical to make very narrow scans of the medium (μ seconds). The method of performing narrow scans is known in practice as continuous profiling.

7.1.1 Main Functionality Aspects

The hardware involved in ground penetrating radar is generic and simple. Typical units consist of three essential elements, namely, the transmitting unit, the receiving unit and the recording and/or display units. The transmitter unit is a pulse generator which outputs a polarized short duration (1-20ns) voltage pulse onto a broadband antenna. The receiving unit consists of a receiving antenna which acquires reflected signals in a time frame window. It is common to use the same antenna for transmitting and receiving.

In order to put the timing problem into perspective, it is necessary to consider the propagation velocities of electromagnetic waves in typical materials. For example, electromagnetic energy propagates at an average speed of 0.3m/ns. The slowest medium encountered is water, which has a speed of propagation of about 0.03m/ns. It is thus imperative that timing mechanisms be available to control the time frame window and hence provide a spatial resolution.

In ground penetrating radar, the typical repetition rate of the transmitted pulse is in the range of 50KHz to 100KHz. Another timing ramp controls the receiving signal which is arbitrary to the user. This timing ramp provides a relation to the desired depth of penetration. Finally, the major timing issue resides in the pulse itself. Depending on the hardware and designs, pulse width can be in the range of 0.5ns to 1000ns.

One of the biggest challenges with ground penetrating radar to date has been presentation of the data in an effective manner. As mentioned earlier, a main component of the radar hardware is its portability which results in a major limitation of the involved technology. As computational costs are reduced to a minimum, only analog basic features are used, mainly, signal filtering and dynamic gain enhancements, to compensate for any signal attenuation.

Invariably, ground penetrating radar records have a significant amount of noise associated with them. Part of this noise is just system noise. The second source of noise can be external, spurious radio frequency interference which is identified as a general harsh background noise on the record. Various type of band pass filtering, zero suppression, biasing of the data and other factors can be utilized to enhance the data presentation and interpretation at a minimum cost.

7.1.2 Operation: A Heuristic Approach

Practically speaking, a ground penetrating radar system can be considered a very simple system. In fact this property has largely contributed to its success in applications in geophysics, mining and other fields. Users have learned to interpret the measured data on site by means of graphics displays where the recorded signal intensity is a function of pre-allocated color map tables.

Others factors which enter into the data display are those that enhance the color map allocation tables such that the eye can discern individual features in the record. Users usually develop their own method of calibration and include color map definitions. There are a large number of variables in a radar survey operation and these various settings of hardware components of the system as well as the survey procedure can be tailored to the particular application under investigation.

The output of a radar survey is a set of continuous sections which show radar reflections versus delay time on one axis and horizontal position on the other axis. The objective of the exercise is to re-map the radar reflections into their true spatial positions under the surface. This involves two aspects of analysing the data. One is just utilizing this record as delay time image. In this case, an event has a known delay time associated with it and a certain spatial position. The delay time is converted to a distance with a pocket calculator by estimating the velocity of the material through which the wave has propagated. Some radar hardware allows distance mapping directly on the display. In practice, the depth in the ground is only an approximate value since the velocity of propagation is usually unknown. Experience with radar data suggests that accurate predictions of propagation velocity can be made with very little effort.

A great deal of additional information is present in the radar record, but requires considerable effort to extract. With high fidelity recording systems, it is sometimes possible to estimate the polarity of the reflection signal, and to discern variations in the character of the recorded signals which indicate variations in the actual geological target. In practice, the understanding of these concepts is limited to those few technical experts in the field who are familiar with the design of the system.

In general, the experienced interpreter, however, can readily identify features in the records which have a unique character and utilize these unique characteristics in order to infer the proper ideas about the investigated site.

7.2 The GSSI SIR 10 System

The SIR 10 radar system has four channels, allowing simultaneous operation of transducers with different center frequencies, gains, ranges and filter settings. Each channel leads to a 16-bit data segment quantization. The system is a based on an 80286 microprocessor and designed to function as an on-board computer and perform automatic control on the radar system. The main components of the system are as follows:

- **Control Functions** Range and gain values, signal position, pulse repetition rate and all alpha-numeric information relevant to the operation of the system.
- **Oscilloscope** Vertical display of the current scan wave form. The wave form is controlled by the time based window.
- **Data Storage** Up to 2.3 gigabyte on 8mm tape drive.
- **Line Scan** Each scan is represented as a column, one pixel wide. The scans begin on the right side of the display and scroll left as new scans are displayed. 16 grey scales or colors represent the amplitude and polarity of the signal.
- **Range Gain** Range adjustment from -26dB to 120dB. Gain curve is represented in its logarithmic form.

Additional electrical and electromagnetic specifications can be found in Table 7.1.

Resolution	60 picoseconds
Range	0 to 20,000 nanosecond
Pulse Repetition Rate	3.2 to 256 kHz
Analog Quantization	8 / 16 bits
A/D Sampling	128, 256, 512, 1024 samples/scan
Clock Synchronization	Crystal Based: 1 nanoseconds
Scan Rates	3.2 to 256
Beam width Maximum	90 degrees aperture
Beam width Minimum	60 degrees aperture
Radiation Power	0.06 to 100 mW

Table 7.1: Basic electrical and electromagnetic specifications of the GSSI SIR 10 radar unit. Electromagnetic specifications are empirical.

7.2.1 Digital Filtering

The SIR 10 radar unit is equipped with digital filters enabling variable filter frequency selections and filter lengths. The two types of filters are the common Infinite Impulse Response Filters (IIR) and the Finite Impulse Response Filters (FIR).

Infinite Impulse Response Filters (IIR) operate by combining new data along with a history of the past data in some average form. The weights applied to this combination determine the bandwidth of the filter. They are popular because they correspond to the analog filters in the real world. IIR filters are recursive and they use past values of the input to attenuate undesirable frequencies.

Finite Impulse Response Filters (FIR) operate by convolving (i.e applying a sliding weighted average) a finite length function with the data. Each data value is multiplied by the corresponding filter value and added together. The advantage of FIR filters is that they can be made symmetric and centered. This means that the output corresponds in time and space to the input, unlike the IIR filters which will be skewed to one side. The FIR filters exhibit a non-recursive behavior which depends only on the current sample.

7.2.2 A Design Aspect in the SIR 10 Radar Unit

This section introduces some theoretical aspects of the SIR 10 radar unit which are related to the transmitted signal. Whether because of practical feasibility or because of a theoretical advantage, the transmitted signal plays a major role in the overall radar behavior. The transmitted signal is known as a pulse and its shape varies with different hardware.

Without getting involved in the electronics design, the SIR 10 radar unit initially generates a sequence of trigger pulses with a pulse width of 1 nanosecond. Each pulse is polarized and transmitted through a dipole antenna. Hence, the real signal is an electrically polarized electromagnetic pulse of a finite duration. In the frequency domain, the spectroanalysis of the transmitted signal is never measured; however, hypothetical spectra can be estimated [12].

Butler [12] estimated the spectrum of the 120MHz antenna with -3 dB attenuation at 120 $\pm$ 80 MHz. Although we can predict the spectrum at 1 GHz, we can also note the possible spectrum width. In general, this wide spectra can result in complex behavior in ground penetrating radar. Actually, similar problems can be compared to the use of the continuous wave (CW) in a frequency sweep model (FMCW) in a wide band impulse response.

7.3 Reflection, Refraction and Scattering

The historical success of lighthouses or any light beacons provides the perfect example for introducing this section. Light emerging from lighthouses has a particular behavior on dense foggy nights, and that particularity is sought in all light beacons which are tracing a light beam. Thus, considering a perfectly semi-transparent medium illuminated by a light source, and provided that the medium is homogeneous, it may become clear that each point in the medium behaves locally and can be visually identified. Light scattering in a semi-transparent medium presents a perfect example by which to introduce energy scattering. Whether related to eye glasses or mirrors, the implications of the concepts of reflection and refraction have major importance tohuman society.

Very little research has dealt with the scattering analysis directly, and this research mainly consists of the polarization matching of the received signal for optimal signal reception and hence identification of amplitude changes [61]. On the other hand, research has indirectly dealt with the scattering influence and its removal with high technology IIR filtering as described in Section 7.2.1.

In reality, scattering can provide tremendous knowledge, and, when compared, exceed the information acquired from the reflection analysis alone. Again, bringing the issue to a visual perspective as in the lighthouse example, the scattering light provides local knowledge within the light beam width similar to identification of the fog from a density perspective. On the other hand, analysing any reflected light can only provide an average for the fog density.

It is thus an advantage to analyse the scattering electromagnetic wave field in addition to the measured reflection. In fact the measurements consist of the convolution of both scattered and reflected signals. When separated, the reflection analysis provides the data for what we call the Impulse Reflection Knowledge Map (IRKM) whereas the scattering analysis provides the data for the Polarized Scattering Map (PSKM).

7.3.1 The Impulse Reflection Map (IR)

When an electromagnetic field is incident upon a boundary, in general it will split up into reflected and refracted fields. For a wave striking a separation interface of materials, Snell's laws [48] [45] state that

1. The angle of incidence is equal to the angle of reflection, $\theta_1 = \theta_2$.

2. $\nu_1 sin\theta_1 = \nu_2 sin\theta_2$

   where ν_1 and ν_2 are the refractive indices of the two media.

3. For E-polarisation (the case of GSSI SIR 10 radar unit), the reflection coefficient is

$$R_E = \frac{\nu_1 cos\theta_1 - \sqrt{\nu_2^2 - \nu_1^2 sin^2\theta_1}}{\nu_1 cos\theta_1 + \sqrt{\nu_2^2 - \nu_1^2 sin^2\theta_1}} \tag{7.1}$$

   and the refraction coefficient is

$$T_E = \frac{2\nu_1 cos}{\nu_1 cos\theta_1 + \sqrt{\nu_2^2 - \nu_1^2 sin^2\theta_1}} \tag{7.2}$$

4. If $\nu_2 < \nu_1$ and $\nu_1 sin\theta_1 > \nu_2$, total reflection occurs and there is no energy entering into the second medium. In fact the total reflection does not occur completely from the interface, but rather from an imaginary location inside the second medium. This phenomena is known as Göss-Hanchen phenomena and will not be taken into account as it is beyond the scope of this thesis.

In general, the reflected and refracted fields are not only dependent on the reflection and refraction coefficients of the media in question. However, we believe that further investigation in electromagnetic scattering becomes irrelevant to the scope of this research, as many approximations and imperfections have been initially introduced (i.e. signal transmission).

7.3.2 The Polarized Scattering Map (PS)

The problem of computing the wave field propagation can be divided into a double estimation of geometrical properties as well as of electrical preperties. Although most research focuses solely on the geometrical aspects as in Osborn [70], some research has computed the wave field in both geometrical and electrical properties [69] [17].

In general, electromagnetic imaging is susceptible to properties of the media. As in Yu [97], we make the assumption that the subsurface is divided into multiple layers of different resistivity separated by discontinuity surfaces. Let us consider a subsurface that is divided into two parts of different resistivity with a single discontinuity layer. When a radar pulse travels from one medium to another, the pulse will undergo an amplitude modulation (attenuation) along its path. In addition, the radar signal will undergo a velocity change when changing media, and hence the modulation will acquire a phase shift at the discontinuity as the initial radar signal is of minimal phase [92]. Normally the phase shift is associated with a reflection at the interface.

From a signal processing point of view, it has been shown that the signal received at the radar antenna is of a complex exponential form [40] [13] [63]. Complex exponentials play a major role in the analysis of signal processing. Most cases of wave field propagation signal processing involve sets of harmonically-related signals as described in the previous section. In general, a surface wave is generated such that the energy enters the medium along one edge, travels along the interface and then leaves at the other edge; thus the E-field components vanish along the penetrating direction and,

$$E = Ae^{jkt+j\omega t} = Ae^{-j\beta t}e^{j(\omega t-\alpha t)} \tag{7.3}$$

where $e^{-j\beta t}$ represents a damping term and $e^{j\alpha t}$ is the phaser. Here E is time varying and generates a corresponding time variant orthogonal magnetic field which is expressed as

$$H = \frac{k}{\omega\mu}Ae^{-jkt+j\omega t}. \tag{7.4}$$

The surface wave impedance (Z) of the medium is defined as the ratio of electrical and magnetic field and is evaluated as

$$Z = \frac{H}{E} = \frac{k}{\omega\mu} = \sqrt{\frac{\sigma + j\omega\epsilon}{j\omega\mu}} \tag{7.5}$$

In practice, the polarized scattering map wave impedance can lead to a double direction inference such that the medium characteristic is infered from phase knowledge

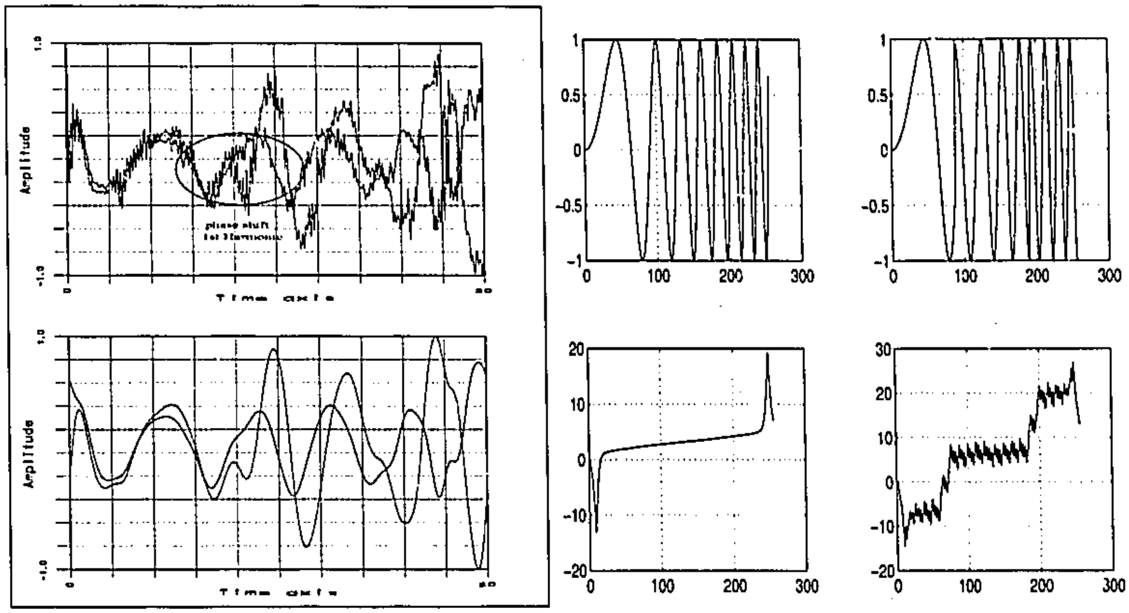


Figure 7.1: Two sets of raw data from sensors are shown in the top left plot whereas the left bottom plot had some noise filtering. The middle and right plots are synthetic, showing phase shift detection in the Fourrier transform. Top middle and right plots are in time domain whereas the bottom middle and right plots are phases in frequency domain

and vice versa.

$$phase \rightleftharpoons impedance$$

Our approach in this proposed research is to base our phase identification on the time series approximations to the basic electromagnetic wave-scattering phenomena rather than fully using Maxwell's wave field equations.

The two plots shown in figure 7.1 represent two real measurements. One measurement was taken in a single medium and the second taken in an additional medium. Since the second medium has a higher dielectric compared to the first medium, noticeable amplitude modulations can be observed. The two plots shown in figure 7.1 are raw data, and the first phase shift lies in the encirclement. In addition to the phase shift, the encirclement also encloses a reflection burst as well as an increase in the scattering polarization.

Figure 7.1 shows a synthetic complex signal with a constant frequency shift on the top left, whereas, on the top right, two phase shifts were added. The bottom plots show their corresponding phase spectrum computed from Fourrier transforms. In fact the Fourier transform between time/frequency is sufficient alone with some minor modifications to estimate any phase shift. However, as our interest is the polarization changes, the phase spectrum obtained from the signals will be used as the basis for the spatial phase maps.

7.3.3 Correction Factors: Range Resolution

An important characteristic of the resolution of short-distance radar is the aperture angle influence of the antenna. Since the radar will acquire measurements within the aperture angle, the detection resolution varies at different angles within the beam. In other words, targets that have been detected at the center beam present different attributes when detected at a different angle within the beam aperture. We adapted this range resolution factor from the airborne long range radar system [85] [94] and derived it for ground penetrating radar application.

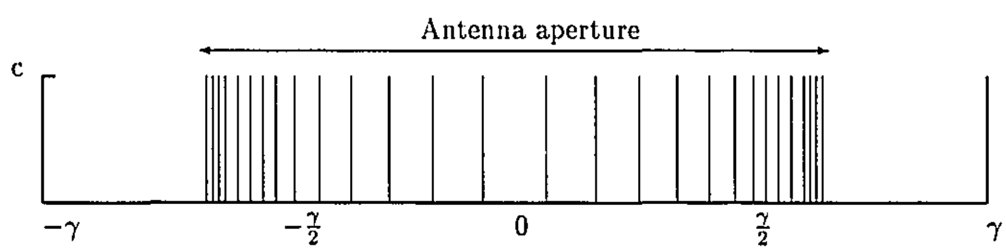


Figure 7.2: A single dimension approximated range resolution R_r for ground penetrating radar. The real aperture is elliptical (two dimensional). γ is in degrees.

Range resolution (R_r) is determined by the beam angle (Figure 7.2) and the pulse length[68]. Range resolution is theoretically equal to one-half the pulse length. The *pulse length* (τ) is the duration of the transmitted pulse and is measured in nanoseconds. It is converted from time into distance by multiplying by the speed of the electromagnetic radiation. The resulting distance is the measure of the *slant range*, or direction in which the energy propagates from the antenna to the target [79]. Range resolution, however, is expressed in *depth range*, which is the penetrated distance in the medium. Dividing the slant range by the sine of one-half of the beam aperture converts the slant-range distance into depth range distance. Hence the equation for the range resolution is

$$R_r = \frac{\tau\ c}{2\ sin\ \gamma}. \tag{7.6}$$

Therefore, given the specifications as in Table 7.1, for a maximum aperture angle 90 degrees ($\gamma = 45\,\mathrm{deg}$), and a pulse width of one nanosecond, the resolution range is 0.21 meters.

7.3.4 Correction Factors: Spatial Distortion

To remedy the spatial distortion that results from velocity change of the electromagnetic wave field in media of different properties, we analyse the (v, x, t) space – velocity/depth/time space. The two-dimensional picture construction is usually based on the assumption that the depth and time dimensions are orthogonal and confine the propagation velocity to a constant value.

The problem is criticized on the assumption regarding the properties of the wave field propagation velocities. We investigate the phase shift and reckon the velocity interval discrepancies. The space created by the velocity, time and depth dimensions hence endures a mapping over the intervals where a medium characteristic is most likely to exist. The velocity ratio can be derived from Snell's law and expressed as

$$\frac{\nu_1}{\nu_2} = \sqrt{\frac{\epsilon_2}{\epsilon_1}} \tag{7.7}$$

ν_1 and ν_2 are the velocities of the wave whereas ϵ_1 and ϵ_2 are the corresponding medium dielectrics. A typical spatial distortion recovery assumes a linear relation between time and depth which produces a gap in the sector where a real velocity change occurred. In reconsidering the aspects of the velocities, the space (x, t) defined by the depth/time dimensions is mapped accordingly and hence new coordinates are formed.

7.4 Knowledge Map (KM) Synthesis

When ground penetrating radar was devised from electromagnetic imaging, major assumptions were carried out and were rarely justified at a later stage. The initial assumptions were basically the consideration of the reflection nature of electromagnetic waves at interfaces and the concerns in imaging were mainly on the reflection aspects, under the influence of seismic work. As seismic sounding had an extremely successful application, it is very logical to use the same concepts and terminology for the ground penetrating radar counterpart applications.

The interpretation of the radar physical measurements has suffered from the fact that many variables involved are not taken into consideration. The methods used often lack statistical justification and a particular method may be chosen because it "works", because it is the only method known, or because it has become popular. On the other hand, very simple applications of statistical inverse methods may lead to significant improvements in the accuracy of estimated analysis results. Moreover, basic adjustment techniques derived from radar principles may also lead to improvements in the resolution of the inverse problem. As this section describes mainly the application of synthesis regularization to radar, the corrections factors mentioned in Section 7.3.4 play a secondary role in the inverse process but a primary importance for any forward problem simulation.

With the models described in equations 7.1, 7.3, 7.5, 7.6 and 7.7, nonlinear inversion of radar data requires computations to generate the *a priori* knowledge map distributions. The separability is based on classical signal analysis. In fact, Veno and Osumi [91] have empirically shown that the received signal from a buried

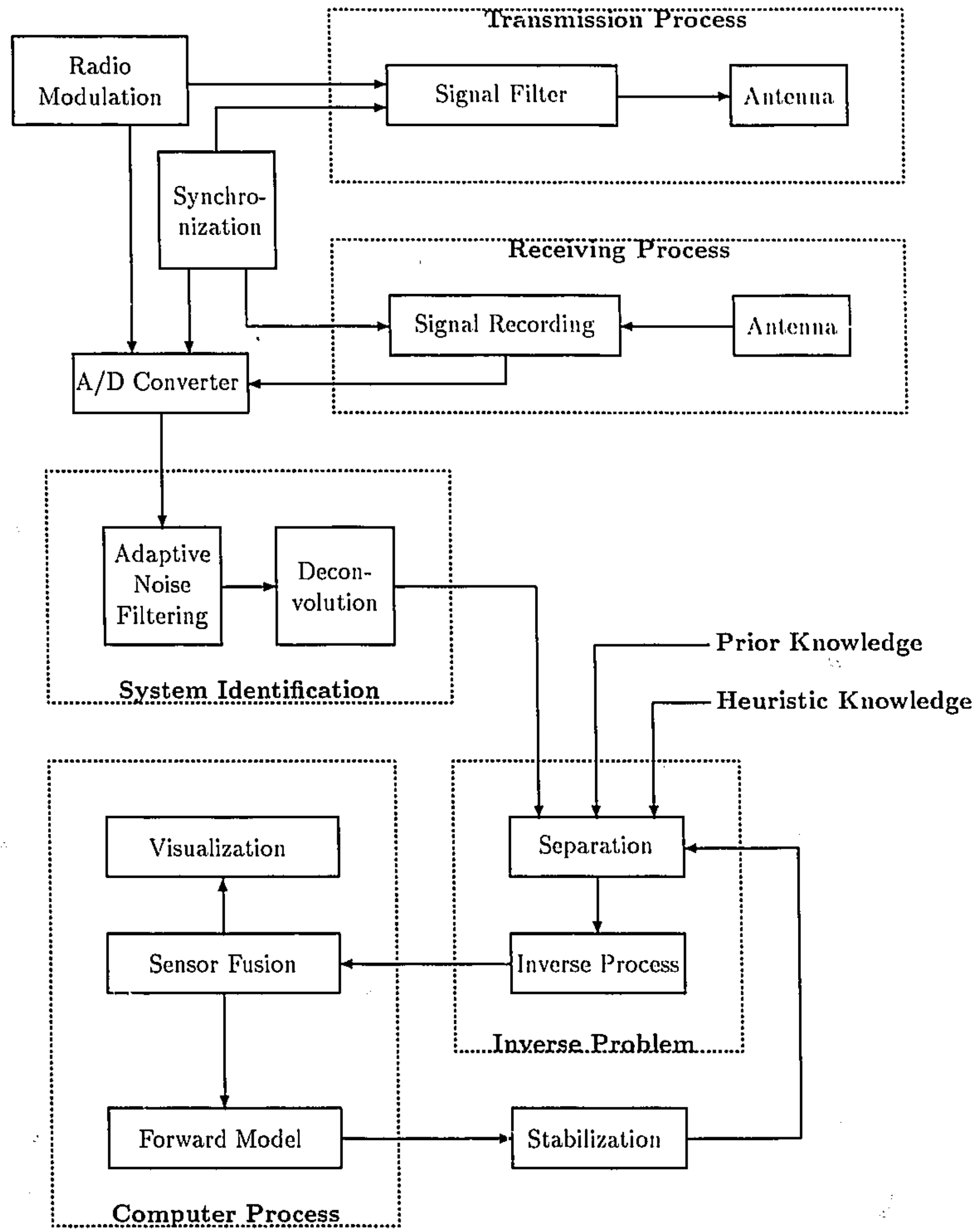


Figure 7.3: Complete overall radar operation associated with the world modeling framework.

object has the following separable form,

$$S(t,\theta,\phi) = A(t)G(\theta,\phi) \tag{7.8}$$

where $A(t)$ is the time series convolution model and $G(\theta,\phi)$ is a function of ϕ, the phase difference of antenna polarization, and θ, the rotation angle between the target axis and the antenna direction (in this thesis the angular rotation is irrelevant and $\theta = 0$). As described in chapter 3, $A(t)$ is the combination of all the impulse responses of the source impulse. Writing the covariance matrix for the synthesis process as in equation 4.14 for the knowledge maps we may then have

$$cov_{f,f^{-1}} = E\left(\begin{bmatrix} [u^i(\mathrm{h}_\phi,t)] - S^{f_\phi} \\ [u^i(\mathrm{h}_A,t)] - S^{f_A} \\ [u^i(\mathrm{h}_r,t)] - S^{f_r} \end{bmatrix}\begin{bmatrix} \left[v^i(\hat{\mathrm{h}}_\phi,t)\right] - S^{f_\phi^{-1}} \\ \left[v^i(\hat{\mathrm{h}}_A,t)\right] - S^{f_A^{-1}} \\ \left[v^i(\hat{\mathrm{h}}_r,t)\right] - S^{f_r^{-1}} \end{bmatrix}^T\right) \tag{7.9}$$

where f_k and f_k^{-1} denote the corresponding forward and inverse operators for the phase modulation, amplitude modulation and reflection knowledge maps.

Although the suggested separation process here is for three components, that is, reflection, phase and magnitude modulation, the basic *a priori* knowledge to drive the synthesis regularization will hence acquire three state variables for the front-end synthesis. First, as for all nonlinear problems, it is important to start infering at a point close to the fastest converging solution. For belief functions, the starting point is selected by the initial maximum entropy search. In fact, in belief functions, the regularization synthesis factor (λ) introduces the advantage of the momentum that forms the strategy of biased beliefs. The locus of the starting point in function of a decreasing λ follows the hyperbola of a forward problem operator starting at the center wavelet with time index zero and splits to the end of first and last wavelets. In other words, for values of λ approaching zero, the minimum entropy search tries to resolve knowledge maps with insufficient data in the beginning of the inversion process whereas, for values of λ approaching unity, the minimum entropy search resolves maximum knowledge first and moves down to insufficient knowledge.

7.4.1 Heuristic Propagation of Evidence

Again, usually there are different pieces of evidence collected from different sources in penetrating radar. One piece of evidence that one should face is that lower L band electromagnetic propagations are constrained by an elliptical aperture as in Table 7.1, which in a matter of fact justifies the multi-sensor fusion approach. Again, assume that a specific sensor z_1 is to confirm some evidence which, according to some inverse operator, results in a set of N possible solutions. If another sensor z_2 happens to confirm one solution of the solutions proposed by sensor z_1, this will result in the dis-confirmation of the rest of the $N-1$ solutions. For penetrating radar, the propagation of evidence is proposed as heuristically structured from the antenna aperture. In fact the experimental calibration of the used antenna happened to be two adjacent ellipses rather than the one as suggested in Table 7.1. Figure 7.4 elaborates on the theoretical and experimental behavior of impulse function of the antenna.

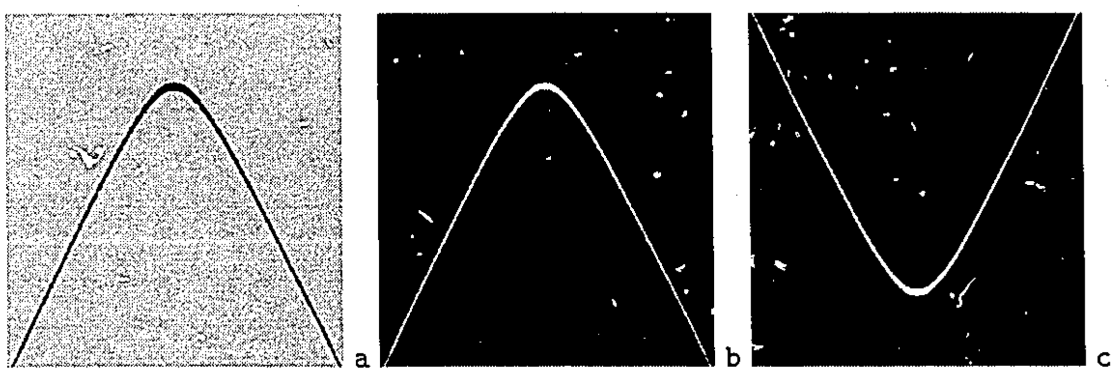


Figure 7.4: Left: The theoretical impulse response. Middle: the experimental impulse response of using the SIR radar unit of 1 GHz. Right: The experimental inverse operator that will collapse the experimental impulse response.

In contrast to most cyclic/acyclic graphs, the proposed evidential propagation function is based on *forward* and *backward* chaining. In our context, a positive proposition propagates backward and a negative proposition propagates forward.

The cross-form propagation for two knowledge units $\{u_i, u_j\}$ sharing a common sensor measurement may be expressed as

$$m(h_j) \rightarrow m(\neg h_i); \qquad m(\neg h_j) \rightarrow m(h_i). \tag{7.10}$$

Here, we devised two types of knowledge propagation among the units:

Type 1: $Bel(u_i|u_j) = BIND(Bel(u_j) \rightarrow Bel(u_i))$

Type 2: $Bel(u_i|u_j) = BIND(\alpha \rightarrow Bel(u_i))$.

α is an arbitrary sensitivity factor, whereas $BIND$ is the combination belief function as described in Section 5.2.2. Since the propagation function forms an acyclic graph, the α factor was introduced for an experimental control in the propagation as in **Type 2**.

Chapter 8

Image Perception

8.1 General Perspective

The field of Computer Vision has already resulted in ample research ramifications in image interpretation, where visual perception aspects have gained a large influence. Numerous mathematical models for representation of the neurophysiological world, whether partial or complete, have been established into concepts and paradigms. The immense complexity and the evolution of the computational aspects of vision have now given way to an increased comprehension and understanding of visual perception of the surrounding world.

Despite the advances in computational vision, informality is frequently encountered in this subject. The focus remains on the evolution of ideas rather than on models, and, likewise, the emphasis is placed more often on the classical foundations of the field instead of the current approaches. Several well established models related to a specific aspect of vision are frequently encountered. Even though the equivalency among these models is heuristic, their applications for this specific purpose are profitable for each suggested model. What is projected in this thesis are some preliminary steps of the current approaches with the intention of perceiving the computed world model.

Different models have described the spatial impression of vision perception in a formal representation such as the *stochastic relaxation* [31] or the *relaxation labeling* [47] concepts. The former presents the approach of Bayesian image reconstruction in hierarchical annealing function and the latter approach uses contextual information for finding consistent labellings of graphs. We have already approached in one form or another in this thesis the essence of both representations in the minimum entropy computation and the Dempster-Shafer belief networks. In fact, in the literature there are many imaging modes and schemes for visualizing information that, when presented under a different perspective with some adaptation, can perform a double task. In our case, that is sensor fusion as well as the basic classification of the image reconstruction. What is needed is an efficient method for quantitative data encoding and integration of the complex information into some perceivable and acceptable form. The investigation in this chapter is based on the consideration of the existence of well-founded mathematical models and paradigms. The attempt is to employ some of these concepts and to integrate information and images in an efficient presentation aiming towards a better visual perception response. The complete image derivation that can be obtained by thorough research in any possible mathematical model is unfortunately not within the scope of this thesis.

8.2 Vision and Image Perception

The initial argument in visual perception is that vision concepts mainly depend on sensed light. Such an assumption leads to the thought that the receptive field is devoted to detecting certain patterns of light and their changes, corresponding to particular relations in the visible world [55].

The regulation of light in the receptive field is considered the low level processing of visual perception. Since the images project different intensities, it is the changes of the intensities which are processed rather that the intensities themselves. Consequently, an abrupt shift in the transition between light and shade produces a contour which is the necessary condition for segregated shapes and forms. Usually, an image in the visual field is coordinated with objects in the visual world, where the perception is initiated originally by the change in brightness, which results in infering cartoon-like forms.

Curves arise from the projection of various kinds of structures in the visible world, such as contours, curvatures and discontinuities. However, the problem is that our world modeling in subsurface imaging is invisible by its physical nature and these curves are not directly observable; rather, they are abstract entities in our imagination, and the example of "a cube contained in a smaller cube" only demonstrates that the primary parameters of the "contained" cube are its curves. All that is observable in images is information about the traces of curves, and the inference upon those traces is under-constrained. An attempt at curve inferencing is not a straight forward problem given the complex situation of the infered knowledge map resulting from the sensor fusion, even though there exists the hope that the qualitative inverse problem algorithms are in some aspects neurophysiological mathematical models. On the other hand, what we try to resolve are concepts of forms that hold as a global symbolic structure and are referred to as the later vision stage, bypassing the discontinuity problems which are still impossible to classify for the visualizing of the inverted data.

8.2.1 Knowledge Interpretation and Pattern Recognition

The first stage in most pattern recognition tasks is feature extraction. Essentially, the problem in our context differs from most pattern classification problems in that we have the feature measurements and statistical distributions among the point estimates. The problem is to group the point estimates into *local features* to result in a local feature detection from the viewer perspective.

A local feature is a subset of pixels at a particular location within an image which form a recognizable pattern in their own right [44]. For example, an edge, a line or any geometric shape or arrangement of lines is a local features. Detection of a local feature within an image may be sufficient in itself to classify the entire image. In some ways, most researchers consider the concept of local feature detection almost another way of stating the fundamental problem of image recognition [99]. In a sense, detecting a local feature is a more complex problem than straightforward pattern recognition which makes the problem of image encoding more complex, since the pattern recognition part is already complex. In other words, the issue is to use approximately the reverse engineering aspects of computer vision paradigms in a way to obey the requirement of *constructing the image given some "a priori" classifications.*

Again, it can be claimed that edges are the result of an early vision process of the gradients of shades. Orientations and curvatures are to be initially extracted, yielding inferable functions for the higher vision process. Likewise, patterns and textures follow a similar behavior and can be occasionally treated as functions of curvatures and orientations, since gradients are also the basis of shading effects in higher dimensions. In addition, most references concur that visual impressions are the effect of interacting gradients [42] [33]. Spacing between edges, whether straight lines or curves, provokes a visual perspective which can induce a considerable change in the inference. Formally speaking, it is agreed that a mathematically simple gradient corresponds to a geometrically complex surface.

8.3 Image Encoding

Encoding means creating visual distinctions among several different types of objects [67]. Many different techniques can be identified, mainly color, shape, intensity and texture. A fundamental issue in any encoding technique is to determine how many possible techniques an image can have at a time. Closely related to quantitative data encoding is any means that calls for the viewer's attention to particular information and the capabilities of the viewer to separate that particular information from the rest of the image.

We will consider here the case of two-dimensional environmental modeling as an extension in the derivation of this thesis. Line segments are used to model collections of subsurface curves. Each segment can be thought of as representing a medium change, although some linear collections of observations may not correspond directly to existing structures. The curve segment models for ground penetrating radar use are appropriate where the medium characteristics can be infered.

What is of interest in this thesis are the geometric properties of images. Indeed, there is no need for the medium characteristics to be involved in image reconstruction rather than being visually infered. There is enough information about the absolute values of the image boundaries which may help in the particular image perception.

The essential constraint in image encoding is an increase in acuity of the image which is governed by the visual perception of orientations and curvatures. To achieve control of certain aspects of image generation, the constraints on the sharpness and diffusion of all the edges are extensive. In addition, as mentioned previously, arbitrary image encoding and filtering are not acceptable, either in computational vision or in general cases.

8.3.1 *CIE* Color Encoding

Levine [56] and Marr [67] support the theory of computational models based on the assumption that the best perception requires smoothness among the point estimates. In other words, the picture must be restricted to patches of uniform color with distinct delineations between them. Although the basic fact of color perception is not completely understood [54], there is a practical need to deal with color for better image perception. Therefore, in 1931 the *CIE* (Commission International de l'Eclairage) adopted a standard based on some assumptions about color definitions. Not to get involved in the neurophysiological aspects of color, the *CIE* psychophysical color mapping is a three-dimensional tristimulus space, *Red-Green-Blue*, thus increasing possible image information encoding and its relative perception.

Color enhancement could be done using pseudocolor, a methodology developed for computer graphics [28]. Pseudocolor images are created by assigning a color to gray level images according to an arbitrary transformation. A quite common color ordering is one based on optimizing information perception; however, selections of pseudocolor mapping are heuristically evaluated.

The important point here is that the *CIE* methodology does provide a practical psychophysical model for defining color as viewed by humans. The model is definitely not unique and does not really attempt to explain the underlying probability distributions behind it. There might be some interest in examining a mathematical model based on optical probabilistic processing, which attempts to deal with the probability distributions gathered from the *a posteriori* knowledge maps.

8.3.2 Segmentation by Thresholding and Stretching

Thresholding is known as a method of separating a foreground from the background in an image. A fixed threshold simply assigns a value of zero to an image pixel if it is less than the threshold, and to unity otherwise. Some researchers do not consider thresholding particularly useful as a segmentation method [1]. At a later stage in the evolution of segmentation in thresholding, the method was revised so that, instead of assigning a constant value, there was stretching of the desired segment and discarding of the rest. The final aspect of thresholding and stretching settled down to a two-valued threshold which assumes that the data points are to be greater than the first value threshold and lower than the second value threshold. The two-valued thresholds always result in either discarding one or two segments, and that is directly dependent on the relative difference of the two threshold limits.

An alternative approach to assigning constant values to the region of interest is to map the desired region over the allowable intensities which results in a *stretching* or normalization of the truncated segment. Thresholding is common among trending techniques and, since our data are classified within probabilistic regions, it is of direct importance to choose an interval of confidence for the inverted data. Whatever segmentation method used, it is helpful to think of the purpose of segmentation.

8.3.3 Encoding

Returning to the essence of the problem of encoding the knowledge map, we can rewrite the final inverted image of the inverted image as the union of two sets of estimates and their corresponding variances or

$$[v^i(\hat{\mathbf{h}}_{\mathbf{k}}, t)] = [v^i(\hat{\mathbf{h}}_{\mu}, t)] \sqcup [v^i(\hat{\mathbf{h}}_{\sigma^2}, t)]. \tag{8.1}$$

On the contrary, as one would expect a cooperative merging involving the point estimate mean and variance, the problem of interpretation and encoding is an independent feature of the two sets of estimates. The coloring scheme proposed in Section 8.3.1 results in thousands of possible colors, hence providing a hierarchal color scheme that would suit the requirements of image encoding and concur with the basic conditions for efficient image perception:

1. Perception is more sensitive to intensity differences than to absolute intensity values.

2. Neighboring point estimates are to be related and to not just occur randomly. The change has to occur randomly.

Before considering the second stage of point estimate grouping, we need to examine how the final image will manifest within the RGB color set. The concepts of edge and color feature enclosure have been introduced above as a means of representing the characteristics of patch contours. It is possible, however, to apply these concepts so that the estimate groups sharing common characteristics may be identified. In accordance with the encoding process, we choose to make a distinction between three different part types - namely the mass $\{m(h)\}$, the mass$\{m(h), m(\neg h)\}$ and the Variance σ^2. The mass $\{m(h)\}$ is defined by the *Blue* color patch configuration, giving rise to the direct confirmation enclosure measurement only, whereas the mass $\{m(h), m(\neg h)\}$ is defined by the *Green* color patch configuration giving rise to the confirmed and disconfirmed enclosure measurement combined, which is mostly governed by evidential reasoning. The third type is defined by the *Red* color patch configuration as a relevance to the point estimate variance. Based on the described estimate grouping, it is possible to intuitively understand the essential nature of what constitutes a part. For example, tracing $\{m(h)\}$ alone provides a relevant aspect of signals of high peaks in radar. This aspect does characterize the *ranging* part of radar. We can speculate that, for a relatively large radio penetration, the $\{m(h)\}$ traces are quite relevant. These operations are shown schematically in Figure 8.1

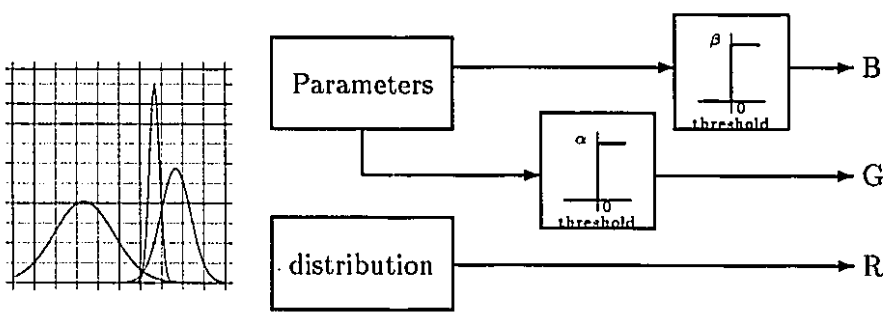


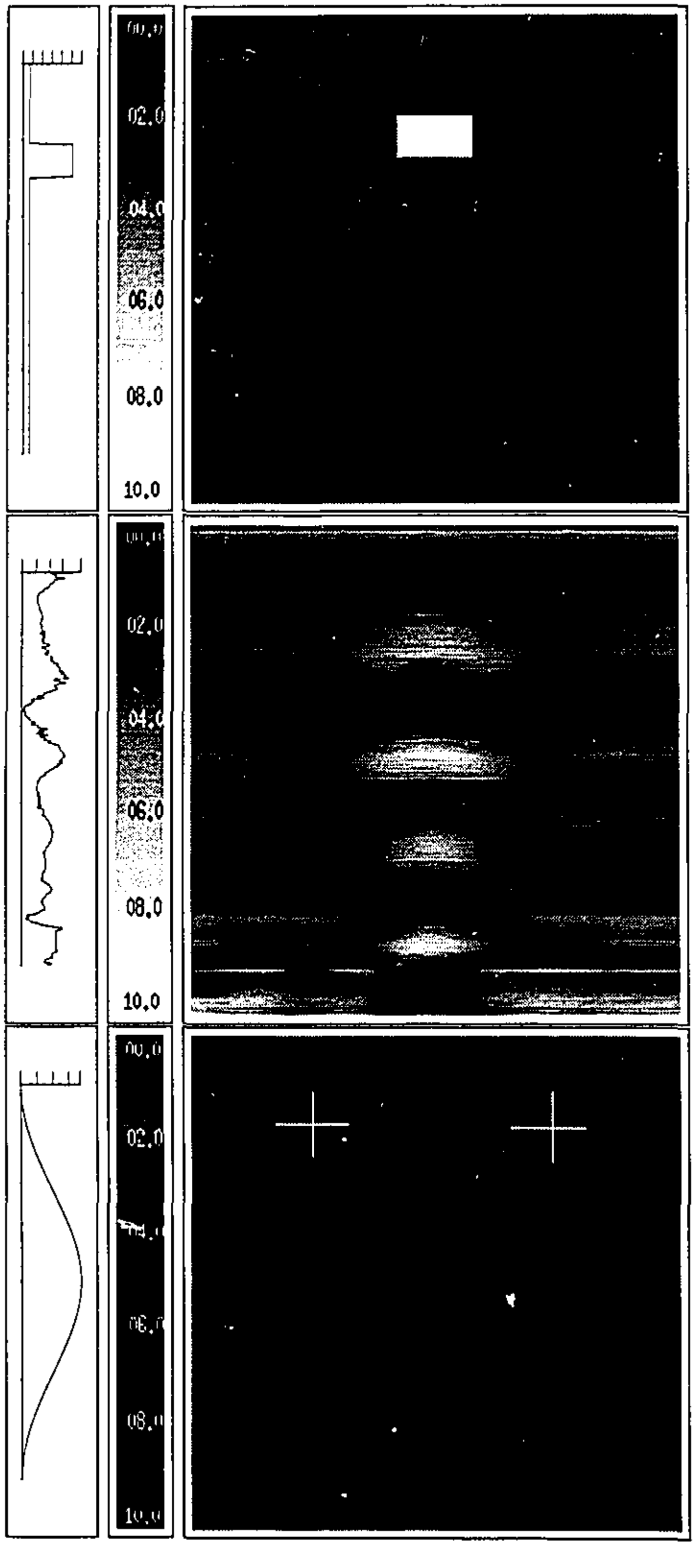
Figure 8.1: Image encoding as combination of threshold segmentation and *RGB* color mapping.

The quality of the image integration will clearly depend on the choice we make in the hierarchy among the three colors where the *Blue* color has priority over the

Green, which in turn has a priority over the *Red*. In other words, for a *Red* color to appear in the final image requires that the belief mass $\{m(h)\}$ is below a certain threshold β and the combined masses $\{m(h), m(\neg h)\}$ are below a certain threshold α. For a *Green* color to appear requires that the belief mass $\{m(h)\}$ is below the threshold β but the combined masses $\{m(h), m(\neg h)\}$ are above a certain threshold α.

8.4 Examples: Image Reconstruction

To evaluate the qualitative performance and robustness of the regularization synthesis, many different tests were performed to evaluate the sensitivity and stability margins throughout simulations and hence a radar simulator was built [35]. However, what could be the foundations of robustness of the regularization synthesis are real radar measurements. The option to evaluate the regularization synthesis through a radar simulator was based on the simplicity of scenario generation. Once the regularization synthesis was "finely tuned", selected real scenarios were designed.

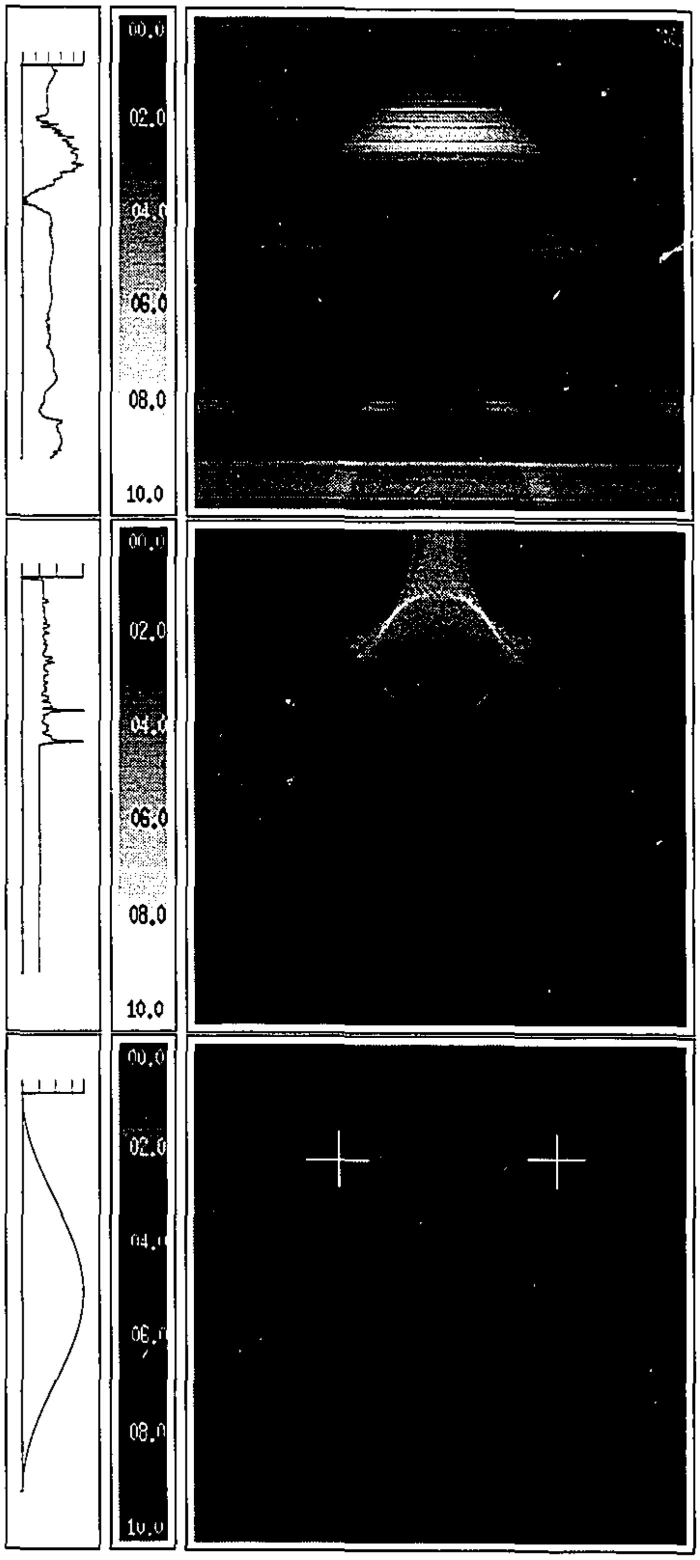


→ Hypothetical image of a metallic rectangular slab in a homogeneous medium. Dimensions are in nanoseconds which results in an approximative ~ 0.5 meters depth.

→ Raw data originating from the radar scans of the rectangular slab.The data had an exponential gain increase as function of depth prior to discretization and hence compensate for attenuation.

→ The inverse solution conforming to the regularization synthesis framework. Data have been dynamically low pass filtered (Butterworth) and deconvolved before the inverse process. The synthesis coefficients are: precision resolution: $\Delta_x = 10^{-4}$ meters; regularization factor (rf): $\lambda = 0.9$, threshold: $\alpha = 0.95$; $\beta = 0.8$.

Figure 8.2: World modeling example: real data analysis of an edged body.



→ Data originating from synthetic simulation of a cylinder. Dimensions are set in nanoseconds which results in 0.5 meters depth.

→ The inverse solution conforming to the regularization synthesis framework. The synthesis coefficients are: precision resolution: $\Delta_x = 10^{-4}$ meters; regularization factor (rf): $\lambda = 0.9$. No coloring scheme for variance.

→ The inverse solution conforming to the regularization synthesis framework. The synthesis coefficients are: precision resolution: $\Delta_x = 10^{-4}$ meters; regularization factor (rf): $\lambda = 0.9$, threshold: $\alpha = 0.8$; $\beta = 0.75$.

Figure 8.3: World modeling example: synthetic data analysis of a curved body.

Part III

Conclusion

Chapter 9

Strategic Review

Remote sensing is a diverse field, with respect to both technology and applications. Chapter 7 of this thesis described the properties of electromagnetic radiation and its interaction with matter. The nature of this interaction, which is specific for different wavelengths of radiation and different types of matter, is detected and recorded by remote sensing systems. This description characterizes the radio detection and ranging system.

At one time, the concept of multiple sensor analysis was popular; for any interpretation project, it was felt that all possible types of analyses should be acquired and interpreted. However, problems in handling the data were not perfect and disappointment was often felt. The idea of presenting a model of the world from an electromagnetic perspective did not fade but rather strongly survived turbulent changes in technology. For many decades, radars system have been considered to be at the leading edge of technology and, in the last decade, ground penetrating radar has been demonstrated to be a successful non-destructive technique.

9.1 A Brief Review

To conclude, in this chapter we recap briefly the essential parts of this thesis and give some guidance about how to proceed thereafter. Essentially the problem is that remotely measured data can be severely filtered convolutions of desired functions. Whether separating essential knowledge from the data or deconvolving plain echo signals, it is a complete mistake to disregard any prior system identification procedures, or to use a forward fitting model procedure, getting impressed by how well the model fits the data, or to use an approximation technique such as stacking inverse operators without recognizing the instability of the model solution.

Extensive work that has been achieved in the inverse problem shows clearly the impossibility of advocating a firm method for achieving such solutions for all problems since the best approach depends not only on the problem but also on the functional form of the data. Either way, optimized numerical technique or not, the fact is that the measured data in radio detection and ranging can never contain enough information to permit unique inversion according to definition 4.1. This is reflected in the classical inversion by solving integral functions where the indeterminacy is of an infinity and form an ill posed inverse problem as in definition 4.4. The most common non-classical techniques – the regularization methods – incorporate information structures of uncertainty.

It is a feature of radio detection and ranging problems that numerical solutions have often been presented by experimenters using heuristic data reduction techniques, often without being aware of the information being discarded. Researchers are impressed by these fast data reduction algorithms where already the inverse problem lacks data. As we have shown in the application of ground penetrating radar, the necessity of analysis of the polarization of the electromagnetic signal can provide as much information as the recorded reflection coefficients.

First, from a given data set, we consider some numerical experiments based on *a priori* information about an initial separable problem. This may include aspects such as the separation of the wave polarization data and the reflection coefficients from the signal intensities as shown in the example of ground penetrating radar. Although there is a good deal of heuristic reasoning in this proposed framework, it is a matter of practical experience that a good representation goes along extracting, rather than submerging, the essential information carried by the data function.

9.2 Approach to the Inverse Problem

In our model, we construct the knowledge space by standard deductive methods of the hyperbolic inverse operator but we may have no source of information about the weights of the propagation (e.g aperture angle). The question is addressed whether it is possible to compute accurately the measure of uncertainty of the inverted knowledge map. For example, it can be proven that the longer the hyperbolic sensor line (the inverse path of a knowledge unit), the larger the evidential space and the more certain a conclusion.

The approach we are taking towards developing an extended framework, which may be used to represent all of the various symbolic and numeric aspects of belief and uncertainty, is to consider a logic of argumentation. We extend the logic so that not just one argument, but all arguments, supporting or opposing a hypothesis are considered in a given decision-making context. That is the logic used to structure the inverse problem. We hold this to be the key component of a practical decision-making system. As arguments are identified, the support they confer on a hypothesis or its negation is aggregated to provide a measure of the degree of belief in the hypotheses of interest. The aggregation operation will depend on the calculus used to represent the uncertainty or vagueness associated with the arguments. The choice of calculus will in turn depend on the representation requirements and the information which is available from the given sensors.

9.3 Claim of Originality

We have described a global approach towards modeling ground penetrating radar measurements and, in particular, image synthesis and integration. One of our secondary emphases is the investigation of *Kalman filters* as a robust sensor-fusion method to perform the inverse image reconstruction, and also to degrade gracefully in front of boundary problems and non-consistent measurements. The combination of the non-discriminating sensor-fusion as a front-end method with the identification process provides a qualitative understanding of the subsurface.

Our main emphasis is based on the theory of *Dempster-Shafer* belief functions; sensor data are structured *in knowledge maps* to estimate the geometrical aspects of the model. In general, belief function estimation is usually described within a probabilistic framework. Here, we basically employ such a framework and we try to restrain our probabilistic interpretation with uncertainty factors which are well expanded in Dempster-Shafer theory. Minimizing the uncertainty will result in series of images classified within probabilistic regions. In our model, we construct the knowledge space by standard deductive methods of the hyperbolic inverse operator but we may have no source of information about the uncertainty. The question is addressed whether it is possible to compute accurately the measure of uncertainty of the inverse knowledge map. For example, it can be proven that the longer the hyperbolic sensor line (the inverse path of a knowledge unit), the larger the evidential space and the more certain a conclusion. A combination of *probabilistic reasoning* theory based on *Dempster-Shafer* belief functions, concepts of information theory and *entropy* driven search has been presented.

It is possible to provide a synthesis framework and derive solutions to the inverse problem in remote sensing for radio detection and ranging. The proposed synthesis methodologies do not claim uniqueness and there is no attempt to do so, but they nevertheless provide a robust effective solution to the inverse problem in world

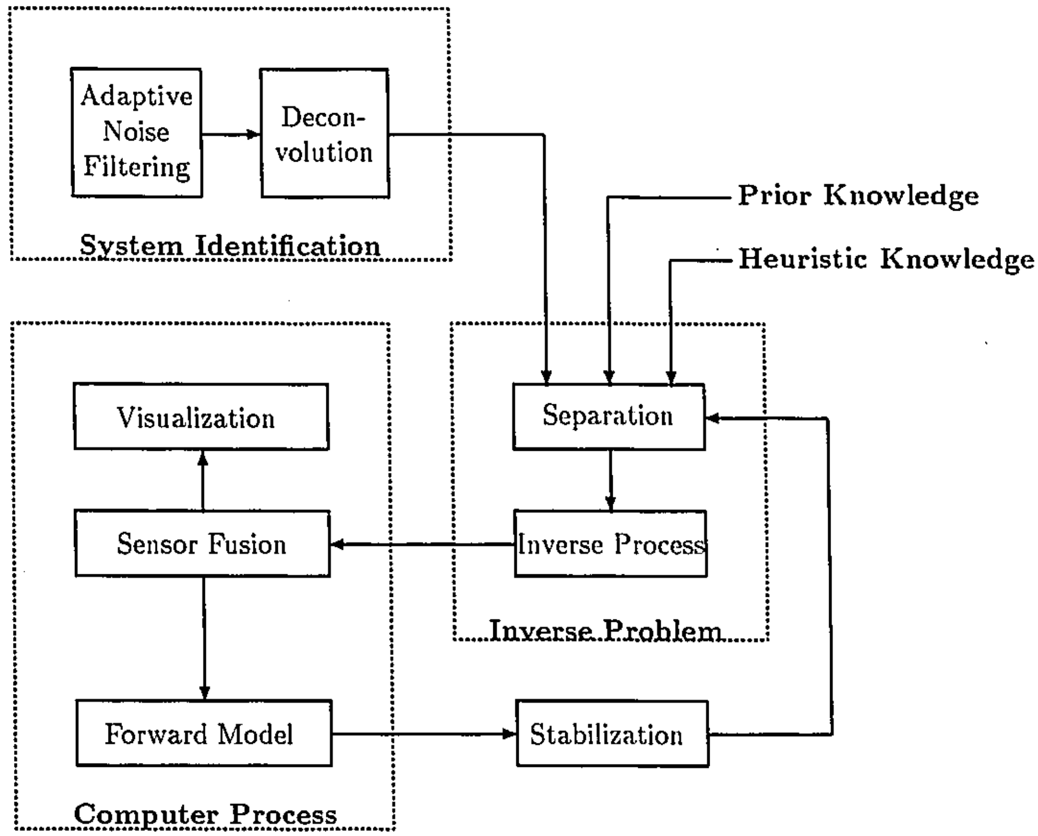


Figure 9.1: Overall operation associated with the world modeling framework.

modeling.

9.4 Recommendations

The science of remote sensing has matured perceptibly over the past decade. Only rarely are striking claims made about some new method describing "computational perfection". The true capabilities and limitations that have always been understood by remote sensing professionals are not generally understood. In the field of resource exploration, for example, people do not expect that remote sensing alone will provide them with the highest quality of information in addition to spatial and temporal efficiency. However, the implementation of advanced technology is not an easy task, especially when the technology itself is so quickly evolving.

The successes of radio detection and ranging have encouraged the development of radar technology, mainly because radar is an active system that supplies its own illumination at different wavelengths. Unfortunately, problems in controlling the illumination direction at the radio wavelength is beyond today's technology (which in fact has an indirect influence on the motivation of this thesis); however, partial control and orientation have provided for additional success of the radar system. There is no doubt that the future of radar systems is in *vision synthesis in the invisible spectrum.*

Ground penetrating radar: There has been for a while a need for height resolution sub-surface imaging that can be obtained rapidly and economically. Initially ground penetrating radar easily took its place among geophysicists and mining engineers who had already become familiar with seismic soundings. Applications for GPR are numerous and include any type of subsurface exploration, geotechnical and archeological investigations, as well as rock mechanics and mine development requirements. Some specific application examples are subsurface mapping and may include rock type changes, fracture identification and soil stratigraphy. Also geotechnical and archeological investigations can highly benefit from high resolution subsurface imaging. GPR analysis can detect and map features. Finally, space exploration cannot be excluded from benefiting from ground penetrating radar as radio detection and ranging imaging through satellites have proven that the moon and near-by planets appear favorable for similar applications.

Bibliography

[1] H. C. Andrews and B. R. Hunt. *Digital Image Restoration.* Prentice Hall, Englewood Cliffs, N.J., 1977.

[2] G. E. Backus and G. F. Gilbert. Numerical applications of a formalism for geophysical inverse problems. *Geophysical Journal of the Royal Astronomical Society*, 13:247–276, 1967.

[3] G. E. Backus and G. F. Gilbert. The resolving power of gross earth data. *Geophysical Journal of the Royal Astronomical Society*, 16:169–205, 1968.

[4] G. E. Backus and G. F. Gilbert. Uniqueness in the inversion of inaccurate gross earth data. *Philosophical Transaction of the Royal Society of London*, 266:123–192, 1970.

[5] Yaakov Bar-Shalom and Thomas E. Fortmann. *Tracking and Data Association*, volume 179. Academic Press, Orlando Florida, 1988.

[6] P. M. Barham and D. E. Humphries. Derivation of the Kalman filtering equations from elementary statistical principles. Technical Report TR-69095, National Research Council of Canada, National Aeronautical Establishment, 1969.

[7] R. H. T. Bates, Garden L. K., and T. M. Peters. Overview of computerized tomography with emphasis on future developments. *Proc. IEEE*, 71:356–372, 1983.

[8] L. Brillouin. *Science and Information Theory.* Academic Press, N.Y., 1956.

[9] C. L. Britt. Solution of electromagnetic scattering problem using time domain techniques. *IEEE Transaction on Antenna Propagation*, 3(9):1181–1192, 1989.

[10] B. G. Buchanan and T. M. Mitchell. Model-directed learning of production rules. In *In Pattern-Directed Inference Systems*, pages 23–26. Academic Press, N.Y., 1978.

[11] B. G. Buchanan and E. H. Shortliffe. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, MA, 1984.

[12] D. K. Butler. Elementary GPR overview and scope of workshop. Technical report, U.S. Army Corps of Engineers, Geotechnical Laboratory, U.S. Army Engineers Waterways Experiment Station, Vicksburg, Mississippi, 1992.

[13] R. Cassinis. *The Solution of The Inverse Problem in Geophysical Interpretation.* Plenum Press, New York, N.Y., 1981. pp. 111-136.

[14] Jon F. Claerbout. *Imaging the Earth's Interior.* Blackwell Scientific Publications, Palo Alto, California, 1985. pp. 326-346.

[15] G. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.

[16] I. Daubechies. The wavelet transform, time–frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36:961–1005, 1990.

[17] P. David and J. Voge. *Propagation of Waves.* Pergamon Press, Braunschweig, 1969.

[18] J. L. Davis and A. P. Annan. Ground-penetrating radar for high-resolution mapping of soil and rock stratigraphy. *Geophysical Prospecting*, 37:531–551, 1989.

[19] A. P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society*, 30:205–247, 1960.

[20] A. P. Dempster. *Construction and local computation aspects of network belief functions, in influence diagram, belief nets, and decision analysis.* Wiley, N.Y., 1990.

[21] G. A. DesChamps and H. S. Cabayan. Antenna synthesis and solution of inverse problems by regularization methods. *IEEE transactions on antennas and propagation*, 20:268–274, 1972.

[22] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis.* John Wiley and Sons, New York, 1973.

[23] Hugh F. Durrant-Whyte. Sensors models and multisensor integration. *The International Journal of Robotics Research*, 7(6):97–113, 1988.

[24] R. M. Fano. *Transmission of Information.* MIT Press, Cambridge, Mass., 1961.

[25] A. M. Flynn. Combining sonar and infrared sensors for mobile robot navigation. *The International Journal of Robotics Research*, 7(6):5–14, 1988.

[26] M. R. Forshaw, A. Haskell, P. F. Miller, D. J. Stanley, and J. R. G. Townshend. Spatial resolution of remotely sensed imagery. *International Journal Remote Sensing*, 4:497–420, 1983.

[27] J. Franklin. Well-posed stochastic extensions of ill-posed linear problems. *Journal of Applied Mathematics Analysis*, 31:682–716, 1970.

[28] B. Funt and J. Ho. Color from black and white. *International Journal of Computer Vision*, 3:109–117, 1989.

[29] David Gabor. Theory of communication. *IRE*, pages 429–457, 1942.

[30] T. D. Garvey, J. D. Lowrance, and M. A. Fischler. An inference technique for integrating knowledge from disparate sources. In *Proceedings of IJCAI '81*, pages 319–325., 1981.

[31] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6:721 – 741, November 1984.

[32] Joseph C. Giarratano and Gary Riley. *Expert Systems; Principles and Programming.* PWS-KENT publishing, USA, 1989.

[33] J. J. Gibson. *The Perception of the Visual World.* Greenwood Press, Connecticut, 1974.

[34] C. E. Glass, B. K. Sternberg, and M. M. Poulton. Continuous profiling subsurface monitoring using adaptive pattern recognition. *International Symposium on Mine Mechanization and Automation*, 2:17.29–17.s38, 1991.

[35] D. Goodman. Ground penetrating radar simulation in engineering and archaeology. *Geophysics*, 59(2):224–232, 1994.

[36] Ramesh Gopinath. *Wavelets and Filter Banks - New Results and Applications.* PhD thesis, Rice University, Houston, Texas, 1993.

[37] J. Gordon and E. H. Shortliffe. The Dempster-Shafer theory of evidence. In B. G. Buchanan and E. H. Shortliffe, editors, *Rule-Based Expert Systems.* Addison-Wesley, Reading, M. A., 1984.

[38] Ian J. D. Graig. *Inverse Problems in Astronomy.* Adam Hilger, Bristol, England, 1986.

[39] H. Hu H. F. Durrant-Whyte, B. Y. S. Rao. Toward a fully decentralized architecture for multi-sensor data fusion. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 2, pages 1331–1336, 1990.

[40] H. Haario. *Theory and application of inverse problems.* Longman Scientific & Technical, Essex, England, 1988.

[41] J. G. Hagerdoorn. A process of seismic reflection interpretation. *Geophysical Prospecting*, 2:85–127, 1954.

[42] H. K. Hartline. The response of single optic nerve nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology*, 121:400–415, 1938.

[43] E. Herskovits. *Computer-Based Probabilistic-Network Construction.* PhD thesis, Stanford University, California, 1991.

[44] D. D. Hoffman and W. A. Richards. Parts of recognition. *Cognition*, 18:65–96, 1984.

[45] K. I. Hopcraft and P. R. Smith. *An Introduction to Electromagnetic Inverse Scattering.* Kluwer Academic Publishers, Netherlands, Kluwer, 1992.

[46] J. L. Horner. Optical spatial filtering with the least mean square error filter. *Journal of the American Optical Society*, 59:553, 1969.

[47] R. Hummel and S. W. Zucker. On the foundation of relaxation labeling processes. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 3, May 1983.

[48] D. S. Jones. *The Theory of Electromagnetism.* Pergamin Press, Oxford, 1964.

[49] H. E. Kyburg Jr. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.

[50] P. D. Lax and R. S. Phillips. *Scattering Theory*, volume 26 of *Pure and Applied Mathematics.* Academic Press, Boston, MA, U.S.A., 1989. Revised edition. 309 pages.

[51] B. W. Leach. A Kalman filter approach to navigation on the nae convair 580 aeromagnetics research aircraft. Technical Report LR-604 NRC 19271, National Research Council of Canada, National Aeronautical Establishment, 1981.

[52] B. W. Leach. An introduction to Kalman filtering. Technical Report NAE MISC 57, National Research Council of Canada, National Aeronautical Establishment, 1984.

[53] M. Lehtinen and A. Huuskonene. The use of multipulse zero lag data to improve incoherent scatter radar power profile accuracy. *Journal One of These Pof Atmospheric and Terrestrial Physics*, 48:787–793, 1986.

[54] P. Lennie, C. Trevarther, D. Van Essen, and H. Warsh. *Visual Perception: The Neurological Foundations*, pages 103–127. Academic Press, 1990.

[55] J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H Pitts. What the frog's eye tells the frog's brain. *Proc. Inst. Radio Engrs.*, 47:1940–1951, 1959.

[56] Martin D. Levine. *Vision in Man and Machine*, chapter 3 and 4, pages 89–121. Addison-Wesley, 1984.

[57] N. Levinson. The Weiner RMS error criterion in filter design and prediction. *Journal of Mathematics and Physics*, 25:261–278, 1947.

[58] J. Liu, M. Desmarais, H. Ying, and A. Maluf. Exploring the effectiveness and exactness of evidential inferences with algorithmically induced implication networks: An empirical investigation. *Journal of Artificial Intelligence*, 1995. under review.

[59] J. Liu, M. Desmarais, H. Ying, and A. Maluf. Experimental validations of an implication network induction method. In *International Joint Conference on Artificial Intelligence*, under review.

[60] L. Ljung. *System Identification Theory For The User.* Prentice Hall, New Jersey, 1987.

[61] R.D. Luzitano and T. J. Ulrych. Instantaneous polarization match: Identifies amplitude changes due to anomalous polarization in GPR data. In *Sixth Canadian Symposium on Mining Automation*, 1994.

[62] J. Makhoul. Linear prediction: A tutorial review. *Proceeding of IEEE*, 63:561–580, 1975.

[63] A. Maluf, L. Daneshmend, F. Hasani, and M. Momayez. Image synthesis and perception of ground probing radar. In *Fifth International Conference on Ground Penetrating Radar*, June 1994.

[64] A. Maluf, L. K. Daneshmend, and F. Hasani. Recovering from spatial distortion in electromagnetic imaging. In *Sixth Canadian Symposium on Mining Automation*, 1994.

[65] Ameen Maluf, Michel Desmarais, and Jiming Liu. Entropy computation in belief networks of mixed functions. In *International Conference on Uncertainty in Artificial Intelligence*, 1995. under review.

[66] Ameen Maluf, Michel Desmarais, and Jiming Liu. Uncertainty and entropy computation in belief networks. *International Journal of Intelligent Systems*, 1995. under review.

[67] D. Marr. *Vision.* W. H. Freeman and Company, San Francisco, 1982.

[68] R. K. Moore. Radar fundamentals and scatterometters. In R. N. Colwell, editor, *Manual of Remote Sensing, Second Edition*, pages 369–427. American Society of Photogrametry, 1983.

[69] Raymond Cunningham Murphy. Inverse scattering by conducting circular cylinders. Master's thesis, McGill University, Montréal, Québec, 1971.

[70] Jim Osborn. Automated subsurface mapping. Technical report in robotics, Field Robotics Center, Carnegie Mellon University, 1990.

[71] Athanasios Papoulis. *Probability, Random Variables, and Stochastic Processes.* Series in Electrical Engineering. Communications and Signal Processing. McGraw–Hill, New York, N.Y., 3^{rd} edition, 1991.

[72] J. Pearl. Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 29:241–288, 1986.

[73] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, CA, 1988.

[74] Paul E. Pfeiffer. *Probability for Applications.* Springer Texts in Statistics. Springer–Verlag, Berlin, Germany, 1990.

[75] Mary Moens Poulton. *Neural Network Pattern Recognition of Electromagnetic Ellipticity Images.* PhD thesis, University of Arizona, Arizona, 1990.

[76] J. L. Powell and B. Crasemann. *Quantum Mechanics.* Addison Wesley, Reading, Mass., 1961.

[77] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Processing Magazine*, 8(4):14–38, October 1991.

[78] Enders A. Robinson and Sven Treitel. *Geophysical Signal Analysis.* Prentice-Hall, Englewood Cliffs, N.J., 1980. ch. 1-3.

[79] F. F. Sabins. Geologic mapping of death valley from thematic mapper, thermal infrared, and radar images. In *Third Thematic Conference, Remote Sensing for Exploration Geology*, pages 139–152, Environmental Research Institute of Michigan, 1984.

[80] L. Schwartz. *Théorie des Distributions.* Hermann, 293, rue Lecourbe, 75015 Paris, France, 1966.

[81] G. Shafer. *A mathematical Theory of Evidence.* Princeton University Press, Princeton, NJ, 1976.

[82] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J,* 27:379–423, 623–656, 1948.

[83] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication.* University of Illinois Press, Urbana, Ill., 1962.

[84] Allen Shapiro. *Structured induction in expert systems.* Addison-Wesley, Reading, MA, 1987.

[85] M. I. Skolnik. *Radar Handbook.* McGraw-Hill, N.Y., 1970.

[86] Glenn S. Smith and Waymond R. Scott. A scale model for studying ground penetrating radars. *IEEE Transaction on Geoscience and Remote Sensing,* 27:358–363, 1989.

[87] D. M. Sullivan. Frequency-dependent FDTD methods using z transforms. *IEEE Transaction on Antenna Propagation,* 40(10):1223–1230, 1992.

[88] A. Tarantola. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation.* Elsevier Science Publishers B.V., N.Y., 1987. ch. 3.

[89] A. G. Tijhuis. *Electromagnetic Inverse profiling: Theory and Numerical Implementation.* VNU Science Press BV, Utrecht, The Netherlands, 1987. pp. 156-228.

[90] J. M. Tribolet. Applications of short-time homomorphic signal analysis to seismic wavelet estimation. *IEEE Transaction in Acoustic and Speech Signal Processing,* 26:343–353, 1978.

[91] K. Veno and N. Osumi. Underground pipe detection based on microwave polarized effect. In *Proceedings of the International Symposium on Noise and Clutter Rejection,* Tokyo, 1984.

[92] N. Weiner. *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications.* Wiley, New York, 1949.

[93] N. Weiner. *Cybernetics.* MIT press, Cambridge, Mass., 1965.

[94] R. S. Wing. Structural analysis from radar imagery. *Modern Geology*, 2:1–21, 1971.

[95] T. S. Yeo. *Diffraction by a Penetrable Wedge.* PhD thesis, University of Canterbury, Christchurch, New Zealand, 1985.

[96] M. Yong and D. C. Wilkins. Induction of uncertain rules and the sociopathicity property in Dempster-Shafer theory. In R. Kruse and P. Siegel, editors, *Symbolic and Quantitative Approaches to Uncertainty*, pages 238–245. Springer-Verlag, 1991.

[97] Siu Fung Yu. *Recursive State-Space Approach to Ground Probing Radar Signal Processing.* PhD thesis, University of Lancaster, Lancaster, United Kingdom, 1987.

[98] L. A. Zadeh. Fuzzy logics and approximate reasoning. *Synthese*, 30:407–428, 1975.

[99] S. W. Zucker, R. A. Hummel, and A. Rosenfeld. An application of relaxation labelling to line and curve enhancement. *IEEE Transactions on computers*, C–26:394–403, 1977.

Appendix A

Electromagnetic Wave Characteristics

A.1 Review of Basic Wave Characteristics and Motion

A *Wave* is a disturbance in a medium such that each particle in the medium vibrates about an equilibrium point in a simple harmonic motion. The direction of the vibration is perpendicular to the direction of propagation of the wave, and the wave is called a *transverse wave.* Many characteristics represent the wave behavior. The *wavelength*, written as λ, is the theoretical or measured distance from the crest to crest (or valley to valley) of a transverse wave. It may also be defined as the distance between two particles with the same displacement and direction of displacement. The *amplitude* is the maximum displacement of a particle in one direction from its equilibrium point. The *frequency*, usually written as f, is the number of wavelength (cycles) that pass per unit time. The *period*, usually written as T, is the time required for one wavelength to pass a point. The *velocity*, usually written as c, of a wave refers to its propagation velocity through the medium. The *phase*, usually written as ϕ, is the difference in displacement and direction of a particle due to two different waves, that is two waves are in phase if each particle has the same displacement and direction of motion ($\phi = 0$).

The basic wave characteristics are related to each other by the laws of physics. The wave velocity is proportional to the wavelength and frequency and

$$c = \lambda f. \tag{A.1}$$

The frequency component is inversely proportional to the period and

$$f = \frac{1}{T}. \tag{A.2}$$

Some basic principles apply to the behavior of the wave characteristics. The *Superposition principle* state that the effects of two or more waves on the displacement of a particle are independent. This means the displacement of a particle by a simultaneous wave in a medium is algebraically additive. *Interference* is the summation of the displacements of different waves in a medium. *Constructive interference* is when the waves add up to a larger resultant wave than their original. This occurs maximally when the phase difference (ϕ) is a whole wavelength (λ) which correspond to multiples of 2π. *Destructive interference* is when the waves add up to a smaller resultant wave than either original wave. Variation of the interferences can be extended to closer analysis on the complex harmonics wavelength. *Standing waves* result when waves are reflected off stationary coordinates back into the oncoming waves of the medium, and super-imposition results. Constructive and destructive interference dominate the standing wave's behavior.

A.2 Electromagnetic Wave Propagation

Maxwell's (James C. Maxwell) electromagnetic field equations are [50] [48]:

$$\nabla \times E + \frac{\partial B}{\partial t} = 0 \tag{A.3}$$

$$\nabla \times H - \frac{\partial D}{\partial t} = J \tag{A.4}$$

$$\nabla . J + \frac{\partial s}{\partial t} = 0 \tag{A.5}$$

$$\nabla . B = 0 \tag{A.6}$$

$$\nabla . D = 0 \tag{A.7}$$

where

E: the electric field vector,	H: the magnetic field number,
D: the electric flux density vector,	B: the magnetic flux density vector,
J: the electric current density vector,	s: the electric charge density

and

$$J = \sigma E \tag{A.8}$$

$$B = \mu H \tag{A.9}$$

$$D = \epsilon E \tag{A.10}$$

in addition to

$$\mu = \mu_0 \mu_r, \qquad \epsilon = \epsilon_0 \epsilon_r, \qquad \epsilon_r = \nu^2 \tag{A.11}$$

with

σ : conductivity	ν :refractive index	c : speed of light
μ : permeability	μ_r : relative permeability	μ_0 : permeability in free space
ϵ : permittivity	ϵ_r : relative permittivity	ϵ_0 : permittivity in free space

Index